

Springer Series in Reliability Engineering

Hoang Pham *Editor*

Applications in Reliability and Statistical Computing

 Springer

Springer Series in Reliability Engineering

Series Editor

Hoang Pham, Department of Industrial and Systems Engineering,
Rutgers University, Piscataway, NJ, USA

Today's modern systems have become increasingly complex to design and build, while the demand for reliability and cost effective development continues. Reliability is one of the most important attributes in all these systems, including aerospace applications, real-time control, medical applications, defense systems, human decision-making, and home-security products. Growing international competition has increased the need for all designers, managers, practitioners, scientists and engineers to ensure a level of reliability of their product before release at the lowest cost. The interest in reliability has been growing in recent years and this trend will continue during the next decade and beyond.

The Springer Series in Reliability Engineering publishes books, monographs and edited volumes in important areas of current theoretical research development in reliability and in areas that attempt to bridge the gap between theory and application in areas of interest to practitioners in industry, laboratories, business, and government.

Now with 100 volumes!

Indexed in Scopus and EI Compendex

Interested authors should contact the series editor, Hoang Pham, Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ 08854, USA. Email: hopham@soe.rutgers.edu, or Anthony Doyle, Executive Editor, Springer, London. Email: anthony.doyle@springer.com.

Hoang Pham
Editor

Applications in Reliability and Statistical Computing

 Springer

Editor

Hoang Pham
Department of Industrial and Systems
Engineering
Rutgers, The State University of New Jersey
Piscataway, NJ, USA

ISSN 1614-7839

ISSN 2196-999X (electronic)

Springer Series in Reliability Engineering

ISBN 978-3-031-21231-4

ISBN 978-3-031-21232-1 (eBook)

<https://doi.org/10.1007/978-3-031-21232-1>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To My Father on His 90th Birthday!

Preface

We're living in an era of fast and unpredictable change. Billions of people are connected to each other through their mobile devices. Data is being collected and processed each day like never before. With 5G and IoT set to generate an estimated 1 billion terabytes of data by 2025, companies continue to search for new techniques and tools that can help them practice data collection effectively in promoting their business. A large portion of this data will come from smart devices, smart communities. The era of big data through reliability and statistical computing with almost all applications in our daily life has experienced a dramatic shift in the past two decades to a truly global industry. The forces that have driven this change are still at play and will continue. Most of the products which affect our daily lives are becoming even more complex than ever.

The book consists of 15 chapters that covers a selection of recent developments and applications on various related topics in reliability and statistical computing. The emphasis of this book is on the practical applications of reliability and statistical methods and techniques in various disciplines using machine learning, risk assessment, modeling and optimization, and other computational methods.

All chapters in the book are written by leading researchers and practitioners in their respective fields with a hope to connect the gap between the theoretical and practical computations in the application areas of reliability and statistical computing.

I acknowledge Springer for this opportunity and professional support. Importantly, I would like to thank all the chapter authors and reviewers for their availability for this work.

Piscataway, NJ, USA
September 2022

Hoang Pham

Contents

Forecasting The Long-Term Growth of S&P 500 Index	1
Stephen H.-T. Lihn	
Smart Maintenance and Human Factor Modeling for Aircraft Safety	25
Eric T. T. Wong and W. Y. Man	
Feedback-Based Algorithm for Negotiating Human Preferences and Making Risk Assessment Decisions	61
Silvia Carpitella, Antonella Certa, and Joaquín Izquierdo	
Joining Aspect Detection and Opinion Target Expression Based on Multi-Deep Learning Models	85
Bui Thanh Hung	
Voting Systems with Supervising Mechanisms	97
Tingnan Lin and Hoang Pham	
Assessing the Severity of COVID-19 in the United States	117
Kehan Gao, Sarah Tasneem, and Taghi Khoshgoftaar	
Promoting Expert Knowledge for Comprehensive Human Risk Management in Industrial Environments	135
Ilyas Mzougui, Silvia Carpitella, and Joaquín Izquierdo	
Data Quality Assessment for ML Decision-Making	163
Alexandra-Ştefania Moloiu, Grigore Albeanu, Henrik Madsen, and Florin Popenţiu-Vlădicescu	
From Holistic Health to Holistic Reliability—Toward an Integration of Classical Reliability with Modern Big-Data Based Health Monitoring	179
Fengbin Sun	

On the Aspects of Vitamin D and COVID-19 Infections and Modeling Time-Delay Body’s Immune System with Time-Dependent Effects of Vitamin D and Probiotic 201
Hoang Pham

A Staff Scheduling Problem of Customers with Reservations in Consideration with Expected Wait Time of a Customer without Reservation 219
Junji Koyanagi

Decision Support System for Ranking of Software Reliability Growth Models 227
Devanshu Kumar Singh, Hitesh, Vijay Kumar, and Hoang Pham

Human Pose Estimation Using Artificial Intelligence 245
Himanshu Sharma, Anshul Tickoo, Avinash K. Shrivastava, and Umer Khan

Neural Network Modeling and What-If Scenarios: Applications for Market Development Forecasting 271
Valentina Kuskova, Dmitry Zaytsev, Gregory Khvatsky, and Anna Sokol

Mental Health Studies: A Review 289
Rachel Wesley and Hoang Pham

Editor and Contributors

About the Editor

Hoang Pham is a Distinguished Professor and former Chairman (2007–2013) of the Department of Industrial and Systems Engineering at Rutgers University. Before joining Rutgers in 1993, he was a Senior Engineering Specialist with the Idaho National Engineering Laboratory, Idaho Falls, Idaho and Boeing Company in Seattle, Washington. His research areas include reliability modeling and prediction, software reliability, and statistical inference. He is editor-in-chief of the International Journal of Reliability, Quality, and Safety Engineering and editor of the Springer Series in Reliability Engineering and has been conference chair and program chair of over 50 international conferences and workshops. Dr. Pham is the author or coauthor of 7 books and has published over 200 journal articles and 100 conference papers, and edited 17 books including Springer Handbook in Engineering Statistics and Handbook in Reliability Engineering. He has delivered over 40 invited keynote and plenary speeches at many international conferences and institutions. His numerous awards include the 2009 IEEE Reliability Society Engineer of the Year Award. He is a Fellow of the IEEE, AAIA, and IISE.

Contributors

Grigore Albeanu “Spiru Haret” University, Bucharest, Romania

Silvia Carpitella Department of Manufacturing Systems Engineering and Management, California State University, Northridge, USA

Antonella Certa Department of Engineering, University of Palermo, Palermo, Italy

Kehan Gao Department of Computer Science, Eastern Connecticut State University, Willimantic, CT, USA

Hitesh Department of Computer Science & Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

Bui Thanh Hung Data Science Laboratory, Faculty of Information Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

Joaquín Izquierdo Institute for Multidisciplinary Mathematics, Universitat Politècnica de València, Valencia, Spain

Umer Khan Amity University, Noida, Uttar Pradesh, India

Taghi Khoshgoftaar Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, USA

Gregory Khvatsky Lucy Family Institute for Data & Society, Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

Junji Koyanagi Graduate School of Tottori University, Tottori-City, Tottori, Japan

Vijay Kumar Department of Mathematics, Amity Institute of Applied Sciences, Amity University, Noida, Uttar Pradesh, India

Valentina Kuskova Lucy Family Institute for Data & Society, University of Notre Dame, Notre Dame, IN, USA

Stephen H.-T. Lih Quant Research at Atom Investors LP, Austin, TX, USA

Tingnan Lin Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

Henrik Madsen Danish Technical University, Lyngby, Denmark

W. Y. Man Department of Engineering (Avionics), Heliservices (HK) Ltd, Hong Kong SAR, China

Alexandra-Ştefania Moloiu R&D Department, TypingDNA, Bucharest, Romania

Ilyas Mzougui Faculty of Sciences and Technologies, Abdelmalek Essaadi University, Tangier, Morocco

Hoang Pham Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

Florin Popențiu-Vlădicescu University “Politehnica” of Bucharest & Academy of Romanian Scientists, Bucharest, Romania

Himanshu Sharma Amity University, Noida, Uttar Pradesh, India

Avinash K. Shrivastava International Management Institute, Kolkata, West Bengal, India

Devanshu Kumar Singh Department of Computer Science & Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

Anna Sokol Lucy Family Institute for Data & Society, Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

Fengbin Sun Reliability Engineering, Tesla Inc, Palo Alto, CA, USA

Sarah Tasneem Department of Computer Science, Eastern Connecticut State University, Willimantic, CT, USA

Anshul Tickoo Amity University, Noida, Uttar Pradesh, India

Rachel Wesley Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

Eric T. T. Wong Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China

Dmitry Zaytsev University of Notre Dame at Tantur, Jerusalem, Israel

Forecasting The Long-Term Growth of S&P 500 Index



Stephen H.-T. Lihn

Keywords Forecast model · Economic cycle · Mean reversion · CAPE · Stock market · Wavelet

JEL Classification: C38 · C53 · E32 · E37 · E47

1 Introduction

The U.S. stock market has exhibited amazing resilience in the long run. Its long-term growth is a wonderful story of American capitalism. In the past 200 years, it has produced a consistent real return of about 6.6% per year (Fig. 1 and Siegel [20]). However, this wonderful return comes with many ups and downs every decade. In some cases, the market went down more than 50%. In other cases, the market was stagnant for more than a decade. The longest and largest drawdown in history was from 1929 to 1948. More recently, the peak reached in 2000 had not been surpassed until 2012. Making things more intricate, these two large bear markets were preceded by two strongest ten-year bull markets in history. How do we make sense of them? More importantly, are they forecastable?

Disclaimer: The views in this paper are solely the responsibility of the author. They don't reflect the views of the company, nor does this paper contain any proprietary data from the company, v1.0, Released on September 30, 2021.

S. H.-T. Lihn (✉)

Quant Research at Atom Investors LP, Austin, TX, USA

e-mail: stevelihn@gmail.com

URL: <http://www.linkedin.com/pub/stephen-horng-twu-lihn/0/71a/65>

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Pham (ed.), *Applications in Reliability and Statistical Computing*, Springer Series
in Reliability Engineering, https://doi.org/10.1007/978-3-031-21232-1_1

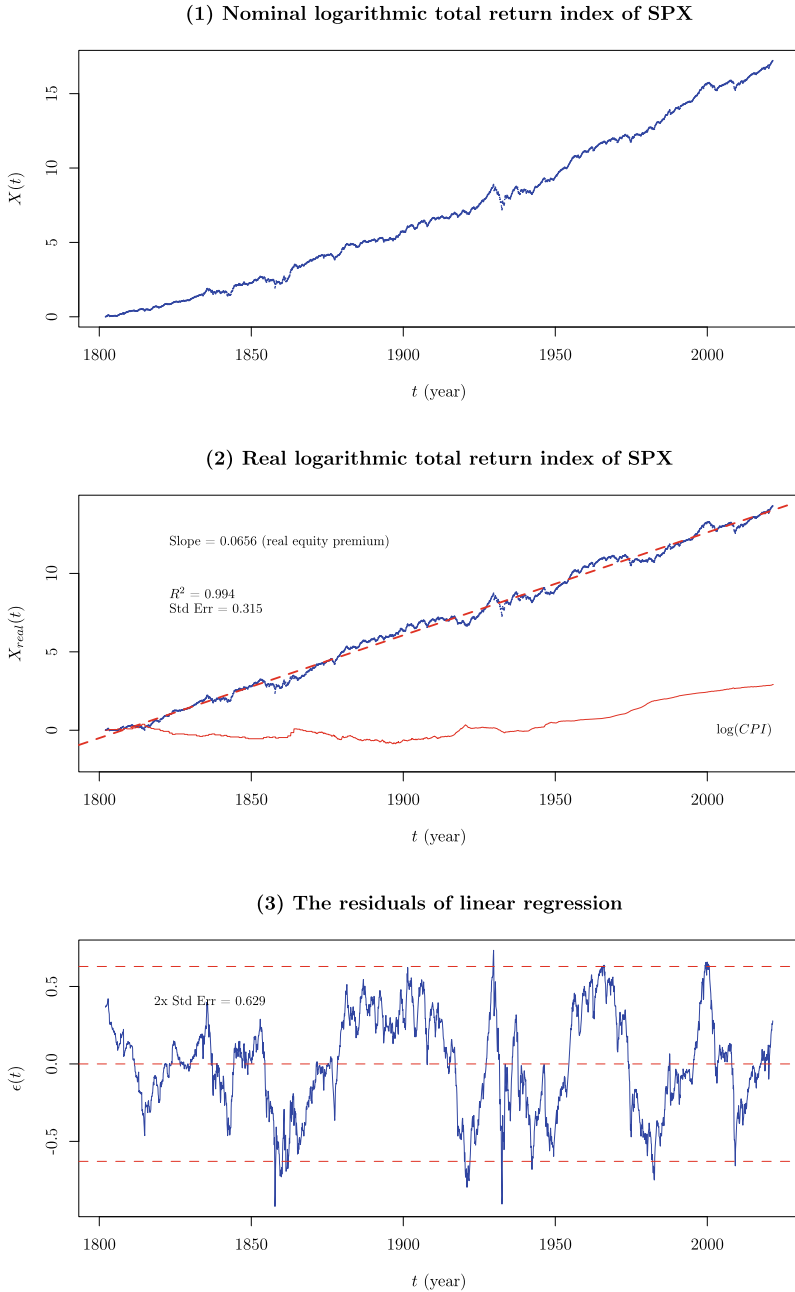


Fig. 1 Panel (1) The nominal total return index for the U.S. stock market $X(t)$ in the logarithmic scale since 1802. Panel (2) shows the real total return index $X_{\text{real}}(t)$. The slope $\beta_{\text{rep}} = 6.55\%$ is the long-term real equity premium over inflation. The linear regression has an impressive $R^2 = 0.994$ with the standard error of 0.32. Panel (3) shows the mean-reverting behavior of the residuals $\epsilon(t)$

For most typical investors, It is believed that the S&P 500 index (SPX) is the best single gauge of the U.S. stock market.¹ This index consists of 500 largest public corporations in the U.S., weighted by their market capitalizations. In a 2017 interview with CNBC,² Warren Buffett said, “Consistently buy an S&P 500 low-cost index fund, I think it’s the thing that makes the most sense practically all of the time.” At the 2021 Berkshire Hathaway annual meeting, he reiterated his conviction, “I just think that the best thing to do is buy 90% in S&P 500 index fund.”³ What is the rationale behind these statements? How much faith should we have in it? What kind of returns can be expected from SPX if we “surrender our freedom”, so to speak, of selecting from thousands of stocks, mutual funds and ETFs. This research is intended to answer some of these questions in an econometric setting.

Recent application of trend filtering technique has revealed linear characteristics of market trends [16]. In the short term, the market process is highly leptokurtotic (kurtosis $\gg 3$) and influenced heavily by the underlying volatility process. In the long term, however, the market process is not a random walk process. It is a mean-reverting process with linear growth. More interestingly, when the time horizon is extended to decades, the mean-reverting process is slightly platykurtic (kurtosis < 3), which is strikingly different from the leptokurtotic random walk process observed in the short term.

The mean-reverting process can be confirmed by the model-free wavelet analysis. The Morlet wavelet [13, 14] provides the ability to decipher the market cycles in a financial time series. By applying the wavelet analysis to both the 10-year and 20-year returns, we are able to show that the U.S. stock market exhibited a 36-year cycle after World War II (WWII).

Next, we review the algorithm developed in [10] that separates the mean-reversion component from the linear growth component in the market process. The mean-reversion component is associated with the “cyclically adjusted P/E ratio”, aka CAPE [1], in a profound way. The nickname of our model is called “jubilee tectonic model”. The “jubilee” name comes from its optimal trend-following window of 45 years and the periodicity of 36 years from the wavelet analysis. The “tectonic” name comes from the hypothesis that there are fault lines in the historical CAPE, which can be calibrated and corrected in this model through statistical learning. Such “model breaks” have been categorically discussed in Chap. 19 of [7]. We apply a more restrictive approach to capture these breaks, and attempt to give them economic interpretation when appropriate.

The forecast of future equity return is an important topic for policy makers and asset allocators. Research from Vanguard [6] found that “many commonly cited signals have had very weak and erratic correlations with actual subsequent returns.” CAPE remains one of the most powerful predictors. Even then, it has explained only

¹ <https://www.spglobal.com/spdji/en/indices/equity/sp-500/>.

² <https://www.cnbc.com/2017/05/12/warren-buffett-says-index-funds-make-the-best-retirement-sense-practically-all-the-time.html>.

³ <https://www.cnbc.com/2021/05/03/investing-lessons-from-warren-buffett-at-berkshire-hathaway-meeting.html>.

about 34% of the time variation.⁴ Recent forecasts using CAPE have been over-pessimistic. The lofty CAPE issue continues to trouble the academic community, as [24] wrote on Project Syndicate: “It is impossible to pin down the full cause of the high price of the U.S. stock market.” In an attempt to address such issue, Siegel [21] studied six variations: reported earnings, operating earnings, and NIPA profits, in combination with price index portfolio and total return portfolio. The R^2 was increased from 34 to 40% in the best case scenario.

In the jubilee tectonic model, the tectonically adjusted CAPE, plus mean reversion and inflation, form the five-factor econometric model that forecasts long-term equity returns with R^2 above 80%. This model produces different predictions for the future: The original CAPE model predicts below average real returns for the next decade. But the jubilee tectonic model predicts much higher returns and very positive outlook for the next decade.

1.1 Objectives

The key points of this chapter are:

- Setup, global linear regression, and equity risk premium
- Wavelet analysis on periodicity
- Channel deviation framework and CAPE
- The 20-year forecast model.

1.2 Data Sources, Tools, and Abbreviations

This chapter uses the **jubilee** package [11] and the **WaveletComp** package [17] in R to produce the analysis. The S&P 500 data in the **jubilee** package is assembled from several original sources. The main data source is from Shiller’s online data website [23]. The excel file “ie_data.xls” contains monthly averaged prices, dividends, and earnings of SPX since 1871.⁵ It also contains consumer price index (CPI) and 10-year Treasury yield (GS10). It derives the real prices, real dividends, and real earnings, and calculates the 10-year CAPE.

The second data source is from Schwert [19], from which we obtain the stock market total return data since January of 1802. The third data source is the annual CPI data since 1800 from Minneapolis FED [12]. The fourth data source is from FRED [9] of St Louis FED, which provides daily and/or monthly online updates for many financial and economic time series.

Frequently used abbreviations are listed below:

⁴ The 40% R^2 cited in [6] is a result of truncating the CAPE data prior to 1926. Such structural break can be explained by this research.

⁵ The word “ie” stands for “Irrational Exuberance”. It is a March 2000 book written by Shiller: [https://en.wikipedia.org/wiki/Irrational_Exuberance_\(book\)](https://en.wikipedia.org/wiki/Irrational_Exuberance_(book)).

List of Abbreviations

CAPE	Cyclically adjusted P/E ratio
CPI	Consumer price index
ETF	Exchange trade fund
GS10	Ten-year Treasury yield
SPX	The SP 500 index
TRI	Total return index.

2 Total Return Index and Equity Risk Premium

We first define the methodology of calculating stock market's logarithmic total return index (TRI). For the long-term analysis, we work with monthly interval $\Delta t = 1/12$. Assume the market index at time t is $p(t)$ and pays dividend $d(t)$ for the period from t to $t + \Delta t$. The total log-return is $r(t + \Delta t) \equiv \log(p(t + \Delta t) + d(t)) - \log p(t)$. And let $CPI(t)$ be the consumer price index (CPI) at time t , we construct the nominal and real TRI in logarithmic scale as

$$\begin{aligned} X(t) &= \sum_{t_1 \leq \tau \leq t} r(\tau), && \text{nominal TRI;} \\ X_{\text{real}}(t) &= X(t) - \log CPI(t), && \text{real TRI.} \end{aligned} \tag{1}$$

where $\{\tau\}$ represents all the months available to our analysis, and t_1 is the inception date of the data, January of 1802.

The above notation of $X(t)$ is the “continuous notation”. Empirically, t is discrete. The “discrete notation” states that, at time t_i , the logarithmic index value is X_i . We use both notations depending on the context and the cleanliness of expression. We follow Shiller's convention that each month is identified by the time fraction of $t_i = y(t_i) + (m(t_i) + 1/2) \Delta t$, where $i = 1, 2, 3, \dots$ is an integer label, $y(t)$ is the calendar year, and $m(t)$ is the month of the year ($m(t) = 0$ for January). X_i is the average price in that month.

Panel (1) of Fig. 1 shows $X(t)$ of the S&P 500 index since January of 1802. The linear trend is obvious, but slightly concave. There are ups and downs. A few of them are quite large. For instance, one in 1860s, one in 1930s, then in 1960–1970s, and more recently in 2000s.

2.1 Equity Risk Premium

The economists often prefer to examine economic quantities in “real” terms, that is, subtracting the effect of inflation. Panel (2) of Fig. 1 shows the more common view in the literatures: the real logarithmic total return index $X_{\text{real}}(t)$ (This reproduces

Figs. 5–4 in [20]). The most notable feature is that $X_{\text{real}}(t)$ can be linearly regressed over the 200-year history, with an impressive $R^2 = 0.994$:

$$X_{\text{real}}(t) \sim \beta_{\text{rep}} t + \alpha_{\text{rep}} + \epsilon(t). \quad (2)$$

The slope β_{rep} is about 6.6% per year (between 1802 and 2021). This is called the **real equity risk premium**. This constant is one of the most celebrated constants in modern financial systems.

However, we must note that no other major equity index exhibits such beautiful linearity over such long history. Geopolitical events, financial bubbles and crashes often caused significant distortion or even disruption to many national indices. Some people may even criticize that the linearity of $X_{\text{real}}(t)$ for SPX carries with it a strong survivorship bias. There is no certainty that it will continue to work, although it has been working quite well for two centuries.

We also note that, on the back of such impressive R^2 is the residuals $\epsilon(t)$ where

$$\epsilon(t) = X(t) - (\beta_{\text{rep}} t + \alpha_{\text{rep}}) - \log \text{CPI}(t). \quad (3)$$

The residuals $\epsilon(t)$ is illustrated in Panel (3) of Fig. 1. Its standard error is $\sigma = \text{Stdev}(\epsilon(t)) = 0.32$ between 1802 and 2021. Thus its 2σ is ± 0.63 , drawn in two red dashed lines. Assume $\epsilon(t)$ is mean-reverting, this implies that $X_{\text{real}}(t)$ will swing around its linear progress $\beta_{\text{rep}} t + \alpha_{\text{rep}}$ between $\pm 2\sigma$ (in 95% confidence) from decade to decade. This large amount of variation is disguised in the semi-log plot of Panel (2).

This work is primarily the study of such “fine structure”. A 0.5 downward move in the log scale translates to approximately 50% market drop in a large recession. This can cause massive blowup for funds and companies that have too much leverage. When the lack of growth is stretched over a decade, it puts a lot of pressure on pensions, endowments, and retirement accounts that have significant cash outflow.

Note that the $\pm 2\sigma$ swings in $\epsilon(t)$ typically span several decades. Each cycle is composed of several recessions, which typically occurred every 4–10 years. Recession forecast is a “shorter-term” activity than what is studied here.

We created a more adaptive algorithm than a global linear regression in (2). It is used to build a forecast framework for $X(t)$ a few years into the future.

2.2 Discussion—A Naive 10-Year Forecast

In Panel (3), we observe several empirical rules from which we can make a naive 10-year forecast. First, $\epsilon(t)$ oscillates between -2σ and $+2\sigma$. At the dot-com peak of 2000, it touched $+2\sigma$. And at the bottom of 2009 financial crisis, it touched -2σ . Amid the pandemic of 2020, $\epsilon(t)$ was approximately at zero.

Assume $\epsilon(t)$ will reach $+2\sigma$ in 2030, the annual rate of change of $\epsilon(t)$ is $\sigma/5$ in 10 years. The annual real return of $X(t)$ will be $\sigma/5 + \beta_{\text{rep}}$. If the annual inflation

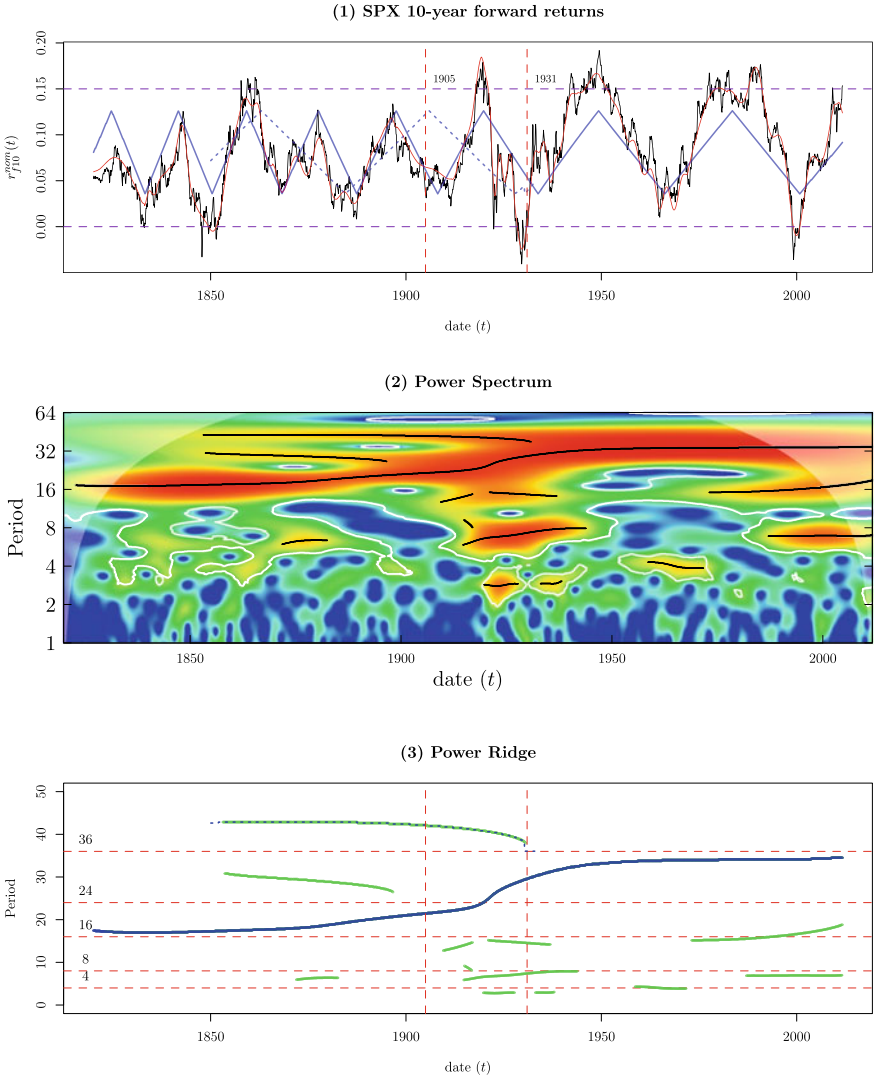


Fig. 2 Wavelet analysis on 10-year forward returns of S&P 500 index, $r_{f10}^{nom}(t)$. The dominant period is 36 years after WWII. The transition period 1905–1931 is marked by the red vertical dash lines, before which the period is shorter, between 16 and 24 years. The 8-year period was very strong for some intervals, e.g. during the Great Depression years, and between 1980 and 2020

is about 3% for the next 10 years, we arrive at the 10-year forward nominal return $r_{f10}^{nom}(t)$ of 16%.

This is a pretty naive estimate. Nevertheless, we will show in Panel (1) of Figs. 2 and 3 that 16% is a reasonable average estimate for a long-term bull market. One

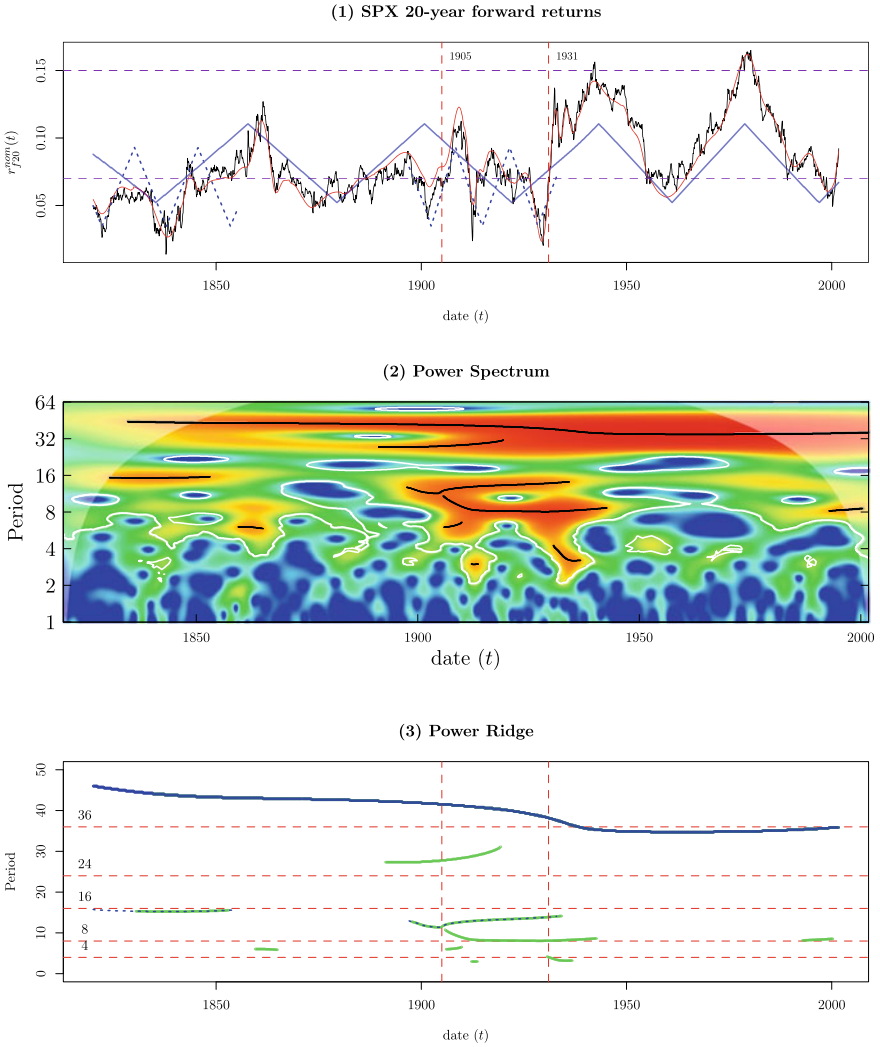


Fig. 3 Wavelet analysis on 20-year forward returns of S&P 500 index, $r_{f20}^{nom}(t)$. The dominant period is 36 years after the transition year 1931, before which there was no clear dominant period

must remember that $X(t)$ has the average annual volatility of about 12–13% between 1950 and 2021. In a good year, the return can reach 30%, but in a bad year, the volatility can be as high as 60%.

3 Wavelet Analysis on the 36-Year Long Term Cycle

In this section, we use the Morlet wavelet [13, 14] to show the 36-year long-term cycle in the U.S. stock market after WWII. This pattern can be observed in both the 10-year and 20-year returns with very little model assumptions. Recognition of such long-term cycle can greatly demystify the behavior of the stock market, e.g. the bull markets in the 1950s, and 1980–90, and the bear markets during 1970s and 2000s.

The **WaveletComp** package in R is used to perform the wavelet analysis. The advantage of this package is its simple user interfaces and beautiful graphical outputs. We briefly explain the main features of the wavelet theory, according to [18].

3.1 Introduction to the Wavelet Transform

The “mother” Morlet wavelet is defined as

$$\psi(t) \equiv \pi^{-1/4} e^{iwt} e^{-t^2/2}, \quad (4)$$

where the “angular frequency” w is set to 6. This is the preferred value in the literatures since it is approximately 2π . This wavelet can be thought of as the composite of a Fourier component e^{iwt} and a Gaussian component $e^{-t^2/2}$. The Fourier component captures the phase of a wave.

The wavelet transform of a time series x_t is defined as its convolution with a set of “wavelet daughters” $\psi\left(\frac{t-\tau}{s}\right)$. The daughters are generated from the mother wavelet by translation in time by τ and scaling by s . Each convoluted wave is

$$\text{Wave}(\tau, s) \equiv \sum_t x_t \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-\tau}{s}\right), \quad (5)$$

where $*$ denotes the complex conjugate. Since x_t in our case is monthly data, τ is shifted in the unit of $dt = 1/12$ (year).

For scaling, the choice of the set of s determines the coverage in the frequency domain, called “periods” $\{s_j\}$. It is a fractional power of 2, a “voice” in an “octave” with $1/dj$ determining the number of voices per octave:

$$s_j = s_{\min} 2^{j \cdot dj}, \quad j = 0 \dots J, \quad (6)$$

where s_{\min} is set to 1 (year), and dj is set to $1/128$. The maximum of s_j is set to 64 (year), which determines $J = 768$. These settings allow us to analyze periods from 1 year to 64 years, that covers our target period of interest: 36 years.

The power spectrum is defined as [4]

$$\text{Power}(\tau, s) \equiv \frac{1}{s} |\text{Wave}(\tau, s)|^2. \quad (7)$$

The power ridges are the s locations of local maximums in $\text{Power}(\tau, s)$ at a given τ [3]. The **WaveletComp** package has a built-in utility to identify statistically significant power ridges in the entire spectrum. For our purpose, the most interesting power ridge is the ridge of global maximum: $\{s_{\max}(\tau) = \text{argmax}_s \text{Power}(\tau, s)\}$.

The instantaneous or local wavelet phase characterizes the periodic phenomena:

$$\text{Phase}(\tau, s) \equiv \text{Arg}(\text{Wave}(\tau, s)), \quad (8)$$

We can follow the phase of global maximum power ridge $s_{\max}(\tau)$ over τ (assume it meets certain continuity condition) to understand the long-term periodicity of the market:

$$\text{Phase}_{\max}(\tau) \equiv \text{Arg}(\text{Wave}(\tau, s_{\max}(\tau))), \quad (9)$$

By transforming the phase via the triangle wave function $f(\theta) = 1 - \frac{2}{\pi} \arccos(\cos(\theta))$, where $\theta = \text{Phase}_{\max}(\tau)$, the periodicity of interest can be clearly illustrated.

The time series can be smoothed and reconstructed by summing over a set of waves:

$$(x_t) = \frac{dj \cdot \sqrt{dt}}{0.776 \cdot \psi(0)} \sum_s \frac{1}{\sqrt{s}} \text{Re}(\text{Wave}(\tau, s)). \quad (10)$$

The reconstruction factor 0.776 is adopted from [25] as an empirically suggested constant for the full reconstruction.

Financial time series is known to have high noise-to-signal ratio. Proper shrinkage during reconstruction (smoothing and/or denoising) can enhance the signal of interest. The wavelet shrinkage is performed by either filtering out s smaller than a certain threshold, or dropping weaker waves according to the strength of the power spectrum.

3.2 Wavelet Regression of the 10-Year Returns

The 10-year forward returns $r_{f10}^{\text{nom}}(t)$ is analyzed in this section. We emphasize that the input data is model-free. The only parametrization is the choice of the return window: 10 years. The wavelet analysis is shown in Fig. 2. From the ‘‘Power Ridge’’ chart in Panel (3), we observe that the dominant period was 36 years after WWII.

In both Figs. 2 and 3, the charting conventions are as follows:

Panel (1) shows the time series x_t ($r_{f10}^{\text{nom}}(t)$ and $r_{f20}^{\text{nom}}(t)$) in the black line, and the reconstructed (x_t) in the red line. The triangle phase $f(\theta)$ of the strongest power ridge is drawn in the solid blue line, and the secondary in the dashed blue line. Two

vertical red dashed lines are drawn at 1905 and 1931—two fault line locations from the 20-year forecast model in Sect. 6.2.

Both the 10-year and 20-year returns could not exceed 15–17% for too long. This level marks the rampant bull market. On the other hand, the 10-year returns rarely went below 0%. The 20-year returns also appear to have a floor at 5–7%.

Panel (2) shows the power spectrum Power (τ, s). The y-axis is the period τ . The color spectrum illustrates the power level where red is high and blue is low. The power ridges are drawn in black lines.

Panel (3) show the power ridges with the guided red dashed lines at the ladders of 4, 8, 16, 24, 36 years. The strongest power ridge is drawn in the solid blue line, and the secondary in the dashed blue line. The remaining ridges in the green lines.

There was a fundamental change in the periodicity before WWI and after WWII. We conjecture this might be related to the transition of the world power from Europe to Washington. Prior to WWI, the period is about 16–24 years, much shorter than 36 years.

3.3 Wavelet Regression of the 20-Year Returns

As we see above, the 36-year period is the natural frequency of the long-term mean-reversion cycles. The regression on the 20-year returns requires the least tectonic adjustments. This gives us the strong incentive to explore the 20-year returns here, even though most financial analysis stops at the 10-year returns.

The wavelet analysis on the 20-year forward returns, $r_{f20}^{\text{nom}}(t)$, is shown in Figure 3. We can clearly observe the 36-year period after the transition year 1931 from the “Power Ridge” chart in Panel (3).

In Panel (1), before 1931, the 20-year returns were pretty flat, around 7%. Most of the smaller fluctuations were smoothed out. In Panel (3), during the Great Depression years, the 8-year period was very strong. But before 1905 and after 1931, there was almost no power distributed in any of the secondary periods. This is consistent with our observation that $r_{f20}^{\text{nom}}(t)$ removed most of the short-term fluctuations and preserved the most important long-term signals.

The 36-year period began to emerge after the 1929 crash. It went through two cycles after WWII. As of this writing, the market is at the bottom of this cycle, and is about to revert from a bear market to a bull market.

4 Channel Deviation Framework

In this section, we lay out the channel deviation framework, in which $X(t)$ is decomposed into the smooth channel moving average $\alpha(T)$, the channel return $R(T)$, and the mean-reverting channel deviation $Y(T)$. We show how the optimal look-back duration $\Delta T_b = 45$ is chosen for the S&P 500 index.

4.1 Mean-Reversion Decomposition

For a given time series that is predominantly in a linear trend, such as the total return index $X(t)$ in (1), we assume it is composed of a linear process and a mean-reverting process. The goal of this framework is to decompose $X(t)$ into these two processes while maintaining causality.

Let ΔT_b be the duration of the look-back channel. At time T , we apply linear regression

$$X(t) \sim \alpha(T) + R(T)(t - T), \text{ where } t \in [T - \Delta T_b, T], \quad (11)$$

to obtain $\alpha(T)$, which is called **channel moving average** (CMA), and $R(T)$, which is called **channel return**. Then we derive the **channel deviation** at time T as $Y(T) = X(T) - \alpha(T)$. One can view $Y(T)$ and $\alpha(T)$ as the decomposition of $X(T)$, where $\alpha(T)$ is linear and non-stochastic, and $Y(T)$ is mean-reverting. $R(T)$ is the instantaneous rate of change of $\alpha(T)$.

$Y(T)$ is of paramount importance in this framework. We will show that log-CAPE mean-reverts in similar pattern and scale to $Y(T)$ in Sect. 5. Since $\alpha(T)$, $R(T)$, and thus $Y(T)$ are causal, they can be used for forecasting after time T , as shown in Sect. 6.

4.2 Closed Form Solution

There are closed form solutions for $\alpha(T)$, $R(T)$, and $Y(T)$ in the discrete notation. (11) is the ordinary least squares (OLS) optimization. Let $\langle t_i - T \rangle$ be the mean of $t_i - T$ for $t_i \in [T - \Delta T_b, T]$, and N is the sample size of t_i , we have $\langle t_i - T \rangle = \frac{N+1}{2} \Delta t \underset{N \gg 1}{\approx} -\frac{1}{2} \Delta T_b$, and $\text{var}(t_i) = \frac{1}{12} (N^2 + N) \Delta t^2 \underset{N \gg 1}{\approx} \frac{1}{12} \Delta T_b^2$. Then

$$\begin{aligned} R(T) &= \frac{\text{cov}(X_i, t_i)}{\text{var}(t_i)} = \text{cor}(X_i, t_i) \frac{\text{stdev}(X_i)}{\text{stdev}(t_i)} \underset{N \rightarrow \infty}{\approx} \frac{\sqrt{12}}{\Delta T_b} \text{cor}(X_i, t_i) \text{stdev}(X_i), \\ \alpha(T) &= \langle X_i \rangle - R(T) \langle t_i - T \rangle = \langle X_i \rangle + \sqrt{3} \left(\frac{N+1}{N} \right) \text{cor}(X_i, t_i) \text{stdev}(X_i) \\ &\underset{N \gg 1}{\approx} \langle X_i \rangle + \frac{1}{2} R(T) \Delta T_b. \end{aligned} \quad (12)$$

The main feature in $R(T)$ and $\alpha(T)$ is the covariance between X_i and t_i in the channel. Given the same $\text{Stdev}(X_i)$, $R(T)$ is maximized by the best $\text{Cor}(X_i, t_i)$, which is 1 when X_i is perfectly linear to t_i .

$\alpha(T)$ is the result of the optimal linear predictor. The first term in $\alpha(T)$ is the moving average $\langle X_i \rangle$. The second term introduces the ‘‘correction’’ for the trend,

which is non-zero as long as $\text{Cor}(X_i, t_i) \neq 0$. The sign of the “correction” is given by the sign of $\text{Cor}(X_i, t_i)$.

Equation (12) leads to the closed form of the channel deviation,

$$Y(T) = X(T) - \langle X_i \rangle - \sqrt{3 \left(\frac{N+1}{N} \right) \text{cor}(X_i, t_i) \text{stdev}(X_i)} \quad (13)$$

This equation is out-of-sample, thus is causal. Also note that this framework is scale independent. The outputs don’t vary much with regard to different data sampling frequency.

4.3 Optimal Choice of Look-back Channel at 45 Years

The look-back channel ΔT_b is the only hyperparameter in this framework. It should be chosen such that the outputs are least biased. The wavelet analysis shows that the channel must be longer than 36 years. Based on our empirical experimentation, we know it is between 30 and 50 years. We provide one version of optimization that we use to determine $\Delta T_b = 45$.

For a given $\Delta T_b < 60$, we calculate $Y(T)$ for all T ’s between 01/1862 and 12/2017. We then calculate the skewness and kurtosis of $Y(T)$ for such ΔT_b . We seek the optimal ΔT_b that produces the lowest kurtosis and zero skewness with a tolerance of randomness. The kurtosis and skewness are shown in Fig. 4.

This turns out to be a relatively simple optimization problem to solve. When ΔT_b is small, the kurtosis is very high and the skewness is negative. As ΔT_b increases, the

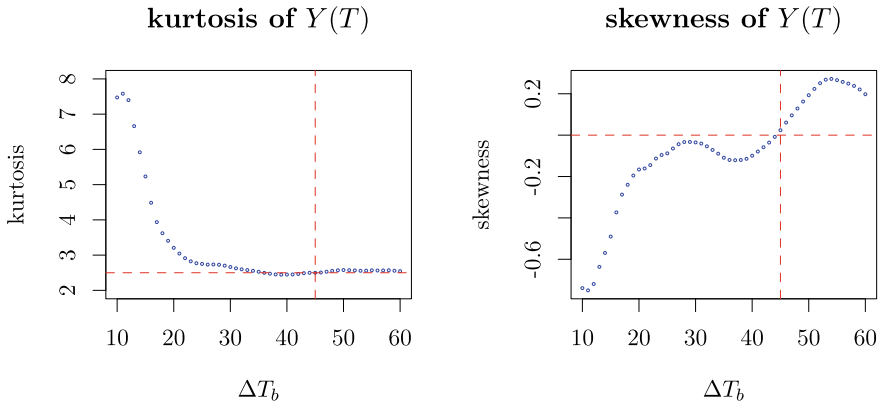


Fig. 4 Optimization of the look-back channel ΔT_b . The left panel shows the kurtosis of $Y(T)$ forms a plateau around 2.5 when $\Delta T_b > 35$. The right panel shows the skewness of $Y(T)$ crosses zero at $\Delta T_b = 45$, which we choose to be the optimal look-back period

kurtosis decreases towards 3 and the skewness increases towards zero. When $\Delta T_b > 21$, the kurtosis decreases below 3, that is, the system transitions from leptokurtotic to platykurtic. When $\Delta T_b > 35$, the kurtosis forms a plateau around 2.5. The kurtosis reaches its minimum of 2.445 at $\Delta T_b = 39$, but the skewness doesn't cross zero until $\Delta T_b = 45$ at which point the kurtosis is at 2.498, slightly higher than the absolute minimum. We determine that $\Delta T_b = 45$ is the optimal choice.

4.4 Discussion on the Outputs

Figure 5 shows the result of $\alpha(T)$, $Y(T)$, and $R(T)$ at $\Delta T_b = 45$. We first note that $Y(T)$ oscillates between ± 0.5 with a periodicity of approximately 40 years. The periodicity is particularly clear by observing the legs of $Y(T)$. The market swings violently during two periods: From 1929 to 1933, the oscillation almost reaches ± 1.0 . From 2000 to 2009, the oscillation is as large as ± 0.75 . We will elaborate more on the periodicity and amplitude of $Y(T)$ in Sect. 5.1.

Secondly, we observe that $R(T)$ has three plateaus in history. The first plateau is at 5% before 1860. The second plateau is at 7.13% from 1880 to 1950. The third plateau is at 10.52% from 1970 to now. The values of plateau are determined by zeroth order `genlasso::trendfilter` utility in R.⁶ At 600 degrees of freedom, we round the output of beta to 3 digits, and select the largest clusters of beta that have repeated more than 50 months. The average of beta from each cluster is the mean of the plateau. The 10.52% return of the third plateau is often quoted in the literatures and media as the long-term expected return of SPX. Here we provide a proper context in terms of $R(T)$.

5 Relation Between Channel Deviation and CAPE

In this section, we show that $Y(t)$, $R(t)$, and CPI have large explanatory power on CAPE, even though their data generating processes (See Chap. 1 of [7]) don't seem to be related at all. The log-CAPE can be decomposed by a four-factor model with a high R^2 .

5.1 Regression of Log-CAPE

Let $\text{CAPE}_{\Delta T}(t)$ denote the ΔT -year CAPE where $\Delta T = 10, 20$. In Panel (1) of Fig. 6, it is shown that $\log(\text{CAPE}_{10}(t))$ and $\log(\text{CAPE}_{20}(t))$ are very similar (the

⁶ See also <https://cran.r-project.org/web/packages/genlasso/vignettes/article.pdf>.

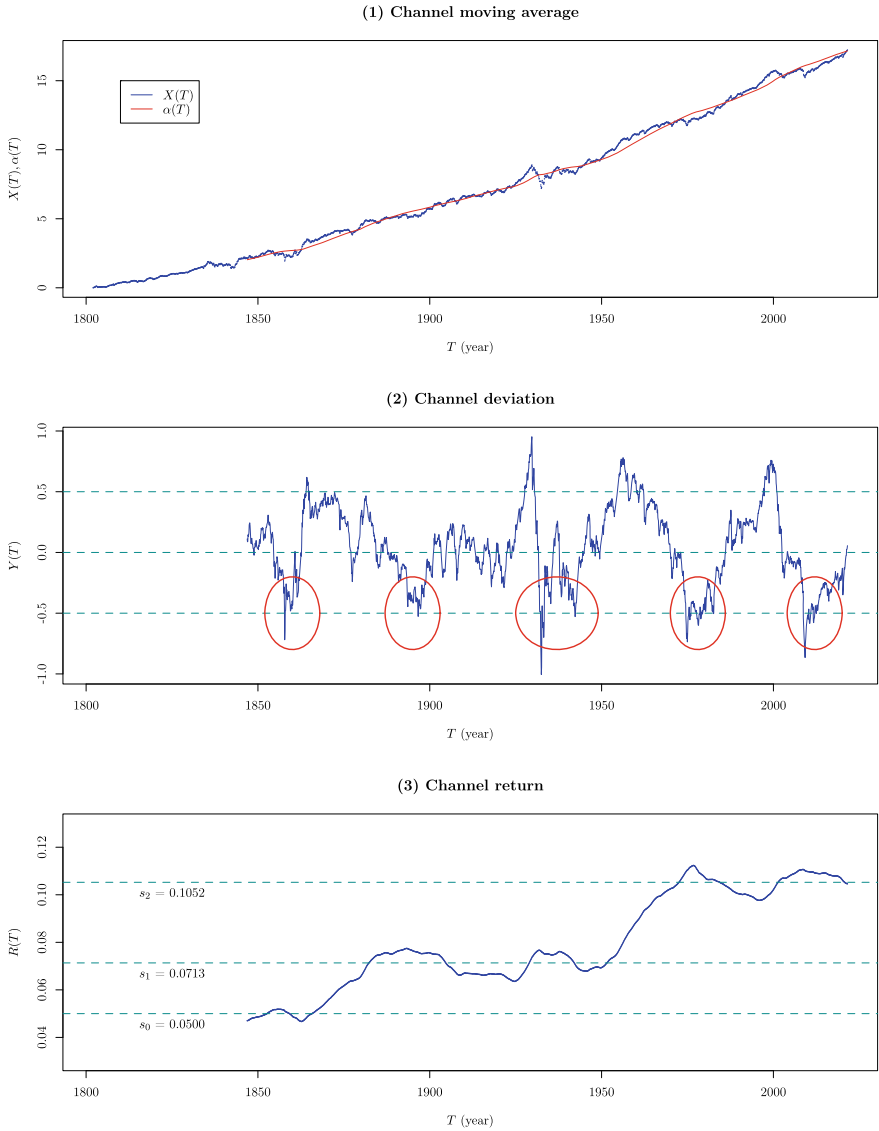


Fig. 5 Optimally decomposed $\alpha(T)$, $Y(T)$, and $R(T)$ at $\Delta T_b = 45$. The legs of $Y(T)$ are drawn in red circles in Panel (2) to illustrate the periodicity. The levels of plateau in $R(T)$, $s_0 = 0.05$, $s_1 = 0.0713$, $s_2 = 0.1052$, are calculated from zeroth order `genlasso::trendfilter` utility. The 10.52% of s_2 is often quoted as the long-term expected return of SPX

blue and cyan lines). Also note that $Y(t)$ is in the same scale of $\log(\text{CAPE}_{\Delta T}(t))$. Hence, we focus on the 20-year model.

And let $\text{CPI}_{10}(t)$ and $\text{CPI}_{20}(t)$ denote the 10 and 20-year log-returns of CPI. That is,

$$\text{CPI}_{\Delta T}(t) = \frac{\log \text{CPI}(t) - \log \text{CPI}(t - \Delta T)}{\Delta T}. \quad (14)$$

In Panel (2), $\text{CPI}_{10}(t)$, $\text{CPI}_{20}(t)$ and $R(t)$ are shown. $\text{CPI}_{10}(t)$ is more volatile than $\text{CPI}_{20}(t)$. $R(t)$ is the long-term moving average of nominal equity returns. It is shifted down by the equity risk premium β_{rep} (6.6%), and we observe it is approximately the long-term (40 years) inflation rate. These three factors constitute the inflation inputs for the regression model.

We perform the following linear regression for t between 1/1881 and 12/2020:

$$\log(\text{CAPE}_{20}(t)) \sim \beta_0 + \beta_1 Y(t) + \beta_2 R(t) + \beta_3 \text{CPI}_{10}(t) + \beta_4 \text{CPI}_{20}(t) + \varepsilon, \quad (15)$$

which results in a high R^2 of 0.82. The summary of linear model from R is shown below:

```
1 a <- lm(log.cape20 ~ eqty.lm.y + eqty.lm.r + cpi.logr.10 +
2 cpi.logr.20, data=df) summary(a)
```

```
1
2 Call:
3 lm(formula = log.cape20 ~ eqty.lm.y+eqty.lm.r+cpi.logr.10+
4     cpi.logr.20, data = df)
5
6 Residuals:
7     Min       1Q   Median       3Q      Max
8 -0.34672 -0.16100 -0.02993  0.18624  0.45322
9
10 Coefficients:
11             Estimate Std. Error t value Pr(>|t|)
12 (Intercept)  1.00789    0.02938   34.30  <2e-16 ***
13 eqty.lm.y    0.93990    0.01713   54.87  <2e-16 ***
14 eqty.lm.r   25.16626    0.36828   68.33  <2e-16 ***
15 cpi.logr.10 -3.84196    0.28691  -13.39  <2e-16 ***
16 cpi.logr.20 -11.73114    0.42353  -27.70  <2e-16 ***
17 ---
18 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
19
20 Residual standard error: 0.1923 on 1555 degrees of freedom
21 (1067 observations deleted due to missingness)
22 Multiple R2: 0.82, Adjusted R2: 0.8196
23 F-statistic: 1771 on 4 and 1555 DF, p-value: < 2.2e-16
```

All four factors are highly significant. More than three quarters of information in log-CAPE is contained in the linear combination of our mean reversion analytics and past inflations.

The result is shown in Panel (3) of Fig. 6.

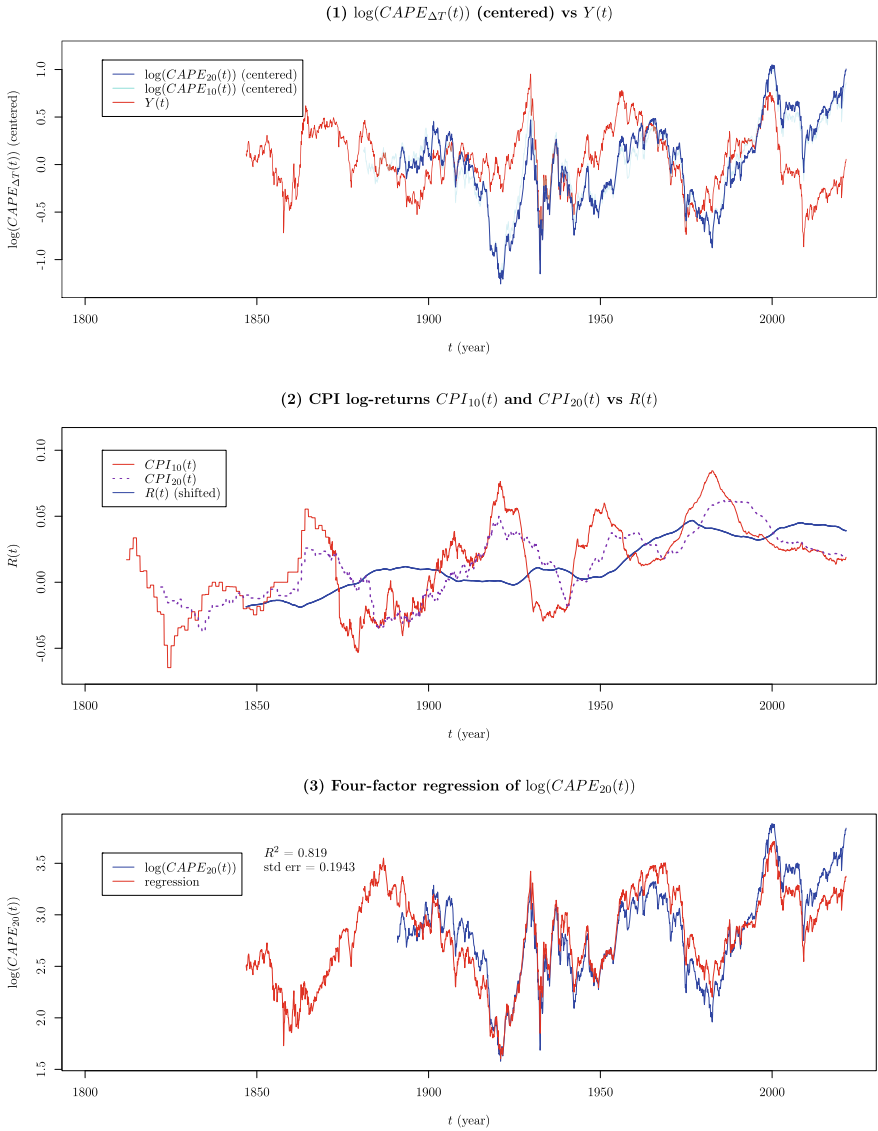


Fig. 6 Linear regression of the 20-year log-CAPE by the four factors: $Y(t)$, $R(t)$, $\text{CPI}_{10}(t)$ and $\text{CPI}_{20}(t)$. Panel (1): Comparison of centered log-CAPE and $Y(t)$, showing their similarity and in the same scale. Panel (2): $R(t)$, $\text{CPI}_{10}(t)$ and $\text{CPI}_{20}(t)$ as the inflation inputs to supplement the differences between log-CAPE and $Y(t)$. Here $R(t)$ is shifted down by the real equity premium. Panel (3): The result of regression on $\log(\text{CAPE}_{20}(t))$ in the four-factor model

5.2 Discussion

We illustrated the inner workings of the four-factor regression by Panel (1) and Panel (2) of Fig. 6. Panel (1) shows that $\log(\text{CAPE}_{20}(t))$ is almost in the same scale as $Y(t)$, and this is confirmed by the coefficient $\beta_1 = 0.94$ in Eq. (15).

There are times that log-CAPE moves below $Y(t)$ (e.g. in 1920s, 1950s, and early 1980s) and other times that log-CAPE moves above $Y(t)$ (e.g. in 1900s and 2000s). Their differences are made up by $\text{CPI}_{10}(t)$ and $\text{CPI}_{20}(t)$. This is confirmed by the negative correlation ($\beta_3 = -3.8$ and $\beta_4 = -11$) in the summary statistics above. This is shown graphically in Panel (2). We observe that, whenever $\text{CPI}_{10}(t)$ and $\text{CPI}_{20}(t)$ are above $R(t)$, $Y(t)$ tends to be above $\log(\text{CAPE}_{20}(t))$, and vis versa.

This anti-correlation between log-CAPE and inflation is one of the two main reasons why CAPE is perceived at a lofty level since 2000. The high CAPE reading is a reflection of ultra-low inflation in the past two decades.

The second reason is that log-CAPE is positively correlated to $R(t)$ with $\beta_2 \approx 25$. Since $R(t)$ is currently at the third plateau, it also contributes to the high level of CAPE. From 1950 to 1970, the market was transitioning from the second plateau to the third, the difference in $R(t)$ is $s_3 - s_2 \approx 3.4\%$. Multiplying it by $\beta_2 \approx 25$, its impact on log-CAPE is 0.85, which is translated to 130% higher CAPE. In 1970s and 1980s, this effect was muted because of the high inflation. Going into 1990s, the high tide of inflation receded and CAPE began to move much higher.

However, in order to justify such high level of equity returns and valuation, it seems to imply that the future inflation will have to be much higher.

5.3 Tectonic CAPE

We introduce the concept of tectonic CAPE, in which we hypothesize wars and national policy changes in the past might have resulted in significant dislocations in the data generating processes of CAPE and CPI (Chap. 19 of [7]). We use nonlinear optimization technique to uncover these dislocations. However, we do this only sparingly so that we don't overfit the data.

At a specific time t_i^{adj} , the amount $\Delta_i \log \text{CAPE}_{20}$ should be added to $\log(\text{CAPE}_{20}(t))$. These adjustments are called the ‘‘fault lines’’, and the adjusted CAPE is called ‘‘tectonic CAPE’’.

Formally, the tectonically adjusted log-CAPE is

$$\log(\text{CAPE}_{\Delta T}^{\text{adj}}(t)) = \log(\text{CAPE}_{\Delta T}(t)) + \sum_{i=1 \dots N} \begin{cases} 0, & t < t_i^{\text{adj}}; \\ \Delta_i \log \text{CAPE}_{\Delta T}, & t \geq t_i^{\text{adj}}. \end{cases} \quad (16)$$

Lihn [10] showed that the 20-year model requires smaller amount of ‘‘fault line adjustments’’ than the 10-year model. The interpretation is that many economic shocks tend to average out much better in 20 years than 10 years.

This fits our understanding of the 36-year cycle from the wavelet analysis. In Figs. 2 and 3, the transition period 1905–1931 are marked from the two fault lines required to adjust the 20-year model. It will be discussed in Sect. 6.2 in more detail.

6 The Tectonic Forecast Model

We first introduce the general form of the factor-based tectonic forecast model. Then we specialize it to equity market forecast in the following sections.

Let f_k be the factors under consideration: $Y(t)$, $R(t)$, $\text{CPI}_{10}(t)$, $\text{CPI}_{20}(t)$, $\log(\text{CAPE}_{\Delta T}(t))$. And $r_{f,\Delta T}(t)$ is the forward log-return of period ΔT that we intend to forecast, in nominal term or real term. In this section, we study the nominal forecast, and ΔT is set at 20 years.

The simplest mathematical form of the forecast regression is to assume the faults $\{\Delta_i\}$ are generic, such that

$$\Delta = \sum_{i=1 \dots N} \begin{cases} 0, & t < t_i^{\text{adj}}, \\ \Delta_i, & t \geq t_i^{\text{adj}}. \end{cases} \quad (17)$$

Plugging Δ to a linear regression, we have

$$r_{f,\Delta T}(t) \sim \beta_0 + \sum_{k=1 \dots 5} \beta_k f_k(t) + \Delta + \varepsilon, \text{ for } t \in [T_{\text{start}}, T_{\text{end}}], \quad (18)$$

where ε is the standard error of the forecast. For a given N , the objective is to minimize the AIC of the regression by a nonlinear optimization in the parametrization space: $\{\beta_k, \forall k = 0, 1 \dots 5\} \cup \{(t_i^{\text{adj}}, \Delta_i), \forall i = 1 \dots N\}$.

However, the faults $\{\Delta_i\}$ in this form is in the unit of the response $r_{f,\Delta T}(t)$. This is saying that the response is dislocated and needs to be corrected. Our preference is to attribute the dislocations to one of the major factors, for instance, to log-CAPE, or $Y(t)$.

In the case of log-CAPE, we prefer to rewrite the regression as

$$r_{f,\Delta T}(t) \sim \beta_0 + \sum_{k=1 \dots 4} \beta_k f_k(t) + \beta_5 \log(\text{CAPE}_{\Delta T}^{\text{adj}}(t)) + \varepsilon. \quad (19)$$

In this form, the faults $\{\Delta_i \log \text{CAPE}_{\Delta T}\}$ are in the unit of log-CAPE in (16). We think that this provides better economic cause-and-effect interpretation. But we acknowledge that the generic representation Eq. (17) is more elegant mathematically, and may conveys an economic significance that is yet to be uncovered.

6.1 The 5-Factor Tectonic Forecast Model for Equity Returns

Let $r_{f,\Delta T}^{\text{nom}}(t)$ denote the nominal forward log-return for the look-forward period ΔT . We construct the 5-factor forecast model as follows:

$$r_{f,\Delta T}^{\text{nom}}(t) \sim \beta_0 + \beta_1 Y(t) + \beta_2 R(t) + \beta_3 \text{CPI}_{10}(t) + \beta_4 \text{CPI}_{20}(t) + \beta_5 \log(\text{CAPE}_{\Delta T}^{\text{adj}}(t)) + \varepsilon_t; \quad (20)$$

where $t \in [T_{\text{start}}, T_{\text{end}}]$ is the in-sample period. The objective is to minimize the AIC of the regression by a nonlinear optimization on the fault lines $\left\{ \left(t_i^{\text{adj}}, \Delta_i \log \text{CAPE}_{\Delta T} \right), \forall i = 1 \cdots N \right\}$. We view our 5-factor model as a meaningful extension of the one-factor CAPE model ([2, 21]).

Although the regression can be performed on both real and nominal returns, we find this model works better for nominal return, because the inflation is accounted for separately from the nominal return.

The nominal return forecast is also advantageous in that it can be translated into index level prediction without the complication of inflation forecast. The predicted forward returns can be easily translated to the predictions of log-index $X_{\text{pred}}(t)$ and future SPX level (dividend included) $p_{\text{pred}}(t)$:

$$\begin{aligned} X_{\text{pred}}(t + \Delta T) &= X(t) + \left[r_{f,\Delta T}^{\text{nom}}(t) \right]_{\text{pred}} \Delta T \pm 2\varepsilon \Delta T, \text{ where } t \in [T_{\text{end}} - \Delta T, T_{\text{end}}]; \\ \text{and } p_{\text{pred}}(t) &= e^{X_{\text{pred}}(t)} e^{\pm 2\varepsilon \Delta T}. \end{aligned} \quad (21)$$

$\varepsilon = \text{stdev}(\varepsilon_t)$ represents the forecast error. The $\pm 2\varepsilon \Delta T$ term indicates the 2-stdev error bounds. Assume $\varepsilon \approx 0.01$ for the 20-year forecast, then $e^{40\varepsilon} \approx 1.49$ and $e^{-40\varepsilon} \approx 0.67$. That is, the price forecast can have 33% to 49% of error. This is consistent with our observation that this forecast model is good for long-term trend line prediction, but not good for predicting short-term ‘‘bubbles and crashes’’ in the 3–5 year timeframe. The reader must be careful in interpreting the forecast in a proper context. The shorter term forecast has to come from other types of models and indicators, e.g. the recession forecasts.

6.2 20-Year Nominal Return Forecast

Since the market has shown the 36-year periodicity, it is natural to expect that the 20-year forecast, $\Delta T = 20$, works better. Only two fault lines are used before WWII:

i	t_i^{adj}	$\Delta_i \log \text{CAPE}_{20}$
1	1904.96	-0.984
2	1930.73	-1.499

We are impressed that there is no adjustment after 1931. However, we are not able to interpret what specific historical events are associated with the fault lines. The 1930 fault line is likely related to the turbulent 1930s. It seems that twenty years are long enough to smooth out many impacts in history. Only the largest dislocations remain.

The $r_{f20}^{\text{nom}}(t)$ regression achieves very high R^2 of 0.86. The summary statistics is shown below:

```
1 c20 <- lm(eqty.logr.f20 ~ eqty.lm.y + eqty.lm.r + cpi.logr.10 +
2 cpi.logr.20 + log.cape20.adj, data=df) summary(c20)
```

```
1
2 Call:
3 lm(formula = eqty.logr.f20 ~ eqty.lm.y + eqty.lm.r +
4 cpi.logr.10 +
5     cpi.logr.20 + log.cape20.adj, data = df)
6
7 Residuals:
8     Min       1Q   Median       3Q      Max
9 -0.041961 -0.007638 -0.000008  0.006443  0.046165
10
11 Coefficients:
12             Estimate Std. Error t value Pr(>|t|)
13 (Intercept)  0.0944513  0.0020000  47.23  <2e-16 ***
14 eqty.lm.y    -0.0241555  0.0010935  -22.09  <2e-16 ***
15 eqty.lm.r    0.4109853  0.0243705  16.86  <2e-16 ***
16 cpi.logr.10  0.3739050  0.0166545   22.45  <2e-16 ***
17 cpi.logr.20 -0.9270231  0.0268912  -34.47  <2e-16 ***
18 log.cape20.adj -0.0246708  0.0003967  -62.19  <2e-16 ***
19 ---
20 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
21
22 Residual standard error: 0.0111 on 1323 degrees of freedom
23 (315 observations deleted due to missingness)
24 Multiple R2: 0.8591, Adjusted R2: 0.8585
25 F-statistic: 1613 on 5 and 1323 DF, p-value: < 2.2e-16
```

Heuristically, we can express the 20-year nominal return as follows:

$$r_{f20}^{\text{nom}}(t) \approx \text{const} - \frac{1}{20} \left(\frac{1}{2} Y(t) + \frac{1}{2} \log(\text{CAPE}_{20}^{\text{adj}}(t)) \right) + \frac{2}{5} R(t) + \frac{1}{3} \text{CPI}_{10}(t) - \text{CPI}_{20}(t). \tag{22}$$

Log-CAPE and $Y(t)$ contribute to the future returns almost equally. The higher they are now, the less the returns in the future. $\text{CPI}_{20}(t)$ is also anti-correlated with a coefficient of nearly -1.0 . On the other hand, $R(t)$ and $\text{CPI}_{10}(t)$ are like the momentum factors for the future returns.

The 20-year result is shown in Fig. 7. Panel (1) shows the 5-factor regression without fault lines (red line). Note that R^2 of 0.51 is too low for prediction purpose.

Panel (2) shows the 5-factor regression with tectonic CAPE (red line), with R^2 of 0.86 and $\varepsilon \approx 0.01$. Both are high enough for a forecast model.



Fig. 7 Forecasting 20-year nominal equity returns with the 5-factor tectonic CAPE model. Only two fault lines before WWII are needed to correct the dislocations. Panel (1) shows the regression without fault lines, which produces much inferior R^2 and ε . Panel (2) shows the regression with fault lines. The R^2 of 0.86 and $\varepsilon \approx 0.01$ are satisfactory for a forecast model. Panel (3) shows the result of $X_{\text{pred}}(t)$. Panel (4) shows the result of $p_{\text{pred}}(t)$

In Panels (3) and (4), the regressed $r_{f20}^{\text{nom}}(t)$ is converted to $X_{\text{pred}}(t)$ and $p_{\text{pred}}(t)$ via (21). It predicts a bright future for the next 10 years (red line). The current level of $X(t)$ is near the forecast level. Two green dashed lines indicate the end of Biden's term at 2024 and the next presidential term at 2028. The index price $p_{\text{pred}}(t)$ is forecasted above 5000 at 2024 and above 8000 at 2028.

The positive outlook coincides with the 36-year wavelet cycle that the market is on the verge of entering another great bull market after 2020. However, we note that even a small error in the 20-year return forecast can cause a large difference in the price level forecast.

The purple dashed lines show the $\pm 2\varepsilon$ bounds. Both the dot-com peak and 2009 bottom reached the dashed lines. These are the market extremes of exuberance and panic, that the long-term investors must be aware of.

References

1. Campbell John Y, Shiller Robert J (1988) Stock prices, earnings, and expected dividends. *J Financ* 43(3):661–676
2. Campbell John Y, Shiller Robert J (1998) Valuation ratios and the long-run stock market outlook. *J Portf Manag* 24(2):11–26
3. Carmona R, Hwang W-L, Torresani B (1997) Characterization of signals by the ridges of their wavelet transforms. *IEEE Trans Signal Process* 45(10):2586–2590
4. Carmona R, Hwang W-L, Torresani B (1998) Practical time frequency analysis. Gabor and wavelet transforms with an implementation in S. Academic Press, San Diego
5. Cox JC, Ingersoll JE, Ross SA (1985) A theory of the term structure of interest rates. *Econometrica* 53:385–407. <https://doi.org/10.2307/1911242>
6. Davis J, Aliaga-Díaz R, Thomas CJ (2012) Forecasting stock returns: what signals matter, and what do they say now. Vanguard Research
7. Elliott G, Timmermann A (2016) Economic forecasting. Princeton University Press. ISBN: 978-0-691-14013-1
8. Fan J, Yao Q (2017) The elements of financial econometrics. Cambridge University Press. ISBN: 978-1-107-19117-4
9. FRED (2021) Economic data at Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/>
10. Lihn SH-T (2018) Jubilee tectonic model: forecasting long-term growth and mean reversion in the U.S. stock market. SSRN: 3156574. <http://dx.doi.org/10.2139/ssrn.3156574>
11. Lihn SH-T (2018) Jubilee: forecasting long-term growth of S&P 500 index. SSRN: 3253325. <http://dx.doi.org/10.2139/ssrn.3253325> The R package is available at <https://CRAN.R-project.org/package=jubilee>
12. Minneapolis FED (2021) Consumer price index since 1800. <https://www.minneapolisfed.org/community/financial-and-economic-education/cpi-calculator-information/consumer-price-index-1800>
13. Morlet J, Arens G, Fargeau E, Giard D (1982) Wave propagation and sampling theory—Part I: complex signal and scattering in multilayered media. *Geophysics* 47:203–221
14. Morlet J, Arens G, Fargeau E, Giard D (1982) Wave propagation and sampling theory—Part II: sampling theory and complex waves. *Geophysics* 47:222–236
15. R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
16. Mulvey JM, Hao H, Li N (2017) Machine learning, economic regimes and portfolio optimization. Princeton ORFE Working Paper

17. Roesch A, Schmidbauer H (2018) WaveletComp: computational wavelet analysis. R package version 1.1. <https://CRAN.R-project.org/package=WaveletComp>
18. Roesch A, Schmidbauer H (2018) WaveletComp 1.1: a guided tour through the R package. R package version 1.1. <https://CRAN.R-project.org/package=WaveletComp>
19. Schwert GW (1989) Indexes of United States stock prices from 1802 to 1987. NBER Working Paper No. 2985
20. Siegel JJ (2014) Stocks for the long run. McGraw-Hill. ISBN: 0-07-180051-4
21. Siegel JJ (2016) The Shiller CAPE ratio: a new look. *Financ Anal J* 72(3)
22. Shiller RJ (2014) The mystery of lofty stock market elevations. *New York Times*. <https://www.nytimes.com/2014/08/17/upshot/the-mystery-of-lofty-elevations.html>
23. Shiller RJ (2018) Online data for U.S. stock markets 1871-present and CAPE ratio. <http://www.econ.yale.edu/~shiller/data.htm>
24. Shiller RJ (2018) The world's priciest stock market. Project Syndicate. <https://www.project-syndicate.org/commentary/us-stock-market-highest-cape-ratio-by-robert-j--shiller-2018-01>
25. Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bull Am Meteorol Soc* 79(1):61–78

Stephen H.-T. Lihn Atom Investors LP., USA

Dr. Stephen Lihn is currently the Head of Quant Research at Atom Investors LP., an asset management firm serving institutional investors since 2018. Atom is a spinoff of Novus Partners, a fintech company running a financial analytics platform, where he worked for 10 years. Mr. Lihn holds a Ph.D. in Physics from the University of Maryland and a B.S. in Physics from the National Taiwan University. He also has a research affiliation with the ORFE department at Princeton University since 2015.

Smart Maintenance and Human Factor Modeling for Aircraft Safety



Eric T. T. Wong and W. Y. Man

Abstract The global pandemic has significantly accelerated the need for remote monitoring and diagnostics of airline operations and assets. As passenger and cargo flights are impacted from all directions, maintenance can be the steady, reliable part of the puzzle that helps get things back on track. This chapter explores the aircraft safety challenges that can be addressed with better maintenance technology and human factor modeling. Aircraft safety relies heavily on maintenance. During the COVID-19 recovery phase, airline operators need to focus on the application of a robust management of change process to implement better maintenance technology, identify new aircraft safety risks, determine effective mitigation measures, and implement strategies for deploying changes accordingly. For years aircraft maintenance routines have been carried out in the same manner without change, now with international travel restrictions, social distancing, reduced staff, and limited maintenance funding, the need for smarter ways of doing maintenance is obvious. In this regard smart technology has an important role to play. For instance, IoT data generates the capacity for predictive aircraft maintenance, AI introduces the capacity for smart, deep-learning machines to make predictive maintenance more accurate, actionable, and automatic. AI-enabled predictive maintenance leverages IoT data to predict and prevent aircraft failures. While smart technology enhances aircraft safety through better maintenance performance on the one hand, there are technical and human factor problems induced by COVID-19 on the other. The Safe Aircraft System (SAS) model, based on the Dirty Dozen and SHELL human factor models, is an initiative proposed to minimize such COVID-19 problems. This work shows through a case illustration that SAS modeling is a useful tool in identifying potential hazards/consequences associated with any major or minor changes in flight operations. Hence the synergistic effect of smart maintenance and the SAS model in enhancing aircraft system safety are demonstrated.

E. T. T. Wong (✉)

Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China

e-mail: eric.wong.me@connect.polyu.hk; mmttwong@connect.polyu.hk

W. Y. Man

Department of Engineering (Avionics), Heliservices (HK) Ltd, Hong Kong SAR, China

e-mail: wai-yeung.man@connect.polyu.hk

Acronyms and Glossary

ACI	Airports Council International
AI	Artificial Intelligence
Airworthy	The status of an aircraft, engine, propeller, or part when it conforms to its approved design and is in a condition for safe operation
AME	Aircraft Maintenance Engineer
AMO	Approved Maintenance Organization
AR	Augmented Reality
Continuing airworthiness	The set of processes by which an aircraft, engine, propeller, or part complies with the applicable airworthiness requirements and remains in a condition for safe operation throughout its operating life
CBM	Condition-Based Maintenance
COVID-19	Coronavirus Disease of 2019
DD	Dirty Dozen Human Factor Model
FAA	Federal Aviation Administration
ICAO	International Civil Aviation Organization
IoT	Internet of things
MRO	Maintenance and Repair Organization
NTSB	National Transportation Safety Board
OEM	Original Equipment Manufacturer
OSHA	Occupational Safety and Health Administration
RAMS	Recovery and Modifications Services of Boeing
SAS	Safe Aircraft System
SHELL	Software, Hardware, Environment, Liveware & Live-ware (a human factor model used to analyze the interaction of multiple system components)
SOP	Standard Operating Procedure
Smart Maintenance	An intelligent maintenance system brings together technology, data, analyses, prognosis, and resources aiming for the aircraft and systems to achieve highest possible performance levels and near-zero breakdown
VR	Virtual Reality

1 Introduction

International Civil Aviation Organization [1] estimated that the COVID-19 impact on scheduled international passenger traffic could reach reductions of up to 71% of seat capacity and up to 1.5 billion passengers globally in 2020. Airlines and airports face a potential loss of revenue of up to 314 billion USD and 100 billion USD,

respectively. Under the COVID-19 pandemic, aircraft safety can be improved not only by considering the reliability of an aircraft and its systems, but also the compliance of additional mandatory maintenance requirements associated with long-term grounding/storage of aircraft. Aircraft safety relies heavily on maintenance. When it is not done correctly, it contributes to a significant proportion of aviation accidents and incidents [2]. Some examples of maintenance errors are parts installed incorrectly, missing parts, and necessary checks not being performed. In comparison to many other threats to aviation safety, the mistakes of an aircraft maintenance engineer (AME) can be more difficult to detect. These mistakes are present but not visible and have the potential to remain latent, affecting the safe operation of aircraft for longer periods of time. AMEs are confronted with a set of human factors unique within aviation. Very often, they are working in the evening or early morning hours, in confined spaces, on platforms that are up high, and in a variety of adverse temperature/humidity conditions. The work can be physically strenuous, yet it also requires attention to detail. Because of the nature of the maintenance tasks, AMEs commonly spend more time preparing for a task than carrying it out. Proper documentation of all maintenance work is a mandatory requirement, and AMEs typically spend as much time updating maintenance logbooks as they do their work. Human factors awareness can lead to improved quality, an environment that ensures continuing worker and aircraft safety, and a more involved and responsible work force [1].

Human error can be a consequence of design flaws, inadequate training, incorrect procedures, use of outdated operation or maintenance manuals among other technical, environmental, and organizational factors. The drawback of the increasing advance in aircraft technology along with job responsibilities, irregular sleep and the stressful shifts can consequently cause fatigue and work overload problems for both pilots and maintenance personnel. On this basis, in recent decades, the accurate comprehension of psychological and physiological characteristics associated with human errors have become of significant interest. With the incidence of the COVID-19 pandemic and the concomitant impact on airline operation due to international travel restrictions and social distancing, workload conditions for AMEs have become very irregular and this seriously affected aircraft system safety.

This work is aimed to advocate the application of smart technology in aircraft maintenance and the use of human factor modeling for the enhancement of aircraft system safety. First, the impact of COVID-19 pandemic on commercial aviation operation is described. Despite various COVID-19 challenges, approved maintenance organizations (AMOs) could make use of the unprecedented chance offered to make a change to traditional aircraft maintenance practices with the aid of smart technology. Recent aviation applications of smart technology and associated contribution to aircraft safety through maintenance are identified. Important components of the Safe Aircraft System (SAS) model for enhancing aircraft safety are described, followed by a case study illustrating the model implementation in respect of a business operation contemplated by more and more airlines: the transport of cargo in passenger aircraft.

2 Impact of COVID-19 Pandemic on Commercial Aviation

The airline industry has been hit hard, and has laid off thousands of pilots, AMEs, flight attendants, and ground support staff. In the case of the world-renowned Cathay Pacific Airways, there were 26,500 staff working for Cathay Pacific and Cathay Dragon globally. To survive COVID-19 pandemic, Cathay Pacific decided to lay off 6000 workers and kill off Dragon affiliate. This action accounts for reducing 24% of its headcount [3]. Many aircraft that are no longer in operation have been placed by the airline or lessor into a manufacturer-approved storage regime to preserve maintenance and asset value.

It was estimated that globally there were over 18000 aircraft sent to storage in July 2020 (see Figure 1) [4].

Gradually, as travel restrictions are being lifted and as operators are preparing to resume passenger flights and demand increases, operators will need the aircraft that have been parked/stored and return them back to service. Due to the huge number of aircraft involved and the limited supporting resources available to perform the maintenance work due to the COVID-19 crisis, airlines are expected to experience difficulties and increased risks. Airline management needs to identify the potential hazards arising from recovery and new business operations, developing control measures to mitigate the associated risks, and thus ensuring aircraft system safety.

Aircraft repairs and maintenance during the pandemic offer a chance of innovation, and everyone concerned can be asked to adopt divergent thinking and develop new ways of performing maintenance activities. Hitherto maintenance routines have been done “the same old way” without a second thought, now with social distancing, fewer staff, and less maintenance resources, the need for smarter ways of doing

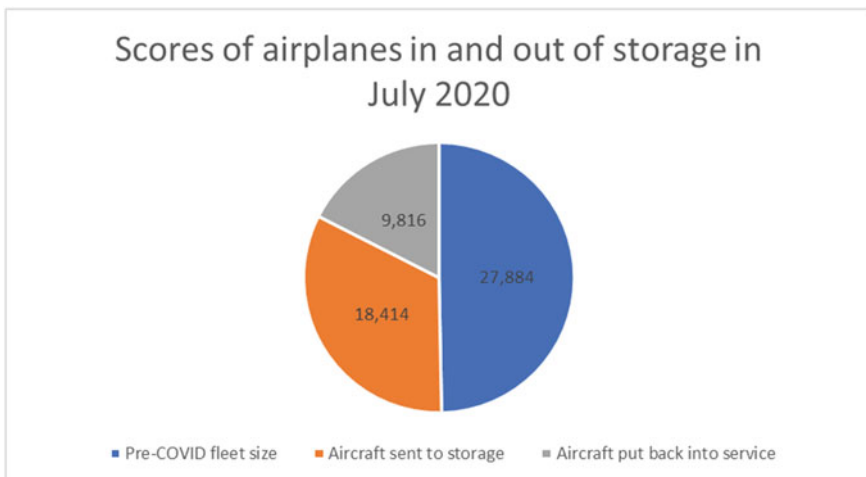


Fig. 1 Scores of airplanes in and out of storage [4]

maintenance is imminent. Getting the job done with the added COVID-19 challenges, maintenance staff have had to examine what really needs to be done, to be more specific on what to inspect, and how to inspect it to ensure the compliance of mandatory airworthiness requirements. To achieve these aims, more and more airlines are employing smart technology in aircraft maintenance.

3 Smart Technology in Aviation

Smart technology is central to the future of the air transport industry. Smart technology has been playing a major role to support airports and airlines preparing to recover from the crisis, which is likely to accelerate the adoption of automation and artificial intelligence in the aviation sector. Automation in the aviation industry is gaining momentum due to rapid advancements in the fields of robotics. Technologies like machine learning and natural language processing are all part of the AI landscape. Each one is evolving along its own path and when applied in combination with data analytics and automation can help airlines achieve their goals of ensuring aircraft safety. With the implementation of smart technology, airports can offer much better passenger services and AMOs can predict potential maintenance failures on aircraft before any mishaps occur. The following section provides an account of smart technology applications and their contributions to aircraft safety.

A. *Smart technology applications in aviation*

1. Airport terminal robotics, automated vehicles, and drone delivery:

- Self-driving guide robots
Robots in airport terminals are becoming a more common sight and among some of the recent examples are Fraport's new self-driving guide robot, called YAPE,¹ for luggage transportation; "Airstar" robot at Incheon Airport; Munich Airport's Josie Pepper; and British Airway has plans to test AI-powered autonomous robots at Heathrow Terminal T5 to further enhance punctuality for passengers. The robots are capable of interacting in different languages with the help of translation technology. They can provide answers to a variety of different questions, including real-time flight information.
- Automated vehicles on the airfield and baggage-related robots
A prominent example is Vanderlande's end-to-end baggage logistics solution FLEET,² deployed at Rotterdam Hague Airport, and tested at Hong Kong International Airport to further improve the efficiency of the baggage handling process, enhance ergonomic working conditions for ground staff and future-proof the airports' baggage handling operations.
- Drone delivery
Edmonton International Airport (EIA) entered a new strategic partnership with Drone Delivery Canada (DDC) that will see the airport

become a hub for drone cargo deliveries in Western and Northern Canada. This is expected to be the world's first regularly scheduled drone delivery service from an airport.

2. Airport AI and Machine Learning

During the past couple of years, the air transport industry has been showing a great commitment to realizing the full potential of artificial intelligence (AI) with a plethora of use cases. On one side, we have seen airlines and airports adopting chatbots to communicate with passengers, and on the other to improve operations. In terms of chatbot applications, last year AirAsia developed and launched its AirAsia Virtual Allstar (AVA),³ a continuously learning AI-powered chat platform.

DeltaAir Lines claimed that it is applying AI-driven machine learning on a scale that's never been done before by an airline. The proprietary AI-driven platform analyses millions of operational data points—from aircraft positions to flight crew restrictions to airport conditions—to create hypothetical outcomes that help Delta's staff make critical decisions before, during and after large-scale disruptions.

KLM Royal Dutch Airlines has developed a suite of advanced optimization tools for the Operations Control Centre to help set up robust schedules by implementing smart tail assignment,⁴ manage and solve disruptions, and help with decision-making. The benefits for passengers are clear—minimizing the impact of disruptions through real-time updates, reducing baggage delays and personalizing information that has been provided to the customer through digital channels.

3. Rapid development of 5G technology

- Recent developments in 5G technology are fuelling the new decade of innovation that will change business as we know it today. The technology lowers data latency offers more stability and connects a huge number of devices at the same time. In the aviation industry, the technology will be instrumental to satisfy the need for fast connectivity inflight and at airports; demand for predictive maintenance through data shared by the connected aircraft; and growing demand for a comfortable inflight experience.

Airports Council International (ACI) World's director of security, facilitation and IT Nina Brooks sees two main benefits of 5G in airports: enhancing existing applications (such as passenger processing, baggage management and airport operations), and enabling new technologies (such as edge computing, internet of things, digital twin creation and artificial intelligence).

"The speed and flow of data, and the sheer potential to use historic, predictive and real-time data, offers solutions that seemed light years away, only a few years ago," says Brooks. "These advances will enable much stronger collaboration between airports, airlines, ground

handlers, air traffic managers, suppliers and retailers, and make their interactions seamless and faster.”

As the global aviation and tourism industry starts up again following the Covid-19 crisis, 5G’s arrival is well-timed and will facilitate applications—including contactless solutions—that will help airports and airlines manage Covid-era regulations.

4 Virtual reality (VR) and augmented reality (AR)

- Airline VR

In a saturated market such as the airline sector, virtual reality (VR) and immersive experiences can be a true differentiator. As one of the leading virtual reality suppliers, Inflight VR’s Virtual Reality Entertainment attracted a number of airlines to its portfolio, including Evelop Airlines, SunExpress, and Jin Air, etc.

Renacen is a company that specializes in the use of virtual reality with its 3D SeatMap VR software, which offers a virtual 360° view of the cabin and can be used for seat upselling, crew training, marketing and VR experiences. The technology has been implemented by a number of airlines, including Emirates, Evelop, Austrian, Aigle Azur and Etihad.

- Augmented Reality (AR) in maintenance: Lockheed Martin and NGRain⁵

With the latest advancements, augmented reality on smart glasses is changing the perspective of aircraft maintenance. Now the AR glasses enable mechanics to see hands-free and with interactive 3D wiring diagrams, rather than viewing things on 2D, 20-ft.-long drawings. It also helps users retain the information more efficiently while doing repair work [5].

AR now enables the support of on-site technicians through detailed information and guidance from highly specialized experts. This makes it possible for even hidden risks to be detected by experts through the “see-what-I-see” principle made possible by live video streaming and providing the full view of the technician’s environment. Moreover, technicians are more flexible in performing tasks due to the use of smart glasses that create the ability to perform various processes hands-free without distractions.

As part of effective maintenance for any aircraft, accurately assessing airframe damage resulting from combat or environmental hazards is crucial. Seemingly small factors—such as the depth of a scratch or the distance of a hole from supporting structures—can impact airworthiness, aircraft stealth capability and pilot safety. In an era of reduced spending and personnel, aircraft maintenance efficiency is essential. Recognizing these challenges, Lockheed Martin sought to streamline its damage and assessment process for the F-35 [7].

Traditionally, maintenance technicians would manually assess and track damaged areas by placing a transparent film over these areas

and tracing reference points (such as fasteners and seams) with a marker. They would then cross-reference this information with repair data history captured in a spreadsheet. However, cross-referencing line drawings did not provide the optimal platform to visualize repair information, it was also quite cumbersome and time consuming—leaving more room for maintenance errors.

Lockheed Martin selected NGRAIN to transform its damage assessment and repair system. Working closely with Lockheed engineers, NGRAIN developed the industry's first interactive 3D Virtual Damage Repair and Tracking™ solution. Traditional methods, involving line drawings and cumbersome spreadsheets, have been replaced with streamlined processes integrated with back-end software systems. With these virtual damage assessment solutions, aircraft maintenance technicians can:

- Increase operational availability of equipment: When an aircraft lands, maintenance staff can connect to the database and immediately determine whether the aircraft is airworthy.
 - Work more efficiently with fewer personnel: Using a streamlined process, maintenance staff can reduce the time required to document, assess, and repair damage. In the case of the USAF, because the F-35 and F-22 solutions have similar workflow, aircraft maintenance teams can easily transit between the two platforms.
 - Capture data more accurately: Providing maintenance personnel with the ability to visualize and accurately represent aircraft damage on a 3D model reduces the probability of human errors, which translates into aircraft system safety.
5. Worldwide connection of experts: Boeing and Microsoft HoloLens⁶

Besides hangar maintenance, the use of AR as a remote guidance solution means on-site staff can be connected to licensed technical experts from anywhere in the world. This real-time, cost-effective solution, provides expertise from any remote location, bridging the shortage of available AMEs allowing for multinational line maintenance teams.

For instance, in September 2020, the Australian Department of Defence announced that the No 36 Squadron aircraft technicians were using HoloLens mixed-reality devices with software developed by Boeing for C-17A Globemaster III aircraft maintenance [8]. Traditionally, specialist technicians of Boeing support repair and replacement tasks by travelling to Australia. The technicians are known as the recovery and modifications services team (RAMS). However, technicians were not able to visit this time due to the Covid-19 pandemic and the travel restrictions in place. This situation led to technicians exploring virtual reality with Microsoft HoloLens. Maintenance team supervisor Thomas Lane said:

“Through a secure ‘Cloud’ connection, my team and the technicians in the US can work seamlessly together by sharing screens and see exactly what they are seeing inside the aircraft through iris tracking. The first project was to replace the floatation equipment deployment systems panels inside C-17s, which consist of explosive components that deploy life rafts in an emergency.”

6. Effective staff training

In addition to improving distance collaboration, this technology also finds application in training inexperienced technicians and enabling an employee onboarding process as effectively as possible. Trainees can be adopted in a more engaging and interactive environment by using Augmented Reality.

7. IoT-enabled predictive maintenance: Flight health checks, predictive maintenance, and innovative product designs

The Internet of Things (IoT) is a collection of interrelated computing devices, mechanical or virtual machines, objects, and individuals that have unique identifiers and the ability to generate, exchange and consume data over a network without needing human-to-human or computer-to-computer communication. IoT has developed from the convergence of wireless technologies, micro-electromechanical systems, microservices, and the internet. IoT enables companies to automate processes and minimize labor costs. Aircraft is a classic example of the industrial IoT: engines have hundreds of sensors that transmit gigabytes of data for each flight, which adds up to terabytes of data to analyze across an airline’s fleet.

While IoT data generates the capacity for predictive maintenance, Artificial Intelligence (AI) introduces the capacity for smart, deep-learning machines to make predictive maintenance more accurate, actionable, and automatic.

AI predictive maintenance leverages IoT data to predict and prevent aircraft system failures. AI machine learning can access historical data stored on the cloud and process real-time data gathered through IoT sensors at the edge. Referencing a knowledge bank powered by these sources of data, AI can generate informed analytics to automate commands in real time. In the context of predictive maintenance, AI analytics and commands serve the specific function of optimizing maintenance cycles for aircraft systems. Outfitting an aircraft system with AI predictive maintenance can extend equipment life, improve efficiency among aircraft technicians, and significantly decrease expensive aircraft downtime, minimize flight delays and hazardous work conditions.

- Flight Health Checks—Rolls-Royce and Microsoft Azure IoT Solutions⁷

At a personal level, flight delays are disruptive and costly, but for the airlines the impact is exponentially larger. Minimizing the cost and disruption of maintenance activities is a key focus for these businesses. Rolls-Royce has over 13,000 engines for commercial aircraft in service around the world, and for the past 20 years, it has offered customers

comprehensive engine maintenance services that help keep aircraft available and efficient. The engine maker's TotalCare program allows airlines to pay for the hours they were able to fly rather than for repairs, which means Rolls-Royce has an interest in collecting data that helps airlines improve operations.

As the rapidly increasing volume of data coming from many different types of aircraft equipment overtakes the airlines' ability to analyze and take insight from it, Rolls-Royce has partnered with Azure IoT Solutions to find and develop the power of AI and IoT to accumulate data and find the best maintenance solution. Also, IoT devices with sensors are being used in aircraft engines to check the health of the engines and monitor all the components in real-time. In other words, Rolls-Royce is using Microsoft Azure IoT Suite and Cortana Intelligence Suite⁸ to support current and future generations of its intelligent aircraft engines. Microsoft said that with better understanding of flight operations, fuel usage and maintenance planning, airlines could potentially save millions of dollars per year.

- Predictive maintenance—Airbus and Skywise⁹

Predictive analytics by AI work through maintenance data. AI can interpret and organize data from sensors and send the data in a report which can easily be comprehended. Such algorithm also identifies and reports on potential failures in real-time and arranges proper timelines for repairs.

Companies like the Airbus is taking proactive steps to improve performance and reliability in aircraft maintenance. It is doing this by migrating historical maintenance information from aircraft and fleets to a cloud-based data repository known as Skywise.⁹ Airbus is also installing sensor systems on each aircraft in an airline's fleet to collate and record thousands of data parameters in real-time. After each flight, this data is uploaded to Skywise to be analyzed and to enable maintenance predictions for the future.

- Innovative Product Designs—Aircraft component manufacturers

AI is providing new ways to design and develop lighter, more efficient parts for the aerospace and aircraft maintenance industry. Efficient and lighter components are one of the most sought-after things in aviation. Design algorithms based on AI enable engineers to encompass a set of tools and techniques for creating intricate aircraft product designs from the data. It allows engineers to view multiple design options and features in less time to find the best designs. This approach helps develop new products with more functionality, making aircraft more sustainable and lighter in weight.

B. *Smart aircraft maintenance*

For airlines and AMOs, there are software provides maintenance planning, assignment, tracking, inventory management, and flight operation functionalities to give a fully integrated flight operations and maintenance management option.

The following maintenance situation illustrates how an aircraft maintenance organization can make use of bespoke mobile apps to achieve contactless job assignment and notification with improved productivity while also satisfying regulatory compliance. A summary of smart technology contribution to aircraft safety will follow.

First, maintenance supervisors can use assignment page in the Assignment app to review all the unassigned jobs based on input from the Maintenance Planning Department. Supervisors can also look at the current availability of all mechanics, and their respective shifts so they can determine which mechanic to pick for the job based on proximity.

If the supervisor wants to assign the job to a certain mechanic he can select from the list of available mechanics and look at his information, which will come up on screen with his availability and skills: if a mechanic does not have the right aircraft maintenance license and skill for the job, the system will not allow the supervisor to allocate that job to that person. In fact, the assignment process could be completely automated using the software scheduling and assignment capability. But, if the supervisor prefers to assign the job to a selected mechanic, that can be managed from the screen by clicking the icon. The system will allocate the job to him and send him a notification to his mobile app or smart watch. This is similar to an instant messaging app, avoids any paper handover and provides a complete contactless experience.

After the job has been assigned the mechanic will be granted electronic access to the job card within the mobile app. A desirable digital task card feature allows scanned PDF or OEM (Original Equipment Manufacturer) task cards to be rendered electronically within the mobile app itself.

Once the mechanic is assigned access to the electronic job card, he or she can review the steps and procedures within the job card and then sign-off after job completion by just selecting the check box on the job card. On selecting and checking these boxes and providing the sign-off information, the system will provide a dual authentication pop-up as per regulatory requirements for keeping electronic documents. After sign-off, the system will send an electronic footprint of the signature as a digital footprint for any future compliance references.

From the same mobile app, the mechanic can also gain virtual access to any reference documents or even launch an appropriate training video. This will allow the mechanic instant access to reference information without having to visit an engineering or a maintenance room to access training documents or maintenance manuals.

For defects reporting the machine learning tool in the mobile app could help mechanics expedite the resolution process by automating the form filling,

suggesting the next step, perhaps corrective action, and requesting the required parts. In the defects screen, when the mechanic starts typing the content, say, 'dent on the rotor blade', as the word 'dent' is entered, the AI will automatically list probable discrepancies based on the historical data of this aircraft.

The system will not only help fill in the data about the defect but will also suggest corrective action and, again, help in ordering the parts to rectify the defect. This will help to improve mechanics' productivity where they would otherwise have spent a lot of time filling out the rectification procedure and requesting the parts.

The mobile app may be integrated with other functions such as supply chain where it allows mechanics to request a part and, upon request, it automatically notifies the inventory control department and allocates the part from the request. That helps avoid the manual process of communication between mechanics and the inventory control department and reduces the time on getting the right parts. In the mobile app, the mechanic can type the required part details either by part number or just a description as the system provides a smart look-up. Also, mechanics should be able to track part availability in the inventory from the screen: not only parts in the warehouse, but it can also be parts currently on parked or stored aircraft. This will help the mechanic choose whether to use a spare part from the warehouse or an existing aircraft to help conserve cash. When the request is placed, the system will automatically generate a material request and notify the inventory control department. The system will automatically allocate the part for this request and notify the mechanic once the part is available to pick-up from the warehouse.

One of the key challenges that will remain is to determine how effectively mechanics can manage their work with minimal face-to-face interaction and manual process, while maintaining regulatory compliance. To this end, it is desirable that a Virtual Support feature along with co-browsing capabilities would allow mechanics to interact and share their work in a live mode with their supervisors and colleagues.

These capabilities can also be used to perform virtual inspections and sign-off by inspectors. This is a trend that readers will have noticed recently in the industry where even regulatory authorities are performing virtual inspections and audits of maintenance facilities due to the pandemic situation where they cannot do an on-site visit.

If a mechanic wants to review a damage notice while performing a job; he can instantly capture a picture from a page in the app and share that information with his supervisor as well as making markings or annotations on the image. The supervisor can base a recommendation on this information and will receive an instant notification where they can respond within the chat window or, for further investigation, the mechanic can share their page or screen with the supervisor and show live details of the work that the mechanic is currently performing, allowing the supervisor to review and even perform a virtual sign-off using the co-browsing capability. Mechanics can also hold a video call with their supervisor and live relay the hangar floor through that video call so that

the supervisor can look at the information and then provide a recommendation which will make the communication more effective.

This will help to minimize person to person interaction and help mechanics manage their work efficiently while helping to increase productivity even with the current limited workforce availability. This virtual support platform feature can be extended to another role in the airline organization.

In short, with the aid of smart technology aircraft maintenance activities will look different because better ways with less resources have been found to perform them. Following are the anticipated benefits of smart maintenance and associated challenges:

C. *Anticipated benefits of smart maintenance:*

- Better and efficient inventory control
Through mobile apps, mechanics should be able to track parts availability in the inventory from the screen: not only parts in the warehouse, but it can also be parts currently on parked or stored aircraft.
- Repair visualization and remote guidance
Due to the highly infectious virus and its variants, many airlines and AMOs implement social distancing policies such as reducing the number of staff working at the same time, splitting different teams among staff and work-from-home schemes. Such contingent measures reduce the effectiveness of staff daily work. Aircraft safety highly depends on maintenance. Maintenance activities require good planning, information sharing among the teams and teamwork. Fortunately, smart technology such as AR and IoT helps to regain the effectiveness of maintenance work during the crisis. Even though staff did not work together at the same time, information can be shared and accessed effectively. With the help of AR technology, remote staff/specialists can see what the on-site staff see, which makes communication clearer and more effective. AR technology can eliminate ambiguities when describing technical issues. There is nothing better than seeing what exactly is going on. All this helps to reduce errors due to procedure violations, misinterpretation of data and information, or insufficient training [6].
- Less occurrence of human errors
Owing to repair visualization and remote guidance, AMEs will be less distracted. Inadvertent violations of airworthiness regulations and common maintenance errors would be avoided.
- Real-time monitoring and maintenance
What to maintain, when to maintain and how to maintain are the key elements of ensuring airworthiness. After being hit by several waves of the Covid-19 outbreak, airlines are forced to reduce number of flights and number of staff. However, keeping aircraft airworthy requires a lot of manpower. Reducing staff may degrade the maintenance standard, especially when airlines are preparing to recover from the crisis. AI predictive maintenance can help airlines to recover from the crisis with minimal staff and not compromising aircraft safety. Through AI predictive maintenance aircraft operating data,

system status and fleet utilization could be analyzed to determine what to maintain and when to maintain. For example, one can monitor engine utilization and real-time operating parameters such as engine operating cycles, fuel consumption, oil pressure, exhaust gas temperature and vibration parameters, etc. By analyzing different aircraft operating data, the predictive system can predict which part of the aircraft engine/system is not working properly or working in degraded mode and maintenance plans can then be optimized. Very often troubleshooting is the most time-consuming process among maintenance tasks. To figure out what is going wrong on the aircraft can take up several days. Once the root cause is known, the time to get it fixed can be few hours or less. By using a smart prediction system, defective/degraded systems or components can be pinpointed. It not only minimizes the manpower for troubleshooting and aircraft ground time but also reduces the unnecessary parts removal.

- **Minimization of excessive or inadequate maintenance**

AI predictive maintenance overcome issues with traditional planned maintenance cycles that are based on pre-calculated time schedules where contextual factors may be overlooked. They solve the issue of over maintenance, where costs are wasted on the work time spent changing or repairing parts more frequently than necessary, as well as the unused portions of parts that go unaccounted for. Similarly, AI predictive maintenance overcome issues of poorly designed planned maintenance cycles that may result in under maintenance, where the lifespans of parts are overestimated and system is, therefore, at risk of failure due to overuse. Predictive maintenance with AI also is more effective than reactive maintenance where the repair is made after catastrophic aircraft system failures resulting in profit-inhibiting downtime. Condition-based maintenance activities will reduce time needed to complete work orders, thus reducing work order backlogs.

- Significant cost reduction in overseas travel as AR technology has replaced the need for face-to-face meetings.
- Staff training will be more effective as trainees are more engaging in an interactive environment and team collaboration will be enhanced.
- Better insights into the root cause of system and component failures. AI predictive maintenance system recognizes failure patterns and predict when malfunctions may arise based on the conditions under which they arose in the past. For instance, an AI-equipped aircraft that uses historical data to determine patterns in engine overheating events. Historically, each time the overheating occurred, temperature and speed sensors recognized that the engine was operating at above 475 knots during periods when the outdoor temperature was above 80°F. Through machine learning informed by data-collecting sensors, the AI-equipped aircraft built an understanding of the conditions that surround the overheating events. The next time the aircraft travels in above 80° temperatures, the AI-equipped aircraft can respond to speeds encroaching 475 knots by predictively

executing a command to notify pilots of the potential maintenance issue, and cap speeds at 470 knots to ensure safe functionality in the engine.

In view of the above contribution of smart technology to aircraft safety, airlines and AMOs should take this opportunity to implement smart maintenance to help themselves to recover from this crisis steadily.

D. *Challenges to the application of smart technology*

While the benefits of smart technology are manifold and the industry seems to be rapidly moving towards a smart technology-oriented environment, there are some factors to be considered while applying smart technology.

Firstly, security remains a predominant threat of the use of smart technology. This is because by forging connection between multiple devices within a cloud network, control over system authentication gets diluted. Anyone can access any information from a wide network of connected devices now.

Secondly, related to security, the privacy of data is another major challenge. Within the network, a substantial amount of user data gets released and the users often lose control over their own data.

Moreover, while the overall usage of smart technology is resource efficient, the deployment process entails layers of complexities and can be potentially expensive.

4 COVID-19-Induced Problem

Meanwhile there are two important COVID-19 problems to be resolved, viz. technical and human factors.

Technically, there are risks related to prolonged parking/storage and these aircraft system problems implies a serious impact on airworthiness:

- Sticking/high friction of valves in engine bleed air system leading to pneumatic system issues during flight (e.g., bleed air loss, in one particular case this led to an in-flight shut down).
- Erroneous air data information including contaminated / blocked pitot-static systems and Angle of Attack vanes failure.
- Fuel system contamination caused by non-adapted water drainage intervals or lack of available biocide.
- Emergency batteries, post parking or storage procedures, not at the expected state of charge.
- Depletion of aircraft parking brake accumulator pressure leading to damaged aircraft in ground incident.
- Wildlife nesting in the aircraft/engines while parked/stored, including insects, birds, and rodents.

The second issue is related to staff readiness for work:

- Decreased wellbeing of maintenance professionals during shutdown—The pandemic is a significant source of anxiety, stress, and uncertainty for almost everyone. Worries about unemployment for aviation staff and their relatives may be exacerbated. During the shutdown, with people working from home and therefore isolated from normal support, the personal wellbeing of professionals is likely to have suffered. For those working, this may lead to task distraction/interruption, workload/task saturation, instructions or requirements not followed.
- Maintenance personnel fatigue—With redundancy and furlough reducing the available number of personnel, those left working may have to work additional hours. The preparation for and eventual return to new normal operations will require significant additional effort in comparison with actual normal operations. These may both contribute to rising levels of fatigue.
- Personnel no longer working collaboratively—Significant gaps in working, or working from home, may have reduced people’s ability to work collaboratively. This may exacerbate problems with teamwork and shift turnover while wearing personal protective equipment.
- Reduced adherence to procedures in the new working environment—Reduced operations and underload may create a belief that the level of risk within the operating environment has substantially reduced, causing staff to become less sensitive to risk with the possibility that they are less alert, and procedures are not completely followed.
- Skills and knowledge degradation due to lack of recent practice—The 90% reduction in traffic means that most aviation professionals are not performing their normal tasks, sometimes they are doing a substantially different job, and sometimes not working at all or at a substantially reduced frequency. Workshop and classroom-based training is also not taking place. Together, this creates a reduction in the practical skills and knowledge of aircraft maintenance professionals, and with it associated safety risks.

To augment the safety of aircraft systems, one needs to solve these aircraft system and work readiness issues. To this end the Safe Aircraft System (SAS) model, based on the human factor model DD-SHELL of accident causation is proposed [9].

Briefly the characteristics of the Dirty Dozen and the SHELL models are:

- Dirty Dozen is a concept developed by Gordon Dupont, in 1993, whilst he was working for Transport Canada. It has since become a cornerstone of Human Factors in Maintenance training courses worldwide. The Dirty Dozen model is particularly useful to ensure maintenance staff fitness for work during the pandemic. The Dirty Dozen refers to twelve of the most common human error preconditions, or conditions that can act as precursors, to accidents or incidents. These twelve elements influence people to make mistakes and they are: Lack of communication, Complacency, Lack of knowledge, Distraction, Lack of teamwork, Fatigue, Lack of resources, Pressure, Lack of assertiveness, Stress, Lack of awareness, and Norms [10].
- The SHELL Model was developed by Edwards (1972) and later modified by Hawkins (1987). The basic premise of the SHELL Model is that a person (one of

the Liveware components) does not act alone, but must interact with other Liveware (team members, etc.), Hardware, and Software, in a particular Environment. The interaction of these “components” may have a positive or negative affect on error reduction [11].

SHELL MODEL examples include:

- Software: Procedures, policies/rules, manuals, placards
- Hardware: Tools, equipment, aircraft, work space
- Environment: Physical environment (social distancing, light, noise, weather); Organizational (structure, teams); Political/Regulatory (FAA, OSHA, NTSB); Economic factors (competition, market factors)
- Liveware (Individual): Individual issues of physical health (compulsory COVID tests and rapid antigen test before going to work), knowledge, attitude, stress, and perception of culture
- Liveware (Others): Teamwork, leadership, workplace norms, communication.

5 Safe Aircraft System (SAS) Model

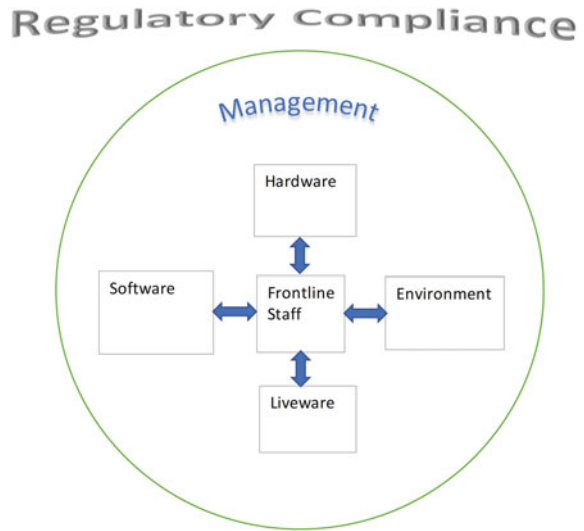
To understand how the SAS modeling process can be used to improve aircraft system safety, it is necessary to define what a system is. Aircraft systems could be flight operations, operational control (dispatch/en-route flight monitoring), maintenance and inspection, cabin safety, ground handling and servicing, cargo handling, and training. Within these systems there are subsystems. Some examples of subsystems include crew scheduling systems, training curricula, maintenance control, deicing, fueling, aircraft fleet, and ground handling.

According to the Federal Aviation Administration (FAA), human error is not avoidable, but it is manageable. The SAS model (Fig. 2) comprises the management function and 4 SHELL model interfaces:

- Frontline staff-Software interface
- Frontline staff-Hardware interface
- Frontline staff-Environment interface
- Frontline staff-Liveware interface.

Frontline staff and Liveware are the human elements or people in the aircraft system. Examples include maintenance technicians, supervisors, pilots, and ground crew, etc. To attain system safety, the four system components (Software, Hardware, Environment, and Liveware) should be carefully adapted and matched with the frontline staff component [12]. Maintenance operations during the pandemic era are affected by various organizational and human factors. The role of the management and the interface between frontline staff and four system components are described below.

Fig. 2 SAS model components



1. Management function

World Health Organization defines health as “a state of complete physical, mental and social well-being and not merely the absence of disease. Besides enforcement of public health control measures, carrier management should minimize workers’ exposure to job-related physical and psychosocial risks and promote healthy behaviors among workers, both job- and lifestyle-relate, e.g., through healthy corporate culture and staff wellness program.

Airline management should note that financial pressure during the pandemic may lead management to seize business opportunities (e.g., charter flights) without a proper risk assessment, or to tolerate violations such as dispatch of flights with unserviceable equipment beyond the acceptable limits. The perceived pressure may also lead crews and maintenance personnel to accept “cutting corners” to avoid costly delays and flight cancellations. Management attention should therefore be given to following risks:

- Flight crew may avoid requesting extra fuel or refrain from performing a go-around when the situation dictates;
- Maintenance staff may overlook required actions, while striving to avoid a delayed departure;
- Frontline staff may believe that deviations from the Standard Operating Procedures (SOP) are justified in the context of the crisis;
- The risks induced by fatigue may also be increased, as operations may stretch flight duty periods to the limits, possibly in combination with reduced rest times (longer sanitary procedures, less expensive hotels far away from the airport, unavailability of local crew transportation, etc.).

2. Frontline staff task readiness

According to the FAA, human factors directly cause or contribute to many aviation accidents. If they are not detected, they can cause events, worker injuries, wasted time, and even accidents. As the core component of the SAS model, the staff concerned must be fit for the project/task assigned and to watch out for any of the following undesirable Dirty Dozen human behavior:

- **Poor or lack of communication skills**
During the COVID-19 period, there may be instances of inadequate/ineffective communication of new rules, Standard Operating Procedures from regulator, airline management or the MRO management.
- **Complacency**
Complacency can be described as a feeling of self-satisfaction accompanied by a loss of awareness of potential dangers. Such a feeling often arises when conducting routine activities that have become habitual and which may be “considered”, by an individual, as easy, and safe. A general relaxation of vigilance results and important signals will be missed, with the individual only seeing what he, or she, expects to see. Lockdown had changed the way we lived since its implementation. Time management and priorities had shifted and a refocus could take time. As a result, over confidence or under confidence may occur.
- **Lack of knowledge or skills**
The regulatory requirements for training and qualification can be comprehensive, and organizations are forced to strictly enforce these requirements. However, lack of on-the-job experience and specific knowledge can lead workers into misjudging situations and making unsafe decisions. Aircraft systems are so complex and integrated that it is nearly impossible to perform many tasks without substantial technical training, current relevant experience, and adequate reference documents. A lot has changed with COVID-19 still around. Ramp procedures, staff screening, first aid and Cardiopulmonary Resuscitation guidelines, aircraft disinfection, etc.—there is a lot of unlearning and new learning. It takes a while to become habituated to the new information and procedures. Systems and aircraft operating procedures can change substantially due to the COVID-19 pandemic procedures and employees’ knowledge and skills can quickly become out-of-date.
- **Distraction**
Distraction could be anything that draws a person’s attention away from the task on which they are employed. Some distractions in the workplace are unavoidable, such as loud noises, requests for assistance or advice, and day-to-day safety problems that require immediate solving. Other distractions can be avoided, or delayed until more appropriate times, such as messages from home, management decisions concerning non-immediate work (e.g., shift patterns, leave entitlement, meeting dates, administrative tasks, etc.), and social conversations. Psychologists say that distraction is the number

one cause of forgetting things: hence the need to avoid becoming distracted and to avoid distracting others.

Change of living patterns during this COVID-19 pandemic period, a sick or unattended family member, increased procedures, new protocols and return to work after a gap are all distractors. Since new systems take time to be seamlessly established; and the dynamics of the problems may call for further changes, one needs to be mindful, agile, and focused.

- Lack of teamwork skills

In aviation many tasks and operations are team affairs; no single person (or team) can be responsible for the safe outcomes of all tasks. However, if someone is not contributing to the team effort, this can lead to unsafe outcomes. This means that workers must rely on colleagues and other outside agencies, as well as give others their support. Teamwork consists of many skills that each team member will need to prove their competence. Some of the key teamwork skills include leadership, followership, effective communication, trust building, motivation of self and others, and praise giving. The dynamics of the pandemic situation are such that everyone is still learning and coping.

- Fatigue

Fatigue is a natural physiological reaction to prolonged physical and/or mental stress. We can become fatigued following long periods of hard work. As we become more fatigued our ability to concentrate, remember and make decisions reduces. Therefore, we are more easily distracted, and we lose situational awareness. Readjusting to COVID-19 working schedules after a break could be difficult initially. Obesity, lack of exercise, and relearning of tasks can also lead to easy physical and mental fatigue.

- Lack of resources

The financial trough may leave fewer people to do more work. Social distancing may allow fewer personal interactions. Constant screening, protection and disinfection measures may burden existing resources thus leading to some gaps in the system. Regardless of the task, resources also include personnel, time, data, tools, skill, experience, and knowledge etc. A lack of any of these resources can interfere with one's ability to complete a task. It may also be the case that the resources available, including support, are of a low quality or inadequate for the task.

When the proper resources are available, and to hand, there is a greater chance that we will complete a task more effectively, correctly, and efficiently. Therefore, forward planning to acquire, store and locate resources is essential. It will also be necessary to properly maintain the resources that are available; this includes the humans in the organization as well.

- Pressure

Pressure is to be expected when working in a dynamic environment. However, when the pressure to meet a deadline interferes with our ability to complete tasks correctly, then it has become too much. Fear of contracting the virus,

restricted movements during layovers, tight schedules due to reduced number of maintenance workers, are all additional pressures while working.

- Lack of assertiveness

Assertiveness is a communication and behavioral style that allows us to express feelings, opinions, concerns, beliefs and needs in a positive and productive manner. Being both unable to express our concerns and not allowing other to express their concerns creates ineffective communications and damages teamwork. Unassertive team members can be forced to go with a majority decision, even when they believe it is wrong and dangerous to do so.

- Stress

There are many types of stress. Typically, in the aviation environment there are two distinct types—acute and chronic. Acute stress arises from real-time demands placed on our senses, mental processing, and physical body such as dealing with an emergency or working under time pressure with inadequate resources. Chronic stress is accumulated and results from long-term demands placed on the physiology by life’s demands, such as family relations, finances, illness, bereavement, divorce, or even winning the lottery. When we suffer stress from these persistent and long-term life events, it can mean our threshold of reaction to demands and pressure at work can be lowered. Thus at work, we may overreact inappropriately, too often and too easily. Financial losses, anger, lack of control, frustration, resentment, lack of confidence, uncertainty of the future and return to new ‘normal’, the need to earn, new ways of living at home, at work and in society, inconvenience caused by wearing PPE suits while at work, required Covid tests, etc. all these would add to stress.

- Lack of awareness

Working in isolation and only considering one’s own responsibilities can lead to tunnel vision; a partial view, and a lack of awareness of the effect our actions can have on others and the wider task. Such lack of awareness may also result from other human factors, such as stress, fatigue, pressure, and distraction. During the COVID-19 period, operation guidelines are changing often, employer may have to change protocols as per national and international requirements and lessons learnt after resuming operations. It is imperative to be aware of the risks and what is the latest expectation of the organization.

- Norms

Workplace practices develop over time, through experience, and often under the influence of a specific workplace culture. These practices can be both, good and bad, safe, and unsafe; they are referred to as “the way we do things round here” and become Norms. Unfortunately, such practices follow unwritten rules or behaviors, which deviate from the required rules, procedures, and instructions. These Norms can then be enforced through peer pressure and force of habit. It is important to understand that most Norms

have not been designed to meet all circumstances, and therefore are not adequately tested against potential threats.

3. Frontline staff/Software Interface

Software are non-physical, intangible aspects of the aviation system that govern how the aircraft system operates and how information within the system is organized. Software may be likened to the software that controls the operations of computer hardware. It includes rules, instructions, regulations, policies, laws, orders, safety procedures, standard operating procedures, customs, practices, conventions, habits, supervisor commands and computer programs. This interface encompasses the connection between humans and the non-physical aspects of the system such as procedures, manual and checklist layout, computer programs and mobile apps.

4. Frontline staff/Hardware Interface

Hardware refers to physical elements of the aviation system such as aircraft systems (including controls, surfaces, displays, functional systems and seating), operator equipment, tools, materials, maintenance platforms, ground power units, etc.

This interface is the most considered when speaking of human-machine systems: the selection and use of correct tools; of smart glass to match the sensory and information-processing characteristics of the user; of ground vehicles with proper powering unit, etc.

5. Frontline staff/Environment Interface

This interface shows the circumstances in which aircraft and system resources (software, equipment, colleagues) operate, made up of physical, organizational, economic, regulatory, political, and social variables that may impact on the frontline staff. Internal aircraft environment relates to immediate work area and includes physical factors such as government health control measures, social distancing, ambient temperature, air pressure, humidity, noise, vibration, and ambient light level. External air transport environment includes the physical environment outside the immediate work area such as weather (visibility/turbulence), terrain, congested airspace and physical facilities and infrastructure including airports as well as broad organizational, economic, regulatory, political and social factors [13]. These aspects of the environment will interact with the frontline staff via this interface.

6. Frontline staff/Liveware Interface

This interface refers to frontline staff interaction with other people in the system. For example, AME with team members, maintenance supervisors, inspectors, maintenance planner, shift managers, station managers, inventory controllers, ground support, other management and administration personnel, etc. This between-people interface considers human performance, capabilities, and limitations [13]. Frontline staff/Management relationships are also within the scope of this interface, as corporate climate and company operating pressures can significantly affect human performance.

6 SAS Modeling Procedure

The SAS modeling process is triggered whenever new operation or changes to systems are being considered. For example, changes to carrier operation could include the reorganization of routes, closing of line maintenance stations abroad, adding or changing contractual arrangements for services, long-term parking and storage of existing fleet or major modifications of existing fleet, or any one of many different types of operations. In fact, the SAS modeling process may also be triggered by any revision of an existing system or a change of operational environment due to the incidence of COVID-19 pandemic.

There are five steps in implementing the SAS model and their outputs need to be documented for regulatory compliance:

1. System description and analysis.
2. Potential consequences identification.
3. System safety risk analysis.
4. System safety risk assessment.
5. System risk controls.

1. System description and analysis

The system description and analysis process should include representatives from management, safety staff, subject matter experts, employees, and representation groups (e.g., unions) formed into workgroups such as safety committees, safety roundtables, safety action groups, or similar titles. Since many, if not most, system changes involve allocation of resources, the accountable executive or other managers with the authority to commit resources should be included in the process.

In conducting a system analysis, the following information must be considered:

- Function and purpose of the system.
- The system's operating environment.
- An outline of the system's processes and procedures.
- The personnel, equipment, and facilities necessary for operation of the system.

Systems analysis is the primary means of proactively identifying and addressing potential problems before the new or revised systems or procedures are put into place. The system analysis should explain the management involvement and interactions between the frontline staff and hardware, software, people, and environment that make up the system in sufficient detail to identify potential hazards and perform risk analyses.

An example of system description and analysis maybe the introduction of a transformed passenger fleet to meet the cargo demand. Several of the carrier's organizational "systems" would be affected, viz. flight operations, maintenance, station, ground and cargo handling, etc. As part of an analysis of the flight operations system, one would need to consider changes to pilot and cabin crew training, scheduling, crew rest, cabin facilities, and other areas.

System analysis should identify and consider activities and resources necessary for the system to function. For example, in transforming passenger aircraft to carry out cargo operations, one would identify for the pilot training system, as one of the affected systems, the activities and resources necessary for pilot training to pilot the transformed passenger aircraft. These may include new simulator exercises, training curriculum, training aids, and instructors.

The above example of transforming passenger aircraft to carry cargo in passenger cabin would entail consideration of several systems and a variety of procedures due to the number of processes that would be affected. However, in a simpler case, such as the introduction of aircraft disinfection procedure due to COVID-19, only those elements of the systems that would be affected by the change would need to be considered. Staff assigned for using cleaning, sanitizing, and disinfecting products would ensure the procedures and products are approved by the public health authority and aircraft manufacturer. The carrier would not be expected to perform a comprehensive analysis of the entire cabin system.

2. Potential hazard/consequences identification

Audits and evaluations bring decision makers important information. However, these tools can be limited by the scope and content of their design. The workforce is an important information source that should be included in the data-gathering process. Frontline employees may observe aspects of the operation or environment that were not expected and were not included in audit or evaluation protocols. In this respect, the employee reporting system can fill in important gaps in the company's data collection process. To be effective, the organization needs to establish and maintain an environment in which employees feel comfortable to report hazards, errors, violations, and concerns, as well as occurrences, incidents, etc., and propose safety solutions and improvements. The accountable executive and management team need to encourage employees to report safety issues and not fear reprisals from management. Policies that assure employees of fair treatment and clear standards of behavior are an essential part of the reporting process. A key aspect of the confidential reporting system is that it is confidential. Therefore, management must define methods for employee reporting and de-identification of sources without losing essential information.

In other words, a carrier would use its own experience (and/or those of others in the industry where available), airworthiness requirements, manufacturers' technical data, and knowledge of department operations to identify potential consequences resulted from its operation. These could include the effectiveness of new procedure training, employees failing to read or misinterpret newly published procedures, supervisors failing to follow the new fatigue risk management procedures, etc. While identification of every conceivable consequence is unlikely, every staff is expected to exercise due diligence in identifying consequences that could foreseeably lead to an aircraft accident.

3. System safety risk analysis

For each identified consequence, the potential for injury and damage that may result from an accident related to operating while exposed to the event is to be defined. To determine potential for injury and damage, one needs to define the likelihood of occurrence of an accident and severity of the injury or damage that may result from the system hazard. It is important to note that the likelihood and severity do not refer to the consequence itself but of a potential occurrence (accident or incident) related to the consequence. Review of accident statistics, failure data, error data (e.g., ramp accident reports), information from the National Aeronautics and Space Administration's Aviation Safety Reporting System or equipment reliability data would help in determining likelihood.

The risk analysis also needs to consider the basis for the estimates of severity. For instance, if the potential consequence could result in controlled flight into terrain, the severity of this outcome is normally major, if not catastrophic. Conversely, tire failures, while potentially leading to a fatal accident, often lead only to aircraft damage.

4. System safety risk assessment.

Once the risk is analyzed, the management must assess whether the risk is acceptable. A common tool used in risk assessment decisions is a risk matrix. A risk matrix provides one with a way to integrate the effect of severity of the outcome and the probability of occurrence, which enables him to assess risks, compare potential effectiveness of proposed risk controls, and prioritize risks where multiple risks are present.

If a risk matrix is used, the carrier should develop criteria for severity and likelihood that are appropriate for their type of operations and their operational scenario. For example, severity levels are sometimes defined in terms of a dollar value of potential damage. The level of detail and complexity of tables and matrices should be adapted to the particular needs and complexities of each organization. It should also be noted that organizations might include both qualitative and quantitative criteria.

Safety risks are conceptually assessed as acceptable, tolerable, or intolerable. Safety risks assessed as initially falling in the intolerable region are unacceptable under any circumstances. Consequences of aircraft system operation are assessed according to their probability (in number) and severity (in alphabet). For example, risk probability 5 indicated a frequent event and risk severity A is catastrophic, B hazardous, and C a major threat to safety. Any system operation consequences being assessed as 5A, 5B or 5C in Tables 1 and 2 are therefore considered unacceptable. The probability and/or severity of the consequences of the hazards are of such a magnitude, and the damaging potential of the hazard poses such a threat to safety, that mitigation action is required or activities are to be stopped. Table 1 depicts a sample risk matrix [11] and Table 2 shows a sample safety risk tolerability.

Hence, if the risk is acceptable, then the system may be placed into operation and monitored in the carrier's Safety Assurance process. If the risk is not acceptable, one

Table 1 Sample safety risk matrix

Risk probability	Risk severity				
	catastrophic A	Hazardous B	Major C	Minor D	Negligible E
Frequent (5)	5A	5B	5C	5D	5E
Occasional (4)	4A	4B	4C	4D	4E
Remote (3)	3A	3B	3C	3D	3E
Improbable (2)	2A	2B	2C	2D	2E
Extremely improbable (1)	1A	1B	1C	1D	1E

Table 2 Sample safety risk tolerability

Safety risk index range	Safety risk description	Recommended action
5A, 5B, 5C, 4A, 4B, 3A	Intolerable	Take immediate action to mitigate the risk or stop the activity. Perform priority safety risk mitigation to ensure additional or enhanced preventive controls are in place to bring down the safety risk index to tolerable
5D, 5E, 4C, 4D, 4E, 3B, 3C, 3D, 2A, 2B, 2C, 1A	Tolerable	Can be tolerated based on the safety risk mitigation. It may require management decision to accept the risk
3E, 2D, 2E, 1B, 1C, 1D, 1E	Acceptable	Acceptable as is. No further safety risk mitigation required

will need to pursue the next process in SAS modeling, which is developing system risk controls.

Risk assessments must involve the levels of management with the authority to make risk decisions, deciding what is or is not an acceptable risk for the systems within their area of operational responsibility. For example, dispatching a flight that presents a high risk (e.g., Countries/areas with active Covid-19 community transmission) might require the Flight Operations Director to approve or authorize the flight. For large scale operational decisions, the accountable executive may be the only appropriate person to make these risk acceptance decisions. Thus, the person responsible for making these risk acceptance decisions will depend on the scope of the proposed change to the operation and the level of risk presented.

5. System risk controls

After events and associated risk are fully understood, risk controls must be designed for risks that the carrier deems unacceptable. Examples of risk controls include new processes, equipment, training, new supervisory controls, new equipment or hardware, new software, changes to staffing arrangements, or any combination of system changes. In short, anything that would lessen the likelihood or severity of

a potential incident/accident. Next, look at the system with the proposed control in place to determine if the level of risk is now acceptable and the proposed control does not introduce unintended consequences or new hazards. Assuming that the controls are approved by all concerned, they were implemented operationally and specifically targeted for performance monitoring by the flight operations, safety, and engineering departments. To facilitate the condition monitoring of aircraft systems, alert levels can be established [14]. The case study below illustrates how SAS modeling can be implemented in an operational context.

7 SAS Modeling Case Study

7.1 Transport of Cargo in Passenger Aircraft

Air cargo services are vital for the economy and for coping with the logistical challenges related to COVID-19, as global supply chains depend on them being unhindered. Air cargo should therefore be able to deliver continuously critical products such as food, medical supplies (e.g., masks, gloves, clothing, etc.) and personal protective equipment (PPE), and other products, which are vital for the functioning of sensitive supply chains. Consequently, more and more operators are considering the transport of cargo in the cabin of their passenger aircraft [15].

7.2 System Description and Analysis

Assuming that a carrier has been looking at new business opportunities using passenger aircraft to transport cargo on passenger seats and in the passenger compartment. To identify potential hazards related to operating flights using transformed cabin configurations the SAS modeling process starts with an analysis of the systems involved, with the addition of a new set of procedures for the transformed aircraft. Besides planning and transforming the passenger cabin, this would include an analysis of the flight operations and training program.

7.3 Potential Hazard/Consequences Identification

Based on various sources of information, it was agreed that the new cargo operation likely to trigger following consequences:

- Incorrect loading and unloading sequence;
- The potential for mis-declared/undeclared or hidden dangerous goods within cargo;

- Inadequate handling of dangerous goods;
- Unrestricted access to all cargo loaded into the cabin;
- The detection of any smoke or fire and firefighting capabilities of personnel in the cabin;
- Qualification and abilities of crew member or other personnel to control and put out fire in cabin;
- Qualification of ground staff to prepare and load cargo in accordance with applicable regulations and instructions;
- Occupational Health and Safety risks associated with the new procedures;
- Operational weight and balance limits exceedance;
- Unsecured/incorrectly loaded cargo;
- Cargo leakage/spillage;
- Loose trolleys or canisters crashing into the cabin;
- Overheating of the galleys' equipment and the in-flight entertainment system;
- The provision, location and storage of sufficient firefighting equipment such as portable breathing equipment, fire extinguishers etc. for use by personnel carried in the cabin.

Additionally, regarding the new cabin cargo storage system, the carrier identifies that pilots, cabin crew, AMEs, load masters and ground crew are not properly trained for this type of operation.

7.4 System Safety Risk Analysis

To analyze the risk associated with this new operation, a team is assembled with representatives from management, engineering department, flight and cabin crew training, flight dispatch, load control and ground handling.

For each of the consequences identified, the risk analysis team has to determine the potential for injury and damage, it needs to estimate the likelihood of occurrence of an accident and severity of the injury or damage that may result from an aircraft accident. Hence the safety risk matrix for each event was formulated for prioritization and the results are shown in Table 3.

7.5 System Safety Risk Assessment

Although risk assessment relating to each of the identified irregularities indicated an acceptable level of safety (i.e., no assessment belonged to categories 5A, 5B or 5C), a management concern is that the procedures and associated training currently in use may not provide an acceptable level of safety for some unexpected events of the new operation and safety risk controls would be required.

Table 3 Safety risk associated with new cabin cargo storage system

Possible consequences	Safety risk index	SAS model interface(s)
Incorrect loading and unloading sequence	4C	Frontline/software Frontline/liveware
The potential for mis-declared/undeclared or hidden dangerous goods within cargo;	4D	Frontline/software Frontline/liveware
Inadequate handling of dangerous goods	4C	Frontline/software
Unrestricted access to all cargo loaded into the cabin	3C	Frontline/liveware
The detection of any smoke or fire and firefighting capabilities of personnel in the cabin	3C	Frontline/software Frontline/liveware
Qualification and abilities of crew member or other personnel to control and put out fire in cabin;	3D	Frontline/liveware
Qualification of ground staff to prepare and load cargo in accordance with applicable regulations and instructions;	3D	Frontline/liveware
Occupational Health and Safety risks associated with the new procedures	3B	Frontline/software Frontline/liveware
Operational weight and balance limits exceedance	3C	Frontline/software
Unsecured/incorrectly loaded cargo;	3D	Frontline/software Frontline/hardware
Cargo leakage/spillage	3B	Frontline/hardware
Loose trolleys or canisters crashing into the cabin	3D	Frontline/software Frontline/hardware
Overheating of the galleys' equipment and the in-flight entertainment system	3D	Frontline/hardware
The provision, location and storage of sufficient firefighting equipment such as portable breathing equipment, fire extinguishers etc. for use by personnel carried in the cabin	3D	Frontline/Environment Frontline/hardware
Familiarity of pilots, cabin crew, AMEs, load masters and ground crew with the new cargo operation with passenger aircraft	5D	Frontline/software Frontline/liveware

7.6 System Risk Controls

With regard to the operational issues and identified potential consequences, safety risk control measures based on the SAS model were developed and these include:

Management actions:

- To ensure the installation of aircraft interior stowage devices would be carried out by a reliable approved maintenance organization.
- Two or more crew members are required to survey and access all areas of the passenger compartment during all phases of flight. Any fire that occurs must be discovered and extinguished immediately utilizing emergency equipment. These crew members are considered additional crew members with specific duties assigned during the flight and would be in addition to the required flight crew members.
- Provision of training for staff concerned, viz. pilots, additional cabin crew, load masters and ground staff.
- To ensure continuing airworthiness of the fleet.
- To consider proper balancing of work and required resources.

Frontline staff task readiness:

- To ensure personal safety with emphasis on public health control measures.
- To ensure individual staff possess the necessary qualification, skills, and experience, have developed a sense of awareness and be a team player.
- To ensure individual staff's readiness for work, i.e., free of stress and fatigue.

Frontline staff/Software issues:

- Set an action plan for potential undeclared dangerous goods that are hidden within the consignment.
- Only flight and cabin crew members that are directly related to the operation of such cargo flights would be permitted. They will be required to be trained with respect to their duties and have training records.
- Set up the cabin cleaning and disinfection and cargo storage (including undeclared or hidden dangerous goods) procedures according to regulatory requirements, aircraft manufacturer and public health authority guidelines.
- When handling hazardous materials, all crew members and ground staff must be aware of the safety precautions.

Frontline staff/Hardware issues:

- The transport of cargo on passenger seats, or, in the passenger compartment, will require adequate restraint systems/means which must address the ground, flight, turbulence, take-off, landing, and emergency landing conditions.
- The additional crew members must be provided appropriate seats that are not located near the cargo (i.e., first rows). These seats must be certified for use during taxi, take-off, landing, flight, and for emergency landing conditions. The additional crew members seating location must be provided with a means of two-way communication with the flight crew members.
- AMEs to ensure the serviceability of operational and emergency equipment.

Frontline staff/Environment issues:

- Aircraft crew concerned should ensure a safe working environment in both the cockpit and cabin, and report any discrepancies as soon as possible.
- Ground staff should ensure a safe working environment on the ramp.

Frontline staff/Liveware issues:

- Flight crew and ground staff: Restrict the access to the cargo deck to authorized qualified personnel.
- AMEs and flight dispatchers: Prepare accurate weight and balance parameters according to the type of cabin cargo configuration.
- Authorized cabin crew and load masters:
 - Ensure storing configuration does not hinder the use of emergency equipment.
 - Ensure the availability of operational and personal protective gear.
 - Strengthen supervision of loading and unloading sequences to guarantee freight is properly stocked.

Prior to implementation of the risk controls, the carrier checked the resulting procedures in flight and cabin simulators to ensure that the risk control measures would mitigate risk to an acceptable level. The solution and simulator results were then submitted to the director of flight operations, head of engineering and director of safety for management review and acceptance of the mitigated risk. As the controls were approved by all concerned, they were then implemented operationally, subject to work readiness and wellbeing of concerned staff. Per the safety assurance requirements, these controls were then specifically targeted for performance monitoring by the flight operations, safety, and engineering departments.

8 Conclusion

Under the influence of COVID-19, airline operators have never faced such an extensive disruption of their operations and cannot rely on previous experience when analyzing the possible scenarios and solutions. Furthermore, they do not have sufficient and reliable data to identify all emerging hazards associated with long-term parked and stored aircraft. Aircraft repairs and maintenance during the pandemic offer a chance of innovation, and everyone concerned should be encouraged to think outside the box to come up with new ways of performing aircraft maintenance. Now with travel restrictions, social distancing, less staff and resources, there is an urgent need for smarter ways of doing maintenance.

Through appropriate application of smart technology, aircraft maintenance is facilitated tremendously, which translates into aircraft system safety. With the aid of augmented reality, less face-to-face meetings are required, remote guidance of world-renowned aircraft experts is readily available, staff training is more effective, and AMEs would reduce errors due to procedure violations, misinterpretation of

standard operation procedure data, or inadequate training. The ability of IoT-enabled predictive maintenance allows real-time aircraft system monitoring and maintenance, permits the use of historical data to prevent future failures, identify component failure causes and patterns, eliminates expensive downtime and consequential damages. Through smart technology AMEs are freed from routine maintenance to focus on more complex problems during recovery, resulting in better aircraft safety.

Meanwhile, there are technical and human factors problems induced by COVID-19 to be tackled. To this end the Safe Aircraft System (SAS) model, based on the human factor model DD-SHELL of accident causation has been proposed to solve various aircraft technical and work readiness issues created by the pandemic. The SAS modeling process is triggered whenever new operation or changes to systems are being considered. Changes to carrier operation could include the rearrangement of flight routes, termination of maintenance stations abroad, outsourcing of contractual arrangements for services on prolonged parked and stored fleet, or major modifications of existing aircraft. In fact, the SAS modeling process may also be triggered by any revision of an existing system or a change of operational environment due to the incidence of COVID-19 pandemic. A case illustration shows that the SAS model would be a useful tool in identifying potential hazards/consequences associated with any major or minor changes in flight operations.

However, the recovery phase may require significant changes in the operator's business model and operational procedures; an inadequate documentation and communication of changes may introduce additional risks. Poor control of changes has significantly contributed to serious incidents and accidents in aviation history. Therefore, during the recovery phase, operators need to focus on the implementation of a robust management of change process to implement smart maintenance and the SAS modeling process.

In summary, the simultaneous execution of smart maintenance and the SAS model would be effective in enhancing aircraft system safety.

Endnotes

1. YAPE is an AI-based transport and delivery robot developed by YapeSrl (the acronyms stand for "Your Autonomous Pony Express"), a company of Italian hi-tech manufacturer e-Novia. The self-driving transport robot helps passengers carry small luggage and navigate through the terminals. Fraport, the operator of Frankfurt Airport (FRA), and e-Novia deployed the little vehicle in FRA's transit area.
2. In November 2018, Vanderlan de achieved a significant milestone for its future-proof baggage logistics solution, FLEET. Beginning live operations at Rotterdam Hague Airport (RTHA)—and supported by on-site handler A via partner. The overall project was aimed at improving the baggage handling process at the airport through using intelligent autonomous vehicle technology.
3. AirAsia has always been a digital-led company and was ahead of the game by launching AVA in January 2019. The chatbot has the capability of managing 25,000 queries at any time in 15 languages.

4. Tail assignment is a tool which can improve aircraft utilization by factoring in crew connections, maintenance, fuel usage, operational costs, and constraints for each aircraft in a fleet.
5. nGRAIN brings AI to damage assessment, serving customers in industries including Aerospace, Defense, and Insurance. Today, thousands of people use NGRAIN every day around the world, including high performance teams at some of the most diverse organizations on the planet, including Lockheed Martin, Raytheon, Microsoft, the US and Canadian governments, R&D labs, schools, and universities. [Online] Available: <https://www.ngrain.com/>. (Accessed 4 May 2022).
6. Microsoft Hololens is an ergonomic, untethered self-contained holographic device with enterprise-ready applications to increase user accuracy and output. [Online] Available: <https://www.microsoft.com/en-us/hololens> (Accessed 4May2022).
7. The Azure Internet of Things (IoT) is a collection of Microsoft-managed cloud services that connect, monitor, and control billions of IoT assets. In simpler terms, an IoT solution is made up of one or more IoT devices that communicate with one or more back-end services hosted in the cloud. [Online] Available: <https://docs.microsoft.com/en-us/azure/iot-fundamentals/iot-introduction>. (Accessed 4May2022).
8. The Microsoft solution to advanced analytics and transformation of data into intelligent action is Cortana Intelligence Suite. Raw data can originate from several sources, including databases, apps, sensors, devices and IoT systems. Various transformations, analytics and machine learning processing can then provide insights, which is presented or delivered to people, apps, and automated systems. [Online] Available: <https://social.technet.microsoft.com/wiki/contents/articles/36688.introduction-to-cortana-intelligence-suite.aspx> (Accessed 4 May 2022).
9. Skywise is an open data platform developed by Airbus for the aviation industry. By making the right information available at the right time, Skywise provides invaluable insights from the massive amounts of data that was previously locked in corporate and functional silos. [Online] Available: <https://aircraft.airbus.com/en/services/enhance/skywise> (Accessed 4 May 2022).

References

1. Guidance for Air Travel through the COVID-19 Public Health Crisis. Int Civil Aviation Organis. <https://www.icao.int/covid/cart/Pages/CART-Take-off.aspx>. Accessed 18 July 2021
2. Human factors in aviation maintenance technician handbook—General. Federal Aviation Administration. file:///C:/Users/User/AppData/Local/Temp/amt_general_handbook.pdf. Accessed 18 July 2021
3. Driskill M (2020) Cathay pacific to cut 6,000 workers and kill off Dragon affiliate to survive COVID-19 pandemic. Asian Aviation. <https://asianaviation.com/report-cathay-pacific-to-cut-6000-workers-and-kill-off-dragon-affiliate>. Accessed 18 July 2021

4. Wyman O (2020) COVID-19 Impact on commercial aviation maintenance. <https://www.oliverwyman.com/our-expertise/insights/2020/mar/COVID-19-Impact-On-Commercial-Aviation-Maintenance.html>. Accessed 18 July 2021
5. Smart glasses in aid of maintenance. The Agility Effect. <https://www.theagilityeffect.com/en/article/smart-glasses-aid-maintenance/>. Accessed 28 June 2021
6. Crescenzo FD, Fantini M, Persiani F, Stefano LD, Azzari P, Salti S (2010) Augmented reality for aircraft maintenance training and operations support. *IEEE Comput Graph Appl* 31(1):96–101
7. Big Data in the hangar: virtual damage assessment & repair tracking for F-35 and F-22 aircraft. Industrial Internet in Action Case Study. Industrial Internet Consortium. file:///C:/Users/User/AppData/Local/Temp/NGRAIN_Lockheed_Martin_case_study-1.pdf. Accessed 18 July 2021
8. RAAF uses HoloLens mixed-reality device for C-17A maintenance. Airforce-Technology. <https://www.airforce-technology.com/news/raaf-uses-hololens-mixed-reality-device-for-c-17a-maintenance/>. Accessed 28 June 2021
9. Po KK, Wong ET (2020) A DD-SHELL HF model for bus accidents. In: Pham H (ed) Reliability and statistical computing: modeling, methods and applications. Springer, pp 151–166. <https://doi.org/10.1007/978-3-030-43412-0>
10. The Human Factors “Dirty Dozen”.. Skybrary. https://www.skybrary.aero/index.php/The_Human_Factors_%22Dirty_Dozen%22. Accessed 18 July 2021
11. International Civil Aviation Organisation (2018) The safety management manual (Doc 9859), 4th edn
12. UK Civil Aviation Authority (2002) CAP 718: human factors in aircraft maintenance and inspection, 1st edn. The Stationery Office (UK)
13. International Civil Aviation Organisation (1993) Human factors digest no 7: investigation of human factors in accidents and incidents, 1st edn
14. Man WY, Wong ETT (2020) Developing alert level for aircraft components. In: Pham H (ed) Reliability and statistical computing: modeling, methods and applications. Springer, pp 85–106. <https://doi.org/10.1007/978-3-030-43412-0>
15. International Air Transport Association (2020) Guidance for the transport of cargo and mail on aircraft configured for the carriage of passengers, 3rd edn

Eric T. T. Wong The Hong Kong Polytechnic University, Hong Kong SAR

Dr. Eric T. T. Wong received an M.Sc. degree in Plant Engineering from the Loughborough University (UK) and a Ph.D. degree from the Leicester University (UK). Prior to joining the Mechanical Engineering Department of the Hong Kong Polytechnic University, he worked for the Hong Kong Civil Aviation Department in air traffic control, aircrew licensing, and examinations. His research interests include Information Communication Technology, aviation human factors, laser surface treatment of metals, risk modeling, and quality management. He was the Program Chair of the ISSAT International Conference on Modeling of Complex Systems and Environments (2007) in Ho Chi Minh City, Vietnam. He is currently an Adjunct Associate Professor at the Hong Kong Polytechnic University and an Associate Editor of the International Journal of Industrial Engineering—Theory, Application, and Practice. He is a Fellow Member of the UK Royal Aeronautical Society and a Spokesman for aviation safety of the Royal Aeronautical Society (Hong Kong Branch). He has achieved over 50 years of membership in the Institution of Mechanical Engineers. He was a Founding Member of the Aircraft Engineering Discipline (Hong Kong Institution of Engineers) and has been appointed as an Honorary Advisor of the China Aircraft Services Ltd. in Hong Kong.

Wai Yeung Man Heliservices (HK) Ltd., Hong Kong

Wai Yeung Man, Simon received a Bachelor's degree in Electronic Engineering from The Chinese University of Hong Kong in 2011 and a Master of Science degree in Mechanical Engineering (Aviation) from The Hong Kong Polytechnic University in 2021. Simon is particularly interested in aviation maintenance, engineering, and human factors studies. He is currently a licensed avionics engineer at Heliservices HK Ltd. He is specialized in helicopter avionics and electrical systems inspection, troubleshooting, fault rectification, and system modifications. He holds a Hong Kong Civil Aviation Department helicopter airframe and engine license.

Feedback-Based Algorithm for Negotiating Human Preferences and Making Risk Assessment Decisions



Silvia Carpitella, Antonella Certa, and Joaquín Izquierdo

Abstract Work equipment risk assessment is essential for guaranteeing health and safety of workers in industrial contexts. Many and varied hazards are involved in the use of equipment, which have to be periodically subject to thorough controls required by law. This research proposes a novel hybrid decision-making framework aimed at integrating flexible negotiation on human preferences. This goal will be achieved by establishing effective feedback exchanges with expert(s) familiar with the field of risk and maintenance of work equipment. We extend a previous research that proposed a user-friendly negotiation procedure to increase consistency of judgments provided by experts about relevant risk factors. The proposed algorithm has been built within the framework of the Analytic Hierarchy Process (AHP), a widely popular Multi-Criteria Decision-Making (MCDM) method. After negotiation of the evaluations of preference with experts to weight the main risk factors of interest, the obtained priorities will be used as a part of the body of input data required for a further MCDM application. This application will use the EElimination Et Choix Traduisant la REalité (ELECTRE) I method as a structured way for planning and implementing maintenance interventions. A real world application will lead towards the selection of the work equipment with associated higher level of risk under diverse risk factors, differently weighted by means of the negotiation process. Apart from the industrial field of reference, our theoretical framework can be applied to solve a wide range of practical decision-making problem.

Keywords Multi-criteria decision-making · Expert feedback · Consistency improvement · Work equipment risk assessment

S. Carpitella (✉)

Department of Manufacturing Systems Engineering and Management, California State University, Northridge, USA

e-mail: silvia.carpitella@csun.edu

A. Certa

Department of Engineering, University of Palermo, Palermo, Italy

e-mail: antonella.certa@unipa.it

J. Izquierdo

Institute for Multidisciplinary Mathematics, Universitat Politècnica de València, Valencia, Spain

e-mail: jizquier@upv.es

1 Introduction

Risk assessment related to the use of work equipment is a process of utmost importance in industry, controlled by the existing national and international regulations in force. Work equipment is defined as any machine, tool or plant—considered as a set of machines—equipment and components necessary for implementing industrial processes. Any work operation related to the use of work equipment must be subject to risk assessment. We may consider, for example, commissioning or decommissioning activities, general use, transport, repair, conversion, maintenance, cleaning operations, assembly, disassembly, and so on. It is also important to evaluate dangerous areas, that are those physical spaces dedicated to work equipment in which the presence of workers constitutes a risk for their health and safety. The employer is obliged to guarantee that work equipment available for workers meet specific compliance requirements. In this context, accurate equipment risk assessment is crucial to understand if replacement and/or repair operations are necessary. When it comes to maintenance, it is also important to establish techniques aimed at understanding how to prioritise interventions on the basis of the risk evaluation carried out according to standardized parameters.

Technical expert assistance is indispensable to make effective decisions on such relevant industrial issues. The present research proposes a novel methodology integrated with Multi-Criteria Decision-Making (MCDM) aimed at establishing effective feedback relations with experts to globally increase the mathematical consistency of their opinions, where necessary. In the decision-making practice, experts are frequently asked to express judgments of preference by comparing pairs of elements. Such a kind of evaluation may lead to lack of consistency when, for instance, the number of elements to be simultaneously compared is large. To achieve scientifically significant results in the field of industrial work equipment risk assessment, we expand a previous conference contribution [8] which developed a flexible tool making use of an algebraic consistency-improving algorithm along with a sensitivity analysis technique to identify which judgment(s), among those expressed by the involved expert(s), contribute most to inconsistency. The proposed framework was applied for weighting the main risk factors influencing equipment assessment by asking decision makers to reconsider only a few *a priori* judgments, instead of rethinking the entire set of previously elicited comparisons. Here, we propose a calculation procedure to achieve a final “adjusted” consistent matrix reflecting as much as possible the input evaluations. Also, we integrate the developed negotiation process as a part of a hybrid MCDM approach to select the work equipment corresponding to the higher level of risk bases of the previously weighted factors. This will enable analysts to formalise important decisions about scheduling and implementing maintenance interventions. The process is herein developed for making relevant decisions about risk assessment. However, our algorithm is extremely flexible and its application can be easily extended to many contexts, both in business and, in general, in real life situations.

Contents are organised as follows. Section 2 develops a literature review on the topics of research. The framework developed in [8] to get the weights of risk factors by means of the negotiation process with the experts is extended in Sect. 3. As already said, the proposed process has been enriched with a new calculation for the final consistent adjusted matrix, to make it even more adherent to the practical perception of decision makers. Section 4 describes the MCDM approach for work equipment selection, on the basis of the weights previously calculated. A real case study is provided in Sect. 5, further expanding the list of significant sub-factors with respect to those we considered in [8]. This is done to make even more exhaustive and reliable the whole stage of risk assessment. Section 6 lastly closes the work.

2 Literature Review

2.1 Consistency and Feedback Exchange in Decision-Making

Various MCDM methods rely on pairwise comparisons (PCs) elicited by a single expert or a panel of decision makers [33]. As underlined in [37], PC-based methods are very popular among decision makers. PCs are implemented by comparing decision-making elements in pairs to establish a relation of preference by evaluating which one of them is preferred over another. PCs have been widely used in various areas of knowledge such as, for instance, medicine, biology, engineering, education, and so on [23]. PC techniques analyse a pair of elements at a time and, once collected the whole body of judgments, judgments are reported in so-called Pairwise Comparison Matrices (PCMs). However, these methods can lead to logically contradictory outputs [19] resulting in lack of mathematical consistency, a property that is possible to check for PCMs.

In this context, mentioning the Analytic Hierarchy Process (AHP) [30] is certainly necessary. AHP is one of the most common MCDM methods, aimed at ranking a set of decision-making elements (i.e. alternatives, criteria and/or sub-criteria) by calculating the related vector of priorities via the Perron eigenvector of PCMs. A plethora of prioritization methods is proposed in the literature [40]. To make decisions reliable and appropriate, pairwise comparisons are required to be acceptably consistent. However, as already said, consistency is a property that not always PCMs have. Humans are indeed not fully coherent when eliciting judgments, due to their natural cognitive limitations. It goes without saying that the higher the number of elements to be pairwise compared, the lower the possibilities for consistency conditions to be met.

Since information gathered from big databases and the Internet is also susceptible to being handled through PCs, the number of elements taken into account can be huge (see [5, 17, 22], among others). On the whole, one of the problems limiting the applicability of pairwise comparisons to large-scale decision problems is the so-called curse of dimensionality. This means that many comparisons have to be issued

or built from a body of information. As a result, PCMs related to real decision-making problems may suffer lack of consistency, and this may have a negative impact on the quality of the final decisions. Keeping in mind that making the right decision is a matter of compromise given the various criteria, often conflicting with each other [24], the final decision may be sub-optimal and possibly quite far from the optimal solution when theoretical assumptions are not met in the practice. With this regard, consistency is a crucial issue in decision-making and among the basic assumptions for making efficient decisions, since it would be unwise to rely on randomly elicited judgments. Inconsistency is then typically explained as a misrepresentation caused by one or more decision makers. When the consistency of a PCM is not satisfactory, it is necessary to improve it [12].

Suitable mechanisms aimed at improving consistency can be implemented and adapted to the problem under analysis [13], but they require the alteration of one or more judgments previously elicited by decision makers. Consistency can be enhanced through various, in particular, as in this paper, projection techniques on vector spaces. However, one has to consider the fundamental importance of sharing any change with the decision makers, since they may not agree with some of the performed changes. Consistent pairwise comparisons may indeed lead to ineffective results under a practical point of view when the evaluations of the intensity of preference are not well matched to the qualitative preference information. In this regard, inconsistency issues have been explored in the literature for some special cases of PCMs [10]. In any case, carrying out a negotiation process capable of balancing mathematical consistency enforcement with the issued judgments reflecting the (expert) reality is essential. It is now clear as decision-making itself is a complex process for various reasons. Among them, we particularly consider the aspects of subjectivity and negotiation. The ability to make a rational use of subjective perceptions of experts is indeed crucial and, in this context, fruitful feedback exchanges [39] ensure the facilitation of discussion and deliberation even among group members to share opinions and eventually agree on a final decision [26].

2.2 MCDM Methods to Support the Problem of Research

Effective equipment maintenance as well as activities integrating workers, resources and business processes concur to optimise productivity of enterprises while pursuing cost effectiveness [6]. With this recognition, not only is the process of risk assessment in industrial workplaces a legal (and a moral) obligation of employers, but also it significantly influences organizational performance [34]. Risk assessment can be extremely complex, especially in dynamic environments [14] and, in general, when dealing with scenarios in which risks may significantly differ from the ones managers are conventionally familiar with [38].

The existing literature shows as MCDM-based approaches can be particularly useful to promote best practices in industry integrating risk assessment [27] and maintenance optimisation [36]. The AHP and the fuzzy AHP (FAHP) have been

largely applied in the field of research [7], and also combined with other MCDM techniques such as the Analytic Network Process (ANP) [11], the Decision making trial and evaluation laboratory (DEMATEL) [18], the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [11], among others. The ANP has been integrated with multi-criteria approaches making use of outranking relations for creating a link between the risk assessment process for complex systems and the choice of the best maintenance policy [9]. In this last case, for example, the usefulness of applying MCDM methodologies belonging to the ELimination Et Choix Traduisant la REalité (ELECTRE) category has been discussed and demonstrated.

ELECTRE methods are quite popular techniques in the MCDM practice of risk management, having been widely used in real life problems [21]. They have been largely improved and supported through mathematical tools aimed at managing uncertainty, with the objective of dealing with those practical cases in which experts are not capable to assess alternatives by providing precise quantitative values [20]. As asserted by Govindan and Jepsen [16], ELECTRE I is suitable for solving the so called *problematic choice*, whose goal consists in selecting just one alternative within a wider set of options. The chosen solution would correspond to the best trade-off under the considered criteria, differently weighted according to experts' perception. ELECTRE I can be then used as a screening method to reduce the set of alternatives [15], even if results coming from the application of different decision techniques may differ when applied to the same problem [25].

3 New Developments Within the AHP Framework

The mathematical tool proposed in [8] and substantially extended in the present research uses a sensitivity analysis of an input inconsistent PCM, by highlighting which judgment(s) previously expressed by the expert should be preferably modified to increase consistency. The process has been herein designed to produce an adjusted PCM of output—as much adherent as possible to the input matrix—by keeping consistency within an acceptable threshold [30]. After setting the problem, the negotiation algorithm is presented next. The process will be first applied to derive the weights of the main factors involved in work equipment risk assessment and, successively, integrated with the ELECTRE I method to highlight the equipment to be maintained with priority. Once again, we claim that the developed algorithm is extremely flexible and can be applied to solve many decision-making problems in many fields of interest.

3.1 Problem Setting

In AHP, pairwise comparisons are formally collected in a PCM by asking to express judgments of preference (typically by using the Saaty's 9-point scale [30]) between

pairs of elements. For a set of n elements, a PCM is an $n \times n$ matrix, $A = (a_{ij})$, whose (positive) entries have to adhere to the properties of homogeneity ($a_{ii} = 1$) and reciprocity ($a_{ji} = 1/a_{ij}$), $i, j = 1, \dots, n$. The problem for matrix A consists in producing a set of numerical values w_1, \dots, w_n expressing the importance of the n elements under analysis. If the expert gives his/her judgments in a completely consistent way, the relations between weights w_i and numerical values translating judgments a_{ij} are simply given by $w_i/w_j = a_{ij}$ ($i, j = 1, \dots, n$), and the PCM A is said to be consistent. For a consistent PCM, the Perron eigenvector (normalized to one) gives the vector of weights [31] of the analysed decision-making elements.

However, considering the limits of human reasoning, a certain degree of inconsistency is always expected. For a non consistent PCM, one has to solve the eigenvector problem:

$$A\mathbf{w} = \lambda_{\max}\mathbf{w}, \quad (1)$$

where λ_{\max} is the principal eigenvalue of matrix A , providing the vector of weights, \mathbf{w} . The so-called Consistency Index

$$CI = \frac{\lambda_{\max} - n}{n - 1}, \quad (2)$$

is used as a common measure of inconsistency and is used to calculate the Consistency Ratio

$$CR = \frac{CI}{RI}, \quad (3)$$

RI being the average consistency index [32]. Consistency is accepted if $CR < 0.1$, otherwise, new judgments should be elicited again until this condition is met.

The literature offers various methodologies aimed at improving consistency, mainly based on iterative optimisation techniques. We herein use a direct, algebraic method, the linearisation technique [3], to get the (fully) consistent matrix A^c closest to a non-consistent PCM A . A simple formula provides it [4] in terms of an algebraic projection p_n of the (entry-wise) logarithm, L , of A :

$$p_n(L(A)) = \frac{1}{n}[(L(A)U_n) - (L(A)U_n)^T], \quad (4)$$

where $U_n = \mathbf{1}_n \mathbf{1}_n^T$ and $\mathbf{1}_n$ is the unit column vector. The (entry-wise) exponential, E , of this projection provides the sought matrix.

We underline as the closest consistent matrix is a synthetic result that may not properly reflect the original opinions provided by the involved expert(s). This is the reason why the process of feedback exchange is fundamental at any time, to assure that the obtained results are not far from the expert reality. In other terms, the closest consistent matrix has to be evaluated by the expert, who will have to adjust its entries where necessary, and this process will be iterated until producing a final matrix with acceptable consistency, representing a reasonable compromise between synthetic consistency and expert judgment. A basic procedure that presents the entire

fully consistent matrix is provided in [2]. Next, we formally develop a more friendly process that simplifies this modifying procedure.

3.2 Negotiation Algorithm for Feedback Exchange with Decision Makers

Figure 1 presents a scheme of the conciliation process between consistency and expert judgments. In this flowchart, trapezoids represent inputs, rectangles correspond to procedures, rhombuses stand for if-else decisions and rounded figures for outputs.

Detailed explanations are now provided.

- Obtaining pairwise comparisons: the elicitation of pairwise comparison judgments (upper left trapezoid) represents the input of the process and initiates the diagram.
- Drawing up the input matrix: judgments of preference are collected into the PCM by using, for instance, the scale proposed in [30] (first upper left rectangle in Fig. 1).
- Checking consistency: this step (second rectangle) uses Eqs. (1), (2) and (3) to calculate CR .
- Establishing consistency acceptability: in this case (lower leg of the rhombus), the PCM is validated; otherwise, a negotiation process will be undertaken with the decision maker, including possible modifications to get a consistency ratio within the acceptable consistency threshold (rhombus's leg 'No'). This process is described next.

The negotiation process is represented in Fig. 2: for a non acceptably consistent PCM (upper rhombus's 'No' leg), this process consists of two main steps.

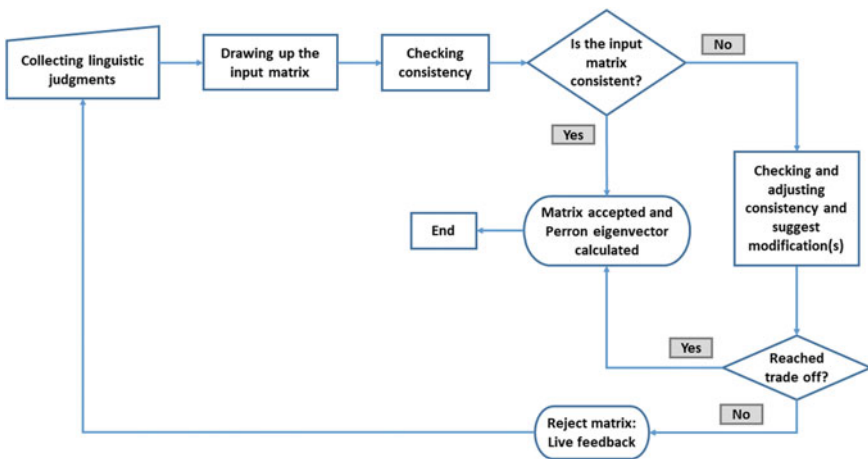


Fig. 1 Conciliation of consistency and expert judgment [8]

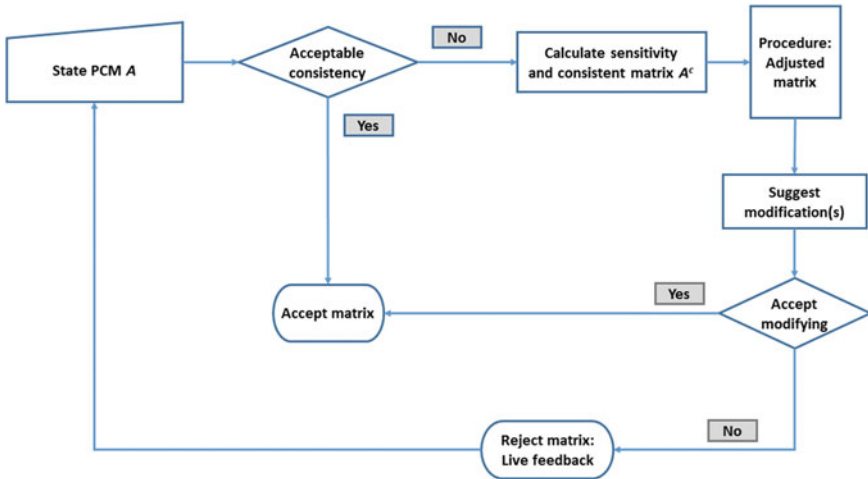


Fig. 2 Process of “checking and adjusting consistency” [8]

- Calculation of sensitivity and closest fully consistent matrix. It includes two calculations:
 - obtaining the closest fully consistent matrix A^c by applying the linearisation process using (4);
 - calculating sensitivity values for the entries of A as described next. Note that just the $M = \frac{n \times (n-1)}{2}$ values over the main diagonal are considered. The sensitivity method applied to rank the most “influencing” judgments is now succinctly presented. Starting from an $n \times n$ PCM A , the method calculates a second matrix D giving the partial derivatives of λ_{\max} with respect to the entries of A , thus identifying entries’ consistency sensitivity. These partial derivatives are given by the formula (Sect. 1.1.1, [35]):

$$D = \mathbf{w}\mathbf{v}^T - A * A * \mathbf{v}\mathbf{w}^T, \tag{5}$$

where: \mathbf{w} is the right Perron eigenvector (normalized to one), associated to λ_{\max} ; \mathbf{v} is the left Perron eigenvector of A , also associated to λ_{\max} , and normalized such that $\mathbf{w}\mathbf{v}^T = 1$; $*$ is the Hadamard (entry-wise) product. The Hadamard product operates on identically-shaped matrices and produces a third matrix of the same dimensions, whose elements (i, j) correspond to the entry-wise product of elements (i, j) of the matrices. The values corresponding to the partial derivatives allow to rank the corresponding entries of matrix A and then to know which one has higher influence on consistency.

- Calculation of the adjusted matrix X . Once the matrix of partial derivatives has been calculated, the process to get the adjusted matrix is given next. This process is intended to make the adjusted matrix as adherent as possible to the input matrix,

by changing as fewer as possible evaluations previously given by the experts. The calculation of the adjusted matrix X involves the input matrix A , the closest consistent matrix to the input matrix A^c and the matrix of partial derivatives D calculated in the previous step, and provides just a convex combination of A and A^c using the directions shown by the gradient given by matrix D . The formula for the calculation is as follows:

$$X = A + \theta \cdot D * |A^c - A|, \quad (6)$$

where θ is the parameter for the convex combination varying between 0 and 1, which has to be suitably fine-tuned. In effect, this factor influences the final calculation of the adjusted matrix and should be properly calibrated so as to guarantee minimum requirements of consistency. On the one side, if the factor is too high, matrix X will tend more to the synthetic matrix A^c that, despite being perfectly consistent, may significantly diverge from the input matrix A elaborated by the expert. On the other side, if the factor is too small, the adjusted matrix X will be more similar to the input matrix, but probably with not significant consistency improvement. Although a minimization process would be possible to get the optimum value for the factor, by making a reduced number of trials it is easy and fast to reach the ideal trade off between matching minimum consistency requirements ($CR \leq 10\%$) and calculating an adjusted matrix close to the input matrix as much as possible. It goes without saying that we expect major changes in correspondence to those judgments with associated higher absolute values in matrix D .

- Validation of would-be modifications by the experts. The modified judgments will be eventually proposed to the expert, who will be invited to agree with the final evaluations. In case of disagreement ('No' leg of lower rhombus in Fig. 2), he/she will be asked to elicit new evaluations, so that a new matrix may be drawn up. At the point in which a decision maker has associated a final adjusted consistent matrix, the process goes back to the central rounded box of Fig. 2, and then through the 'Yes' leg of the lower right rhombus in Fig. 1 leads to its rounded central box, where the matrix is accepted and the process stops after calculating the sought vector of priorities.

4 Hybrid MCDM Perspective

The ELECTRE I method is now proposed to be integrated with the mathematical algorithm carrying out the negotiation process previously discussed. This methodological combination aims to effectively manage the problem of work equipment risk assessment. Specifically, after deriving risk factor priorities by negotiating PCs evaluations with expert(s), the ELECTRE I selects the equipment showing an association with higher risk conditions for workers. This will correspond to the specific industrial equipment for which maintenance interventions should be carefully planned and implemented with priority with respect to other equipment. We remind that priorities

Table 1 ELECTRE methods and their objectives

Method	Objective
ELECTRE I	Selecting the alternative representing the best trade-off under diverse criteria
ELECTRE II ELECTRE III ELECTRE IV	Ranking alternatives to establish a priority index or to draw up a list of candidates
ELECTRE TRI	Sorting alternatives into different classes on the basis of their common features

calculated during the previous stage will represent one of the necessary input data to apply the ELECTRE I.

ELECTRE I belongs to the ELECTRE category of MCDM methods, originally born in France at the end of the 1960s [28]. As already mentioned, ELECTRE methods are grounded upon the so called outranking approach [29]. These techniques aim to establish outranking relations by pairwise comparing alternatives, relations that need to be examined and confirmed by performing two tests in sequence, namely the concordance and the discordance tests. Concordance and discordance tests are respectively aimed at calculating the concordance and discordance indices. The concordance index C_{ij} quantitatively expresses, referring to a specific criterion, the agreement degree about the fact that alternative A_i outranks or equals alternative A_j . The discordance index D_{ij} quantitatively expresses, referring to a specific criterion, the agreement degree about the fact that alternative A_i has a worst score compared to alternative A_j . Different versions of ELECTRE methods have been proposed and developed over the years. The most common ones are described in Table 1 along with the relative objectives.

Integrating the ELECTRE I technique within the negotiation process built within the AHP framework may constitute a reliable methodological support towards the optimisation of work equipment management in industry. Such a hybrid perspective can be of interest for supporting a wider range of decision-making problems in general business and real life situations. The methodological steps necessary to implement the ELECTRE I procedure are described next.

4.1 The ELECTRE I Method for Work Equipment Selection

The ELECTRE I method requires the preliminary collection of the following input data:

1. set of alternatives A_i to be evaluated;
2. evaluation criteria B_k ;
3. vector of criteria weights w_k ;
4. numerical evaluation of alternatives respect to the considered criteria, $u_k(A_i)$.

Table 2 Application of the ELECTRE I method

Phase	Description	Step
I PHASE	Developing an outranking relation characterizing the pair Wise comparisons between alternatives and accomplishing The concordance and discordance tests	1. Correlation sets matrix 2. Correlation indices matrix 3. Correlation test matrix 4. Discordance sets matrix 5. Discordance indices matrix 6. Not-discordance test matrix
II PHASE	Exploiting the outranking relation in order to obtain the final result by means of a specific rule	1. Outranking matrix

Once collected and normalised this input data in a proper input matrix, the procedure, whose development is summarised in Table 2, can be initialised.

The first phase is made of six steps. The first group of three steps refer to the accomplishment of the concordance test whereas the second group of three steps refer to the not-discordance test. All the six steps are detailed next.

- 1st STEP: building the correlation sets matrix. The correlation sets are determined by pairwise comparing alternatives. These sets correspond to:

$$J(A_i, A_j) = J^+(A_i, A_j) \cup J^-(A_i, A_j); \quad (7)$$

where $J^+(A_i, A_j)$ represents the sets of criteria for which the first alternative A_i has a higher evaluation with respect to the second one A_j , and $J^-(A_i, A_j)$ represents the sets of criteria for which the first alternative A_i has an equal evaluation respect to the second one A_j . The output of the first step will be a squared matrix in which correlation sets are compiled for each pairwise comparison.

- 2nd STEP: building the correlation indices matrix. The correlation indices correspond to:

$$C_{ij} = \frac{w^+(A_i, A_j) + w^-(A_i, A_j)}{\sum_k w_k}; \quad (8)$$

where $w^+(A_i, A_j)$ corresponds to the sum of criteria weights belonging to the set $J^+(A_i, A_j)$, and $w^-(A_i, A_j)$ corresponds to the sum of criteria weights belonging to the set $J^-(A_i, A_j)$. The denominator $\sum_k w_k$ corresponds to the sum of all criteria weights (which is equal to 1). The output of the second step is a squared matrix in which correlation indices are given for each pairwise comparison.

- 3rd STEP: building the correlation test matrix, which is the following Boolean matrix:

$$T_C(A_i, A_j) = \begin{cases} 1 & \text{if } C_{ij} \geq C^* \\ 0 & \text{if } C_{ij} < C^* \end{cases} \quad (9)$$

where C^* represents a threshold arbitrarily fixed to establish if an alternative actually outranks the other one. The threshold should be fixed within the range $[0.5; (1 - \min C_{ij})]$. The output of the third step will be a squared Boolean matrix in which the 1's indicate the outranking relations that have passed the concordance test.

- 4th STEP: building the discordance sets matrix. After having analysed concordance principles, it is also necessary to consider the discordance ones by taking into account discordant criteria. The discordance sets are determined by pairwise comparing alternatives and correspond to $J^-(A_i, A_j)$. In particular, $J^-(A_i, A_j)$ represents the sets of criteria for which alternative A_i has a lower evaluation with respect to alternative A_j . The output of the fourth step will be a squared matrix in which discordance sets are given for each pairwise comparison.
- 5th STEP: building the discordance indices matrix. The discordance indices are calculated by pairwise comparing alternatives:

$$D_{ij} = \begin{cases} 0 & \text{if } J^-(A_i, A_j) = \emptyset \\ \max[u_k(A_j) - u_k(A_i)] & \text{with } k \in J^-(A_i, A_j) \end{cases} \quad (10)$$

where $u_k(A_j)$ and $u_k(A_i)$ are the evaluations of the two alternatives under the criterion B_k belonging to the discordance set $J^-(A_i, A_j)$. The output of the fifth step will be a squared matrix in which discordance indices are provided for each pairwise comparison.

- 6th STEP: building the not-discordance test matrix, which is the following Boolean matrix:

$$T_D(A_i, A_j) = \begin{cases} 1 & \text{if } D_{ij} \leq D^* \\ 0 & \text{if } D_{ij} > D^* \end{cases} \quad (11)$$

where D^* represents an arbitrarily fixed threshold, expressing the maximum accepted value of discordance to not deny the hypothesis that A_i outranks A_j . The output of the sixth step will be a squared Boolean matrix in which the 1s indicate the outranking relations that have passed the not-discordance test.

Both the concordance test and not-discordance test matrices constitute the main outputs of the I first phase and the inputs of the second phase of the ELECTRE I method. The second phase is carried out by means of the following rules: A_i outranks A_j if both the concordance and the not-discordance tests are passed. The following outranking matrix has to be built:

$$S(A_i, A_j) = \begin{cases} 1 & \text{if } C_{ij} \geq C^* \wedge D_{ij} \leq D^* \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

The best alternative will be eventually determined by observing the outranking matrix. Specifically, this alternative outranks all the other options and is not outranked by anyone.

The case study proposed in the next section demonstrates the applicability of the presented methodology and its effectiveness to solve real problems in industry.

5 Industrial Case Study

The present case study focuses on a manufacturing company operating in the food industrial sector and located in the South of Italy. The objective of study consists in carrying out an effective work equipment risk assessment by selecting the equipment in major need of maintenance according to suitable factors, differently weighted. Such a selection aims to optimise the level of safety and security within the company while pursuing an effective organisation of maintenance interventions. This last aspect will have the double purpose to both minimising the risk of equipment shut down and allocating economic resources for maintenance over a given time horizon.

Apart from the mentioned reasons, compliance of work equipment with safety requirements is a crucial aspect that must be verified according to the existing standards. In addition, the employer is required by law to carry out an exhaustive assessment aimed at formally and quantitatively evaluating all the potential risks for employees working with the equipment. Risk assessment has to thoroughly analyse, for each working station, dangerous situations potentially occurring during tasks execution. It is also necessary to highlight which workers may be critically exposed to risks in the area where a given equipment operates.

We then proceed by first applying our new algorithm for weighting those factors that are relevant for the risk assessment process. Secondly, the selection of the most critical work equipment will be performed by means of ELECTRE I. We specify that we are going to find a good solution for our problem, that is not the optimal alternative in an absolute sense. The final choice will certainly depend on experts' experience and personal perception. It means that different stakeholders may have even significantly different opinions, potentially leading to divergent results for different applications of the same procedure. This is the reason why we underline the importance of paying particular attention to the selection of the expert to be involved. Not only will the decision maker have to show proven experience in the practical field of reference, but also he/she will have to be as much as honest as possible about the issues under analysis, without omitting details to the best of his/her knowledge and trying to improve critical aspects.

After a careful evaluation, we have decided to involve the expert in charge of the whole maintenance function. Having been working for the company for decades, this expert knows with plenty of details features of work equipment as well as all the different stages of maintenance interventions in terms of practical execution. He has also been trained to promptly react in the case of any emergency should involve one or more critical equipment. With these preliminaries, we have organised several one-to-one brainstorming sessions with our expert for carrying out the evaluation on the whole. The first step has been collecting evaluations of pairwise comparisons about risk factors, that have been chosen according to standard procedures, as detailed next.

5.1 Calculating Risk Factors Weights Via the Negotiation Process

The quantitative risk evaluation procedure for work equipment applied by the company is described in Appendix A of document [1]. This document considers the development of the procedure according to three main factors: material (M), environment (E), organisation (O). The first factor refers to risks related to physical work equipment and their use by workers. The second factor considers the specific features of workplaces where equipment (and consequentially workers) operate. Lastly, the third factor is connected to the ability of personnel to master risk occurrence, especially in terms of expertise and organisational issues. We can observe as all the relevant aspects for risk analysis are taken into account, including human factors, something that is not still exhaustively considered in the existing literature, despite clearly impacting safety and security at a general level. In any case, it is clear that the influence of the mentioned aspects on the process can be different. This is the reason why the management of the company has decided to upgrade the process of risk assessment for some of its core systems. The objective of the first stage of analysis consists indeed in supporting the company by calculating the degrees of importance associated to the three main risk factors on the basis of judgments of preference provided by the involved expert. In this way, the risk assessment will be more precise, by considering the diverse contribution of the three factors instead of assuming them as equally weighted.

Each risk factor is split into various sub-factors [1]. In Table 3 we give and synthetically describe those most significant sub-factors derived after a screening process. The reported list extends the list of sub-factors analysed in [8].

The nine sub-factors have been pairwise compared by the expert, and linguistic relations of preference have been numerically translated via the Saaty scale [30]. Specifically, numerical evaluations corresponding to the preference evaluations expressed by the maintenance expert according to the mentioned Saaty scale have been then collected in the input PCM, A (Table 4).

Matrix A , having a $CR = 11.5 > 10\%$, is not acceptably consistent, so that some judgments need to be adjusted. To such an aim, we apply the negotiation algorithm to calculate an adjusted matrix acceptably consistent and adherent as much as possible to the original input matrix. We first calculate the closest consistent matrix A^c as a synthetic result by using (4). This matrix is given in Table 5. It can be easily noticed as the calculated evaluations significantly differ from the original ones.

The matrix of partial derivatives of λ_{\max} with respect to A 's entries, D , is calculated by (5) and provided in Table 6. This matrix gives us an actual idea about those judgments having a higher impact on (in)consistency, which are those with higher absolute values.

We can observe as pairwise comparison corresponding to entry (E_2, E_3) influences consistency most, having associated the highest absolute value, followed by comparison (M_1, O_2) , then (O_2, E_2) , (O_1, E_2) , and so on. The process of negotiation is initialised by calculating the adjusted matrix X , obtained via (6) and provided in

Table 3 Factors and sub-factor for risk assessment

Factor	Sub-factor	Description
<i>M</i>	<i>M</i> ₁	Dangerous events and potential injuries derived from the use of work equipment
	<i>M</i> ₂	Frequency and duration of exposition of workers during equipment functioning
	<i>M</i> ₃	Difficulty of preventing dangerous event or limiting the related injures
<i>E</i>	<i>E</i> ₁	Hazardous events related to the particular workplace physical location
	<i>E</i> ₂	Inadequate work conditions determined by poor lighting or excessive noise
	<i>E</i> ₃	Possible overstress sources determined by such factors as microclimate
<i>O</i>	<i>O</i> ₁	Scarce personnel qualification potentially leading to improper equipment use
	<i>O</i> ₂	Inadequate physiological factors connected to hazardous work conditions
	<i>O</i> ₃	Poor workflow organisation and scarce communication with the management

Table 4 Input PCM *A*

<i>A</i>	<i>M</i> ₁	<i>M</i> ₂	<i>M</i> ₃	<i>E</i> ₁	<i>E</i> ₂	<i>E</i> ₃	<i>O</i> ₁	<i>O</i> ₂	<i>O</i> ₃
<i>M</i> ₁	1.00	1.00	1.00	1.00	2.00	3.00	3.00	5.00	4.00
<i>M</i> ₂	1.00	1.00	2.00	3.00	4.00	2.00	4.00	4.00	1.00
<i>M</i> ₃	1.00	0.50	1.00	0.50	2.00	2.00	2.00	2.00	2.00
<i>E</i> ₁	1.00	0.33	2.00	1.00	0.50	0.50	1.00	1.00	1.00
<i>E</i> ₂	0.50	0.25	0.50	2.00	1.00	3.00	0.33	0.33	0.50
<i>E</i> ₃	0.33	0.50	0.50	2.00	0.33	1.00	1.00	1.00	1.00
<i>O</i> ₁	0.33	0.25	0.50	1.00	3.00	1.00	1.00	2.00	0.50
<i>O</i> ₂	0.20	0.25	0.50	1.00	3.00	1.00	0.50	1.00	2.00
<i>O</i> ₃	0.25	1.00	0.50	1.00	2.00	1.00	2.00	0.50	1.00

Table 5 Closest consistent matrix A^c

A^c	M_1	M_2	M_3	E_1	E_2	E_3	O_1	O_2	O_3
M_1	1.000	0.919	1.526	2.347	2.957	2.652	2.423	2.565	2.244
M_2	1.088	1.000	1.661	2.553	3.217	2.884	2.636	2.790	2.441
M_3	0.655	0.602	1.000	1.537	1.937	1.737	1.587	1.680	1.470
E_1	0.426	0.392	0.650	1.000	1.260	1.130	1.032	1.093	0.956
E_2	0.338	0.311	0.516	0.794	1.000	0.897	0.819	0.867	0.759
E_3	0.377	0.347	0.576	0.885	1.115	1.000	0.914	0.967	0.846
O_1	0.413	0.379	0.630	0.969	1.220	1.094	1.000	1.058	0.926
O_2	0.390	0.358	0.595	0.915	1.153	1.034	0.945	1.000	0.875
O_3	0.446	0.410	0.680	1.046	1.318	1.192	1.080	1.143	1.000

Table 6 Matrix D of partial derivatives of λ_{\max} w.r.t. a_{ij}

D	M_1	M_2	M_3	E_1	E_2	E_3	O_1	O_2	O_3
M_1	0.000	-0.018	0.095	0.216	0.172	-0.068	-0.161	-0.841	-0.540
M_2	0.018	0.000	-0.066	-0.109	-0.229	0.166	-0.382	-0.347	0.208
M_3	-0.095	0.016	0.000	0.150	-0.040	-0.061	-0.122	-0.103	-0.148
E_1	-0.216	0.012	-0.598	0.000	0.130	0.110	-0.000	0.012	-0.013
E_2	-0.043	0.014	0.010	-0.521	0.000	-1.040	0.080	0.089	0.054
E_3	0.008	-0.041	0.015	-0.439	0.116	0.000	-0.032	-0.021	-0.044
O_1	0.018	0.024	0.031	0.000	-0.721	0.032	0.000	-0.299	0.074
O_2	0.034	0.022	0.026	-0.012	-0.803	0.021	0.075	0.000	-0.395
O_3	0.034	-0.208	0.037	0.013	-0.217	0.044	-0.295	0.099	0.000

Table 7. We specify that, after various iterations, the value of parameter θ has been fixed to 0.216, that leads to a consistency ratio $CR = 9.997\%$ for the adjusted matrix, within the allowed threshold of 10%.

We eventually submitted matrix X to final expert approval, with positive outcome, and we lastly proceeded by calculating the vector of weights as the Perron eigenvector of matrix X . The last column of Table 7 shows the sought vector for the sub-factors, which we aggregate to get main factors' weights: $w_M = 51.50\%$, $w_E = 23.40\%$ and $w_O = 25.10\%$. This weights will serve as a part of input data for the following ELECTRE I application.

Table 7 Adjusted matrix X

X	M_1	M_2	M_3	E_1	E_2	E_3	O_1	O_2	O_3	Weights
M_1	1.000	1.000	1.011	1.063	2.036	2.995	2.980	4.558	3.795	0.189
M_2	1.000	1.000	1.995	2.990	3.961	2.032	3.888	3.909	1.065	0.203
M_3	0.989	0.501	1.000	0.534	1.999	1.997	1.989	1.993	1.983	0.123
E_1	0.941	0.335	1.874	1.000	0.521	0.515	1.000	1.000	1.000	0.088
E_2	0.491	0.252	0.500	1.918	1.000	2.528	0.342	0.344	0.503	0.072
E_3	0.334	0.492	0.501	1.942	0.396	1.000	0.999	1.000	0.999	0.074
O_1	0.336	0.257	0.503	1.000	2.926	1.001	1.000	1.939	0.507	0.082
O_2	0.219	0.256	0.502	1.000	2.910	1.000	0.516	1.000	1.904	0.081
O_3	0.263	0.939	0.504	1.000	1.988	1.001	1.973	0.525	1.000	0.088

Table 8 Input data for the ELECTRE I application

Weights →	0.515	0.234	0.251
Alternatives ↓	M	E	O
WE_1	4.00	4.00	3.00
WE_2	2.00	4.00	3.00
WE_3	3.00	3.00	3.00
WE_4	4.00	3.00	4.00
WE_5	5.00	4.00	4.00

5.2 Selecting the Priority Work Equipment Via the ELECTRE I

We are now selecting the most critical equipment within the set of five machines belonging to the core subsystem of the packaging plant of the company, called “cardboard boxes line”, completely characterised in [9]. The five machines have been labelled as follows: WE_1 , cartoning machine; WE_2 , bundler; WE_3 , shrink-wrapped; WE_4 , palletizer; WE_5 , wrapping machine. Table 8 provides the evaluation of the five alternatives under the criteria previously weighted. Numerical evaluations have been attributed once again with the help of the expert within the range [1, 5], increasing values indicating higher criticality.

Tables 9, 10 and 11 respectively give the correlation, discordance and outranking matrices. In detail, the correlation indices matrix and the discordance indices matrix (Tables 9 and 10), respectively collecting the correlation and the discordance indices, have been obtained by applying formulas (8) and (10). The outranking matrix reported in Table 11 has been obtained by formula (12). Results have been eventually double checked and validated by means of the J-Electre-v2.0 software for multi-criteria decision aid (<https://sourceforge.net/projects/j-electre/files/>).

Table 9 Correlation matrix

C_{ij}	WE_1	WE_2	WE_3	WE_4	WE_5
WE_1	0.000	1.000	1.000	0.749	0.234
WE_2	0.485	0.000	0.485	0.234	0.234
WE_3	0.251	0.766	0.000	0.234	0.000
WE_4	0.766	0.766	1.000	0.000	0.251
WE_5	1.000	1.000	1.000	1.000	0.000

Table 10 Discordance matrix

D_{ij}	WE_1	WE_2	WE_3	WE_4	WE_5
WE_1	0.000	0.000	0.000	0.333	0.333
WE_2	0.667	0.000	0.333	0.667	1.000
WE_3	0.333	0.333	0.000	0.333	0.667
WE_4	0.333	0.333	0.000	0.000	0.333
WE_5	0.000	0.000	0.000	0.000	0.000

Table 11 Outranking matrix

$S(WE_i, WE_j)$	WE_1	WE_2	WE_3	WE_4	WE_5
WE_1	0.000	1.000	1.000	0.000	0.000
WE_2	0.000	0.000	0.000	0.000	0.000
WE_3	0.000	0.000	0.000	0.000	0.000
WE_4	0.000	0.000	1.000	0.000	0.000
WE_5	1.000	1.000	1.000	1.000	0.000

The ELECTRE I application shows as the equipment WE_5 (wrapping machine) is not dominated by other alternatives, exposing higher degree of criticality for the packaging process as well as for human safety. This equipment should be maintained with priority.

5.3 Discussion of Results and Practical Insights for Management

Practical managerial implications of the obtained results are briefly discussed. Weights have been calculated for the three main risk factors used for work equipment risk assessment. The calculation has been led by adhering as much as possible to the preferences issued by a stakeholder who has proven relevant experience in the field. The “material” factor (M) mostly influences the risk assessment process for

Table 12 Matrix of differences $X - A$ between adjusted and input matrices

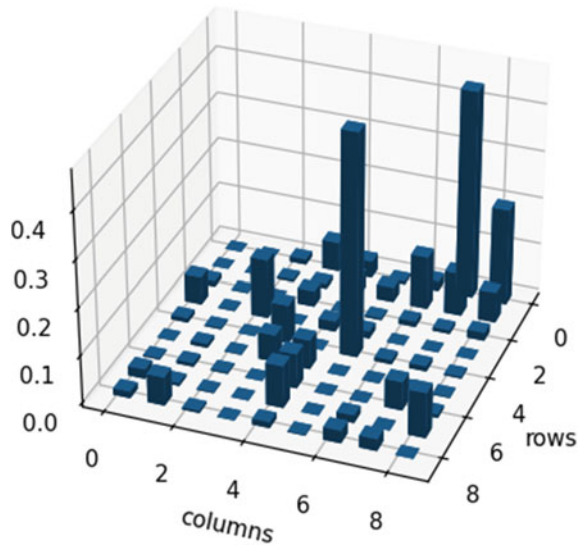
$X - A$	M_1	M_2	$-OM_3$	E_1	E_2	E_3	O_1	O_2	O_3
M_1	0.000	-0.000	0.011	0.063	0.036	-0.005	-0.020	-0.442	-0.205
M_2	0.000	0.000	-0.005	-0.010	-0.039	-0.032	-0.112	-0.091	0.065
M_3	-0.011	0.001	0.000	0.034	-0.001	-0.003	-0.011	-0.007	-0.017
E_1	-0.059	0.001	-0.126	0.000	0.021	0.015	-0.000	0.000	-0.000
E_2	-0.009	0.020	0.000	-0.082	0.000	-0.472	0.008	0.010	0.003
E_3	0.001	-0.008	0.001	-0.058	0.062	0.000	-0.001	-0.000	-0.001
O_1	0.002	0.007	0.003	0.000	-0.074	0.001	0.000	-0.061	0.007
O_2	0.019	0.006	0.002	-0.000	-0.090	0.000	0.016	0.000	-0.096
O_3	0.013	-0.061	0.004	0.000	-0.012	0.001	-0.027	0.025	0.000

work equipment, followed by the “organization” (O) and “environment” (E) factors. Aspects related to work equipment management, above all including safety and security, have hence to be controlled with particular attention to pursue effective risk minimization.

Some considerations need to be formulated by comparing the adjusted matrix X (Table 7), consistently obtained by our negotiation procedure, and the input matrix A (Table 4), originally issued by the expert. To such an aim, we can observe the matrix of differences given in Table 12, showing which comparisons have been altered most with respect to original evaluations. Results have been synthesised in Fig. 3, indicating as the comparisons (E_2, E_3), (M_1, O_2) and (M_1, O_3), with associated differences of respectively 0.472, 0.442 and 0.205 in absolute values, have been altered most. We can conclude that these values (corresponding to the highest columns of Fig. 3, showing absolute values of the differences) are quite small, indicating that input evaluations have been just minimally adjusted to guarantee consistency without distorting experts’ practical opinions.

The following ELECTRE I application, led on the basis of the calculated weights, has driven towards the selection of the alternative WE_5 as a work equipment to be maintained with priority. This selection has been confirmed to be in line with company needs, since the wrapping machine is the equipment closing the cardboard boxes line of the packaging line, from where the packed product is dispatched to the storage area by the operators, by means of the use of dedicated forklifts. It is then clear the importance of the correct functioning of this specific machine for workers’ safety as well as for optimising the whole packaging process, guaranteeing the continuity of service. Suitable periodic maintenance should be carefully planned and implemented within company procedures.

Fig. 3 Absolute values of differences



6 Conclusions

This study proposes a novel iterative procedure to carry out an effective negotiation with experts involved in complex business decision making processes. As a main field of application, we herein consider the work equipment risk assessment, the last one being such a crucial topic in industrial contexts. However, the application can be easily extended to any real life field. Starting from an inconsistent input PCM collecting human pairwise evaluations of preference, the main goal of research consists in calculating priorities of elements while reaching a trade-off between mathematical consistency and maximum adherence to practical reality. The proposed procedure is based on a sensitivity analysis signalling those comparisons most “influencing” to be used as drivers for consistency improvement. Then, a calculation based on the gradient operator will be performed to calculate a final adjusted matrix. This matrix will be characterised by the achievement of minimum consistency requirements by performing minimum numerical variations of judgments elicited by experts. As a result, the decision maker will be asked to agree with the final adjusted matrix to formalise the process. The approach has been built within the framework of AHP and integrated with a further MCDM application based on ELECTRE I. The last one is used to solve the decision making problem by selecting one final alternative according to the criteria weights calculated during the previous negotiation stage.

The proposed approach has been applied to an industrial case study aimed at supporting the process of risk assessment for work equipment management in a real Italian company, operating in the food industry. An effective feedback relation has been established with the expert in charge of equipment maintenance towards an easy calculation of priorities for the main elements of analysis. These elements, e.g.

risk factors, have been chosen from standard procedures of risk evaluation followed by the company. Once calculated risk factors' weights, the equipment in main need of maintenance has been selected by the ELECTRE I.

References

1. Arcari C, Bernazzani A, Mazzari M (2009) LINEE GUIDA Per l'APPLICAZIONE DEL D. Lgs. 81/08 N. 05 - Requisiti di conformità e valutazione del rischio delle attrezzature. Servizio Sanitario Regionale Emilia Romagna
2. Benítez J, Delgado-Galván X, Gutiérrez J, Izquierdo J (2011) Balancing consistency and expert judgment in AHP. *Math Comput Modell* 54(7–8):1785–1790
3. Benítez J, Delgado-Galván X, Izquierdo J, Pérez-García R (2011) Achieving matrix consistency in AHP through linearization. *Appl Math Modell* 35(9):4449–4457
4. Benítez J, Izquierdo J, Pérez-García R, Ramos-Martínez E (2014) A simple formula to find the closest consistent matrix to a reciprocal matrix. *Appl Math Modell* 38:15–3974
5. Bozóki S, Csató L, Temesi J (2016) An application of incomplete pairwise comparison matrices for ranking top tennis players. *Euro J Oper Res* 248(1):211–218
6. Bulut M, Özcan E (2021) A new approach to determine maintenance periods of the most critical hydroelectric power plant equipment. *Reliab Eng Syst Saf* 205:107238
7. Butdee S, Phuangsalee P (2019) Uncertain risk assessment modelling for bus body manufacturing supply chain using AHP and fuzzy AHP. *Procedia Manuf* 30:663–670
8. Carpitella S, Certa A, Izquierdo J (2021) Flexible negotiation process to adhere to human preferences; a case of work equipment risk assessment. In: 26th ISSAT international conference on reliability and quality in design, pp 261–265
9. Carpitella S, Mzougui I, Benítez J, Carpitella F, Certa A, Izquierdo J, La Cascia M (2021) A risk evaluation framework for the best maintenance strategy: the case of a marine salt manufacture firm. *Reliab Eng Syst Saf* 205:107265
10. Čerňanová V, Koczkodaj WW, Szybowski J (2018) Inconsistency of special cases of pairwise comparisons matrices. *Int J Approx Reason* 95:36–45
11. Ekmekcioğlu Ö, Koc K, Özger M (2021) Stakeholder perceptions in flood risk assessment: a hybrid fuzzy AHP-TOPSIS approach for Istanbul, Turkey. *Int J Disaster Risk Reduct* 60:102327
12. Finan JS, Hurley WJ (1997) The analytic hierarchy process: does adjusting a pairwise comparison matrix to improve the consistency ratio help?. *Comput Oper Res* 24:8749–755
13. Franek J, Kresta A (2014) Judgment scales and consistency measure in AHP. *Procedia Econ Financ* 12:164–173
14. George PG (2021) Evolution of safety and security risk assessment methodologies to use of bayesian networks in process industries. *Process Saf Environ Protect*
15. Gershon M, Duckstein L, McAniff R (1982) Multiobjective river basin planning with qualitative criteria. *Water Resour Res* 18(2):193–202
16. Govindan K, Jepsen MB (2016) ELECTRE: a comprehensive literature review on methodologies and applications. *Euro J Oper Res* 250(1):1–29
17. Gu W, Basu M, Chao Z, Wei L (2017) A unified framework for credit evaluation for internet finance companies: multi-criteria analysis through AHP and DEA. *Int J Inf Technol Decis Making* 16(03):597–624
18. Khalilzadeh M, Shakeri H, Zohrehvandi S (2021) Risk identification and assessment with the fuzzy DEMATEL-ANP method in oil and gas projects under uncertainty. *Procedia Comput Sci* 181:277–284
19. Lehmann EL, Romano JP (2006) Testing statistical hypotheses testing statistical hypotheses. Springer Science&Business Media

20. Liao H, Yang L, Xu Z (2018) Two new approaches based on ELECTRE II to solve the multiple criteria decision making problems with hesitant fuzzy linguistic term sets. *Appl Soft Comput* 63:223–234
21. Liu X, Wan S-P (2019) A method to calculate the ranges of criteria weights in ELECTRE I and II methods. *Comput Ind Eng* 137:106067
22. Lopez JCL, Noriega JJS, Chavira DAG (2017) A multi-criteria approach to rank the municipalities of the states of Mexico by its marginalization level: the case of Jalisco. *Int J Inf Technol Decis Making* 16(02):473–513
23. Ng CT, Lee W, Lee Y (2020) Logical and test consistency in pairwise multiple comparisons. *J Stat Plan Inference* 206:145–162
24. Ocampo-Melgar A, Gironás J, Valls A (2018) A rule-based approach for preventive identification of potential conflictive criteria in mining operations in Chile. *J Clean Prod* 184:559–568
25. Ok K, Okan T, Yilmaz E (2011) A comparative study on activity selection with multi-criteria decision-making techniques in ecotourism planning. *Sci Res Essays* 6(6):1417–1427
26. Pérez IJ, Cabrerizo FJ, Alonso S, Dong Y, Chiclana F, Herrera-Viedma E (2018) On dynamic consensus processes in group decision making problems. *Inf Sci* 459:20–35
27. Pham BT, Luu C, Van Dao D, Van Phong T, Nguyen HD, Van Le H et al (2021) Flood risk assessment using deep learning integrated with multi-criteria decision analysis. *Knowl-Based Syst* 219106899
28. Roy B (1968) Classement et choix en présence de points de vue multiples. *Revue française d'informatique et de recherche opérationnelle* 2:857–75
29. Roy B (1990) The outranking approach and the foundations of ELECTRE methods. In: *Readings in multiple criteria decision aid*. Springer, PP 155–183
30. Saaty TL (1977) A scaling method for priorities in hierarchical structures. *J Math Psychol* 15(3):234–281
31. Saaty TL (2003) Decision-making with the AHP: why is the principal eigenvector necessary. *Euro J Oper Res* 145(1):85–91
32. Saaty TL (2008) Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. *RACSAM Serie A. Mat* 102(2):251–318
33. Safarzadeh S, Khansefid S, Rasti-Barzoki M (2018) A group multi-criteria decision-making based on best-worst method. *Comput Ind Eng* 126111–126121
34. Stefanović V, Urošević S, Mladenović-Ranisavljević I, Stojilković P (2019) Multi-criteria ranking of workplaces from the aspect of risk assessment in the production processes in which women are employed. *Saf Sci* 116:116–126
35. Stewart GW (2001) *Matrix algorithms: Volume II: eigensystems matrix algorithms: Volume II: Eigensystems*. SIAM
36. Syan CS, Ramsoobag G (2019) Maintenance applications of multi-criteria optimization: a review. *Reliab Eng Syst Saf* 190:106520
37. Szybowski J, Kułakowski K, Prusak A (2020) New inconsistency indicators for incomplete pairwise comparisons matrices. *Math Soc Sci* 108:138–145
38. Tabatabaee S, Mahdiyar A, Ismail S (2021) Towards the success of building information modelling implementation: a fuzzy-based MCDM risk assessment tool. *J Build Eng*. In press
39. Tian ZP, Nie RX, Wang JQ, Zhang HY (2018) A two-fold feedback mechanism to support consensus-reaching in social network group decision-making. *Knowl-Based Syst* 162:74–91
40. Zhang J, Kou G, Peng Y, Zhang Y (2021) Estimating priorities from relative deviations in pairwise comparison matrices. *Inform Sci* 552:310–327

Silvia Carpitella California State University, Northridge, CA

Dr. Silvia Carpitella holds a Ph.D. in Technological Innovation Engineering and a Ph.D. in Mathematics. She has been an Assistant Professor at the Department of Manufacturing Systems Engineering and Management of the California State University, Northridge, since August 2022.

Silvia's activity has been internationally recognised by means of two prizes awarded to her doctoral thesis, the first one in Spain and the second one in Italy, and through the Otto Wichterle award, granted from the Czech Academy of Science for the contributions to the advancement of scientific knowledge in her field.

Antonella Certa University of Palermo, Palermo, Italy

Dr. Antonella Certa is an Associate Professor at the Engineering Department of the University of Palermo, Italy. Her research interests are mainly focused on project management, multi-criteria analysis, and maintenance optimization models. Antonella has been teaching the subject "Production and industrial plant management" since the year 2013. She participated in national and international research projects supported by competitive grants programs.

Joaquín Izquierdo Universitat Politècnica de València, Valencia, Spain

Dr. Joaquín Izquierdo, Ph.D. in Mathematics and Full Professor (currently Emeritus Professor) of Applied Mathematics at the Universitat Politècnica de València (UPV). Educational career developed in the UPV and research activity as the Director of the research group FluIng, devoted to mathematical modelling, knowledge-based systems, and DSSs in Engineering. Author and editor of several books and author or co-author of more than 300 research papers, book chapters, and contributions to international events. Tutor of 16 Doctoral Dissertations, wide experience in consulting and R&D projects.

Joining Aspect Detection and Opinion Target Expression Based on Multi-Deep Learning Models



Bui Thanh Hung

Abstract Aspect-based Sentiment Analysis (ABSA) is an advanced task as well as technique which is developed based on sentiment analysis. Aspect-based sentiment analysis is a text analysis technique that categorizes data by aspect and identifies the sentiment attributed to each one. Aspect-based sentiment analysis can be used to analyze customer feedback data by associating specific sentiments with different aspects (e.g. the attributes or components) of a product or service. This has attracted increasing attention in the recent few years in Natural Language Processing and has broad applications in both research and business. In this research, we apply joining aspect detection and opinion target expression using multi-deep learning methods: RNN, LSTM and CNN and we do experiments on Vietnamese VLSP2018 dataset. Both of the results for the aspect detection as well as opinion target expression achieve the best results on the CNN model.

Keywords Aspect-based sentiment analysis · Opinion target expression · Aspect detection · Deep learning · RNN · LSTM · CNN · Word2vec

1 Introduction

With the advancement of technology and digital revolution, nowadays many businesses have transformed digitally to invest more on online sales channel and conversational commerce (such as websites, social networks, forums, etc...). In addition, the customers especially the young generations are more vocal and like to interact with a brand than ever. They not only enjoy surfing the brands' website, their Facebook page, share information with their friends, but also love to leave feedback, comments, reviews about the products and services as well as the experience they had with the brands. Those comments and reviews are certainly the valuable insights for the businesses to let them know what they are doing well and what still needs to improve. However, it might be extremely impossible to analyze thousands of

B. T. Hung (✉)

Data Science Laboratory, Faculty of Information Technology, Industrial University of Ho Chi Minh City, 12 Nguyen Van Bao Street, 4 Ward, Go Vap District, Ho Chi Minh City, Vietnam
e-mail: buihanhhung@iuh.edu.vn

customer comments, feedback or reviews manually—especially on a granular level. Therefore, the aspect-based sentiment analysis plays an important role to solve this problem by determining the user’s sentiments in each specific aspect. That said this has become a hot topic and attracted increasing attention from both the business and research in recent few years.

In aspect-based sentiment analysis, there are three basic problems such as aspect detection, opinion target expression and sentiment polarity. In this research, we propose joining aspect detection and opinion target expression using multi-deep learning methods. We do experiments on RNN, LSTM and CNN models. The rest of the paper is constructed as follows: related works are reviewed in Sect. 2; Sect. 3 presents our proposed method; our experiments are described in Sect. 4; Sect. 5 gives conclusion and future direction.

2 Related Works

For a long period of time, analyzing emotions in aspects has attracted the attention of many researchers and businesses. Around the 2000s, opinion mining and emotion analysis was introduced and immediately became the hot topic of natural language processing. Marzieh et al. [1] performed sentiment analysis on the Sentihood dataset built by reviews and comment on landmarks and neighborhoods in the city of London on the Yahoo. Aspect-based Sentiment Analysis ABSA for English and other languages (non-Vietnamese) was originally introduced in SemEval 2014, 2015, 2016 with many published datasets and methods [2–4]. Since then, there has been a variety of researches and studies to be introduced across the world to propose methods to solve this problem. Recently, in aspect-based sentiment analysis, deep learning method has been presented as a fine-tuning method, especially using RNN, LSTM and CNN models [5–7] which outperform the previous state-of-the-art results.

For the Vietnamese aspect-based sentiment analysis, there has been a number of studies such as: Son et al. [8] presented a study on the exploitation of opinions based on the aspect of product using Vietnamese syntax rules; Thuyet al. [9] built a Vietnamese corpus for the restaurant data domain at the sentence level and solved the aspect detection by combining the hand-labeled corpus and the English-Vietnamese translation model; Vo et al. [10] built a dataset named “Vietnamese Sentiment” and introduced a new model that integrates the Convolutional Neural Network (CNN) and LSTM; Bui et al. [11] proposed hybrid deep learning approach for aspect detection.

In 2018, the Vietnamese Language and Speech Processing (VLSP) community organized a contest and published the corpus of ABSA in restaurant and hotel data domain of Huyen et al. [12]. The corpus is built for the opinion analysis problem in aspect with 4751 comments for the restaurant domain and 5600 comments for the hotel domain.

Speaking of our research, what makes it different with the others is the joining aspect detection and opinion target expression based on multi-deep learning models.

3 The Proposed Method

The proposed model includes the data preprocessing, converting the text to word embeddings and training model with multi deep learning algorithms: RNN LSTM and CNN. The framework of our proposed model is shown in Fig. 1.

3.1 Multi Deep Learning Models

Recurrent Neural Network—RNN

RNN has proved as an effective method for natural language processing. This method uses previous information to present the task. However, RNN cannot work well in long term dependencies. Figure 2 shows a cell of RNN.

With h_t is current state, h_{t-1} is previous state and x_t is input state, the current state is calculated by the formula:

Fig. 1 The proposed model

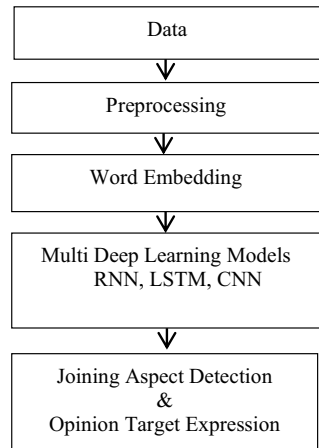
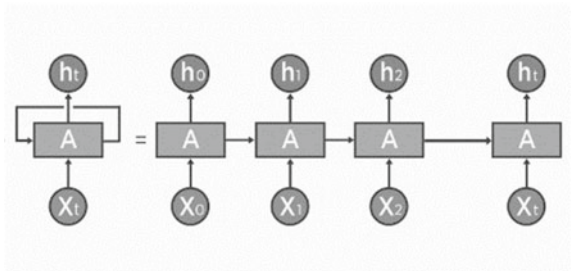


Fig. 2 RNN model



$$h_t = f(h_{t-1}, x_t) \quad (1)$$

Long Short Term Memory—LSTM

Long short-term memory model is modification of RNN [13–15], this model works well on long term dependence with 3 gates: forget gate f_t , input gate i_t and output gate o_t , the memory cell is updated at time t with the value h_{t-1} of the memory cell output from the previous time step $t - 1$ the input value x_t as follows:

$$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (2)$$

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \quad (3)$$

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (4)$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (5)$$

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (6)$$

$$h_t = o_t \tanh(c_t) \quad (7)$$

LSTM is also initialized by a pre-trained word2vec. Then the neuron in the LSTM model will learn on these vocabularies and give a vector representation for the input comments. The model results are presented as shown in Fig. 3. In LSTM, we use 128 of number of neurons. We use the Attention class to choose the words that express aspect and opinion effectively. The number of the Attention layer is also 128. The dropout value is chosen to be 0.5 and the activation function is Relu.

Convolutional Neural Network—CNN

Convolutional neural network is a form of multilayer neural network; each layer belongs to one of three types: convolution, subsampling, fully connection [16–18]. The convolution mathematical operation is done in the Convolution layer. Subsampling or the Pooling layer is responsible for the reduction of the size (spatial) of the convolved feature. The Fully Connected layer comprises the weights and biases together with the neurons and is used to connect the neurons between two separate layers. The CNN convolutional network has been studied and applied to solve many natural language processing problems effectively. In this research, we apply the CNN model to detect aspects and expression. Figure 4 shows the architecture of the CNN model in our proposed method. The model consists of the six main components: Input, Word embedding, Convolution, Pooling, Fully connected and Output layer. Each layer has the following functions:

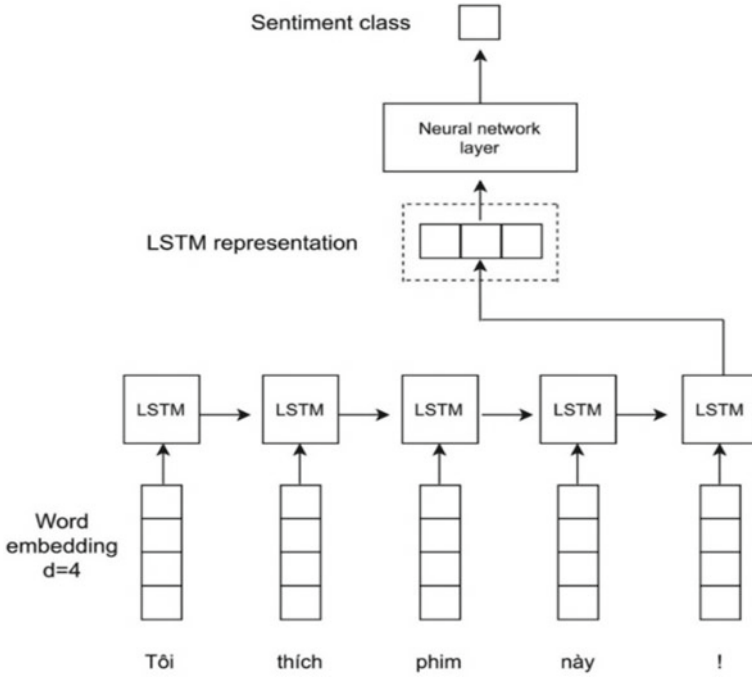


Fig. 3 LSTM model

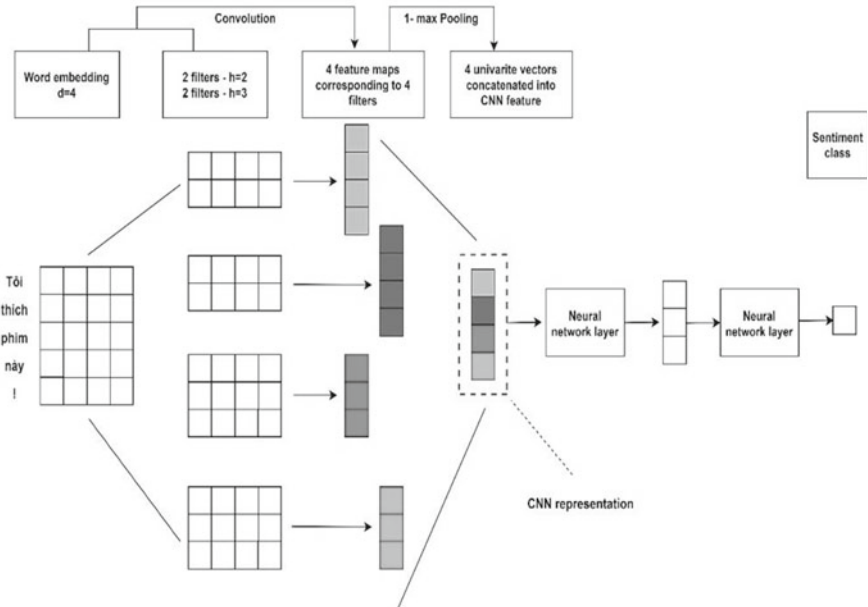


Fig. 4 CNN model

- Input: Each comment after going through the preprocessing step will be represented as numeric vectors with fixed length, we use the longest comment which is the dimension of the fixed vector. Comments that are not long enough will be added the <PAD> value automatically.
- Word embedding: we use Fasttext, a pre-trained word embedding published by Facebook AI Research.
- Convolution layer: In this layer, we apply 128 filters with different sizes through the matrix of the word embedding. Specifically, the filter sizes used in this class are 3, 4 and 5.
- Pooling layer: we use a combination of max pooling and average pooling for each convolution layer to extract important features.
- Classifier: After extracting the important features of the pooling layer, we pass these features through the classifier with the Relu activation function.
- Output layer: Each aspect will be represented as a four-hot vector

The dimension of the word2vector is 300, the filter is 128 and the window size is 3, 4, 5 respectively. Each layer has 300 neurons. We use Relu activation function and Adam is optimization algorithm.

3.2 *Joining Aspect Detection and Opinion Target Expression*

Each aspect will be represented as a four-hot vector with the following information:

- If the value 1 is on the first position of a four-hot vector, the comment will not be assigned to the considered aspect.
- If the value 1 appears on the position from 2 to 4, then the aspect category will be assigned to the comment along with the opinion target expression positive, neutral and negative, respectively.
- Example: The type of considered aspect is Food#Quality, the output vector is vector [0, 1, 0, 0], it will be interpreted as a comment assigned the Food#Quality aspect and opinion target expression is positive.

We use the process to get the result of joining aspect detection and opinion target expressions as follows:

Step 1: Prepare dataset.

Step 2: Process the dataset with following steps:

- Return to lowercase to reduce the size of the vocabulary set
- Remove a lot of spaces because comments are collected directly from the website, so there are a lot of punctuation errors
- Replace numbers with special characters to reduce vocabulary size
- Remove punctuation, characters in the entire comment
- Extract words on comments.

- Step 3: Extract features using Fasttext pre-trained word2vec.
- Step 4: Use multi deep learning models: RNN, LSTM and CNN to train and return the results.
- Step 5: Joining Aspect Detection and Opinion Target Expression results.

4 Experiments

4.1 Dataset

We use the ABSA dataset of VLSP 2018 data [12]. This dataset is collected from various articles on the social media. There are three aspects in this dataset: Location, Organization and Person. Figure 5 shows details about the dataset.

The list of 12 aspect labels is presented as follows:

- Food#Quality: the food, the quality, the taste of the food.
- Food#Style&Option: the serving option or the dish style.
- Food#Prices: the detailed price of each dish.
- Drink#Quality: the quality of the drink.
- Drink#Prices: the detailed price of the drink.
- Drink#Style&Options: the recipe, the serving option or the detail of the drink.
- Service#General: the service style of the staff, the restaurant, the park in general.

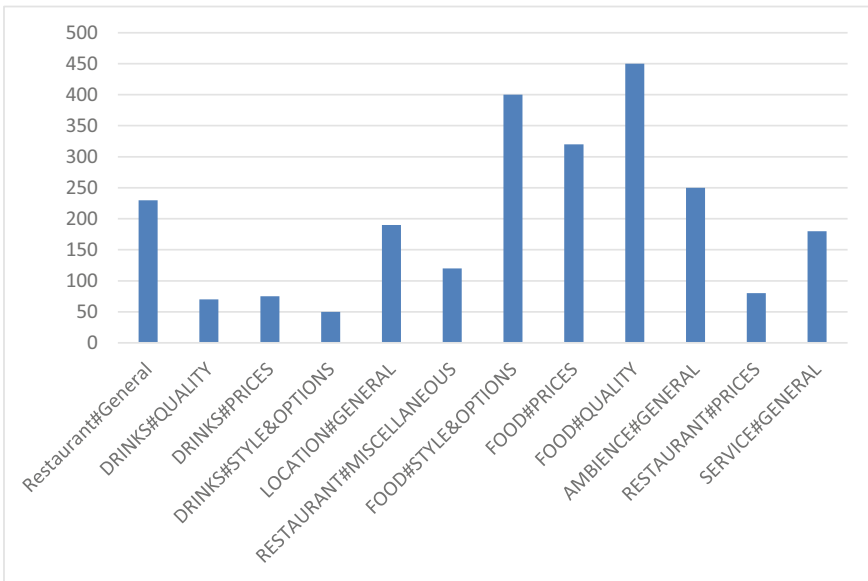


Fig. 5 VLSP2018dataset

Table 1 The dataset

Dataset	Number of comment	Number of aspect
Train	2961	9034
Validation	1290	3408
Test	500	2419

- **Ambience#General**: the restaurant’s atmosphere and space.
- **Location#General**: the location of the restaurant, the view of the restaurant.
- **Restaurant#General**: refers to the restaurant in general.
- **Restaurant#Prices**: the general cost of the restaurant or other service costs.
- **Restauran#Miscellenous**: restaurant problems like restrooms, etc.

We divided the dataset to train, test and validation as shown in Table 1. We used Vietnamese FastTextword2vec of Facebook AI Research. We installed in Python and used deep learning framework of the Keras library [19], Tensorflow support [20]. We evaluated this task by Precision, Recall, F1 score and Accuracy measurement calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where: TN: True Negative, TP: True Positive
 FN: False Negative, FP: False Positive.

To evaluate the aspect detection and the joining aspect detection and opinion target expression, we did these experiments separately. We received the results of aspect detection as shown in Tables 2 and 3 shows the result of joining aspect detection and opinion target expression.

Experimental results show that CNN and LSTM give better results than the RNN on the test set. Specifically, for the aspect detection problem, the RNN method achieved Precision of 83.77%, Recall of 66.56% and 74.18% of F1-score; LSTM with 82.43%, 79.26% and 80.82%, respectively. And CNN gave the best results with 84.49%, 80.82% and 82.61%, respectively.

Table 2 The results of aspect detection

Model	Precision	Recall	F1 score
RNN	83.77	66.56	74.18
LSTM	82.43	79.26	80.82
CNN	84.49	80.82	82.61

Table 3 The results of opinion target expression

Model	Precision	Recall	F1 score
RNN	62.59	49.73	55.43
LSTM	60.23	55.72	57.89
CNN	60.46	57.83	59.12

For the Opinion Target Expression problem, the RNN model achieved 62.59%, 49.73% and 55.43% for Recall, Precision and F1-score respectively. The results of the CNN model are 60.46%, 57.83% and 59.12%, respectively.

Based on the result, it proves that the CNN model produces the best results in both of aspect detection and Opinion Target Expression. Next, we analyze the results CNN model because this method gives the best results for both problems. We will discuss in more detail the results of the CNN model on aspect detection as well as Opinion Target Expression in Figs. 6 and 7.

From the Figs. 6 and 7, it could be seen that the high performance of aspect detection on Food#Quality, Service#General, Restaurant#General and Food#Style&Options. Because these labels have a large amount of labeling data in the training dataset, when it comes to the training, those results are effective. Meanwhile, labels with a small number of trainings scored as low such as Drink#Prices, Restaurant#Prices. Therefore, to increase the efficiency of the result, we need to add more labeled data. From the result, we saw that, the result of aspect detection affects

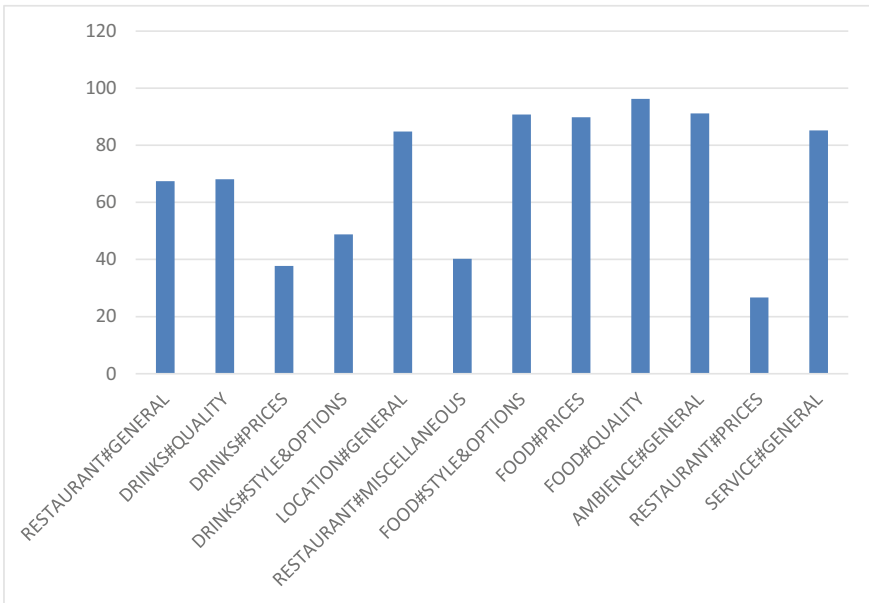


Fig. 6 The results of CNN model on aspect detection

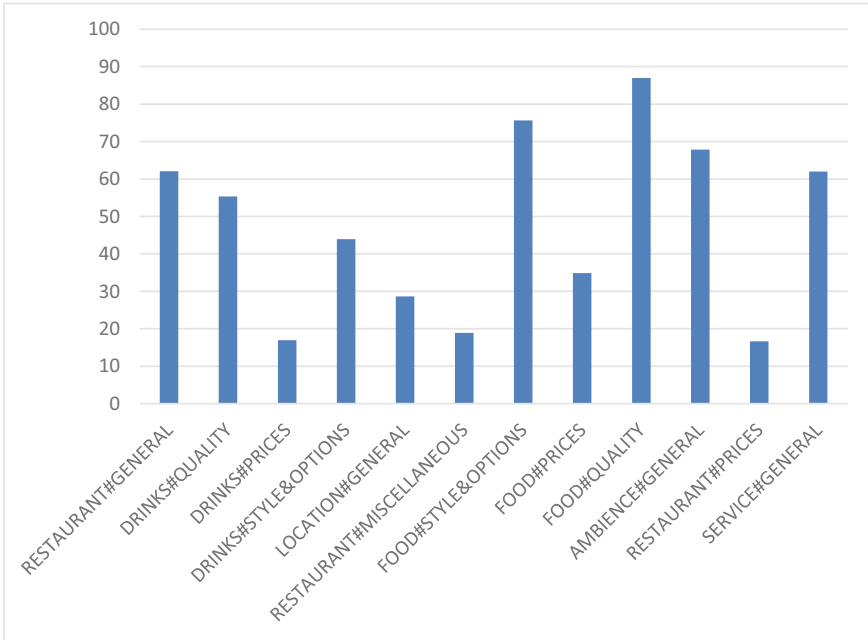


Fig. 7 The results of CNN model on Opinion Target Expression

to Opinion Target Expression, so to improve Opinion Target Expression result, we have to improve aspect detection.

5 Conclusion

In this research, we proposed the Joining Aspect Detection and Opinion Target Expression. We did experiments on multi deep learning methods: RNN, CNN and LSTM. The results of the CNN method are very promising and have potential for further research. In the future, we will add more labeled data, change the appropriate parameters of deep learning model and test on other deep learning modelsto improve the results.

References

1. Saeidi M, Bouchard G, Liakata M, Riedel S. Sentihood (2016) Sentihood: targeted aspect based sentiment analysis dataset for urban neighbourhoods. In: Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical Papers, pp 1546–1556
2. Pontiki M, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, Al-Smadi M, Al-Ayyoub M, Zhao Y, Qin B, De Clercq O, Hoste V (2014) SemEval-2014 task 4: aspect based sentiment analysis. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval), pp 27–35
3. Schouten K, Frasinca F (2016) Survey on aspect-level sentiment analysis. *IEEE Trans Knowl Data Eng* 28(3):813–830
4. Maria P, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, Al-Smadi M, Al-Ayoub M, Zhao Y, Qin B, De Clercq O, Hoste, Apidianaki M, Tannier X, Loukachevitch N, Evgeniy (2016) SemEval-2016 task 5: aspect based sentiment analysis. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval)
5. Zhang Y, Wallace B (2015) A sensitivity analysis of convolutional neural networks for sentence classification. [arXiv:1510.03820](https://arxiv.org/abs/1510.03820)
6. Wang X, Liu Y, Sun CJ, Wang B, Wang X (2015) Predicting polarities of tweets by composing word embeddings with long short-term memory. *ACL*
7. Liu P, Joty S, Meng H (2015) Fine-grained opinion mining with recurrent neural networks and word embeddings. *EMNLP*
8. Le HS, Van Le T, Pham TV (2015) Aspect analysis for opinion mining of vietnamese text. In: Proceedings of the 2015 international conference on advanced computing and applications (ACOMP) (ACOMP '15). IEEE Computer Society
9. Thuy NT, Bach NX, Phuong TM (2018) Cross-language aspect extraction for opinion mining. In: 10th international conference on knowledge and systems engineering (KSE)
10. Vo Q, Nguyen H, Le B, Nguyen M (2017) Multi-channel LSTM-CNN model for Vietnamese sentiment analysis. In: 9th international conference on knowledge and systems engineering (KSE), pp 24–29. <https://doi.org/10.1109/KSE.2017.8119429>
11. Hung BT, Semwal VB, Gaud N, Bijalwan V (2021) Hybrid deep learning approach for aspect detection on reviews. proceedings of integrated intelligence enable networks and computing. In: Springer series in algorithms for intelligent systems
12. Nguyen HTM, Nguyen HV, Ngo QT, Vu LX, Tran VM, Ngo BX, Le CA (2018) VLSP shared task: sentiment analysis. *J Comput Sci Cybern* 34(4):295–310
13. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
14. Hung BT (2019) Integrating diacritics restoration and question classification into vietnamese question answering system. *Special Issue on Adv Eng Comput Sci J—ASTESJ* 4(5):207–212
15. Hung BT (2019) Vietnamese question classification based on deep learning for educational support system. In: The 19th international symposium on communications and information technologies, ISCIT
16. Hung BT (2019) Domain-specific versus general-purpose word representations in sentiment analysis for deep learning models. In: *Frontiers in intelligent computing: theory and applications*. Springer, pp 252–264
17. Hung BT (2022) College admissions counseling using intelligent question answering system. In: *Intelligent systems and sustainable computing. Smart innovation, systems and technologies*, vol 289. Springer, Singapore, pp 27–37
18. Hung BT (2022) Link prediction in paper citation network based on deep graph convolutional neural network. In: *Computer networks, big data and IoT. Lecture notes on data engineering and communications technologies*, vol 117. Springer, Singapore

19. Keras. <https://keras.io/>
20. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, sIsard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X (2016) Tensorflow: a system for large-scale machine learning. Tech. rep. Google Brain

Bui Thanh Hung Ton DucThang University, Ho Chi Minh City, Vietnam

Dr. Bui Thanh Hung received his M.S. degree and Ph.D. degree from the Japan Advanced Institute of Science and Technology (JAIST) in 2010 and in 2013. He is currently the Director of the Artificial Intelligence Laboratory, Faculty of Information Technology, Ton Duc Thang University. He has completed 2 projects and published 14 journals, 18 book chapters, 36 International conference papers, and 16 domestic conference papers. He had four best paper awards of FICTA 2018, RICE 2020, ICAIAA 2021, and CIEMA 2022. He was selected as an Excellent Scientific Researcher of CIEMA 2022 and Thu Dau Mot University in 2019, 2020, and 2021. He is a reviewer of many reputed journals and conferences. His main research interests are Natural Language Processing, Machine Learning, Machine Translation, Text Processing, Data Analytics, Computer Vision, and Artificial Intelligence.

Voting Systems with Supervising Mechanisms



Tingnan Lin and Hoang Pham

Abstract This chapter first covers various existing works of voting systems in different aspects, such as design of voting systems, weighted voting systems, weighted voting classifiers and reliability evaluation of voting systems. Then, different inspection policies are reviewed, including periodic inspection, sequential inspection and state dependent inspection. After that, some recent studies on Intervened Decision-making Systems, which are voting systems with inspections and other supervising mechanisms, are presented. Finally, some future research trends are discussed.

1 Introduction

Decision making is one of the most common and important activities in daily life. It is essential in various context, including computer control [21], energy planning [98], relief logistic planning [62], human organization [78], software development [39], communication [86], allocation of remote power feeding systems [91], network resource reallocation [122] and security surveillance [115]. Since the world is full of uncertainty, how to make good decisions based on uncertain information is an eternal topic and a lot of work has been done for different applications.

Since the capability of individuals is limited, it is very hard to make correct decisions all the time. Similar with some digital components, which have two failure modes: fail open and fail close, people can also have different types of errors when making decisions. For example, in statistical tests, people may make Type I errors or Type II errors, that is, people may reject the null hypothesis when it is true or fail to reject it when the alternative hypothesis is true; in economic systems, a profitable idea may be rejected by a company, but it may be implemented by its competitor [96], which will lead to loss; in an expert system [78], a good proposal may be rejected or a wrong proposal may be accepted.

T. Lin (✉) · H. Pham

Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA
e-mail: tl385@scarletmail.rutgers.edu

In order to improve the quality of decisions, ideas from multiple sources can be put together and the final decision can be made based on some particular rule, which is the idea of Voting System. In the 1950s, Neumann [77] put forward the idea that voting mechanism could be used to obtain highly reliable data from multiple unreliable versions. In addition, a group of individuals may be able to do more than a single individual [96]. These are the initial ideas of Voting System. Voting system is a more generalized idea than k -out-of- n system, which is fault tolerant with redundancy. However, given n units in the system, the value of k required to make the final decision correct may not be constant and may depend on the voting rule, inputs and change of system configuration. There are many voting criteria to determine the final results, such as plurality and threshold voting, and their efficiency is different under different situations [82, 83]. Moreover, the decision unit in the system usually has multiple failure modes [45]. Since the function of a voting system is to make decisions, cost-benefit issues will occur due to the decisions produced by the system, which are also not considered in traditional k -out-of- n systems.

Voting systems can be applied in many situations. For example, in a human-organization (HO) system [78], a group of experts will get together to determine whether a proposal should be approved or not; in a financial investment system, a group of agents will determine whether to take the investment option or not [58]. Other applications include imprecise data handling [18], failure detection [107], safety monitoring and self-testing [4], signal processing [24], embedded surveillance systems [5, 6], communication systems [86] and credit card fraud detections [93].

Moreover, the importance of each individual unit can be different in a voting system. For example, in a company, the idea from the CEO or senior staff is usually more important than regular employees; in an academic conference, the opinion from a senior expert with high reputation is usually more important than new graduates. This generates the idea of weighted voting system (WVS). Weighted voting systems have wide applications, including human organization (HO), clustering [22], breast cancer diagnosis [11], text sentiment classification [80] and wireless sensor networks [63, 65]. For example, Bashir et al. [11] have applied several different classifiers to diagnose breast cancers and the final prediction is made based on weighted voting technique, where the weights are determined by the accuracy of each classifier, which the best accuracy achieved is more than 97%, Onan et al. [80] have proposed an ensemble method using majority voting by assigning weights based on the predictive performance of each classification algorithm for text sentiment classification, which the best accuracy achieved is about 99%.

Decision making can be time consuming because correct decisions are always desired but making decisions perfectly can take long time. However, we always expect that good decisions should be made within limited time. In addition, with the use of voting systems to improve the quality of final decisions, the ideas from multiple sources are put together and analyzed, which potentially increases the duration of a decision process even more. However, time is valuable and in many real cases, the time to make a decision is limited and it is impossible to wait until the decision is made naturally without any action, which motivates the idea of supervised decision

process. With some appropriate supervising mechanisms, a decision process can be well controlled, which achieves good balance between quality and efficiency.

The remaining part of this chapter is organized as follows: in Sect. 2, many existing works about voting systems, including general voting systems, weighted voting systems and reliability modeling of voting systems, are discussed; In Sect. 3, various inspection models are reviewed, which provides very constructive inspirations on supervising mechanisms for decision-making and voting systems; Sect. 4 presents some recent and current research about voting systems with supervising mechanisms, mainly focusing on Intervened Decision-making Systems, with a numerical example for demonstration; Sect. 5 draws the conclusion and provides possible future directions.

2 Modeling of Voting Systems

Voting systems play critical roles in modern society because important decisions are usually not made by a single individual but a group of individuals based on certain criteria. Even though a voting system is fault tolerant, if the number of decision units providing incorrect results exceeds a certain limit, the final decision from the voting system will be incorrect. Since the motivation of voting is to improve the quality of decisions, the failure of a voting system can be catastrophic because the system is assumed to be highly reliable and the failure is usually beyond expectation. As a result, it is very important to study the reliability and performability of voting systems. In addition, the rules to make the final decision should be selected properly and the structure as well as relative parameters should be determined carefully to achieve better performance. Related works in various aspects of voting systems will be discussed in this section.

2.1 General Voting Systems

As mentioned in Sect. 1, the idea of voting was first introduced in 1950s by Neumann [77]. Since voting systems are very useful and efficient in many situations, a lot of research has been done for this topic during the past decades. Many papers have expressed that voting systems are more reliable than a single individual, such as Neumann [77], Sah and Stiglitz [96], Kim et al. [33], Kim et al. [35] and Bai et al. [5]. Kim et al. [33] have designed an All Voting Triple Modular Redundancy (AVTMR) and used Markov models to analyze its reliability, availability and MTTF. They have claimed that AVTMR is more reliable than a single modular system. Bai et al. [5] have designed an embedded surveillance system for temperature monitoring with a majority voting rule. Their experiment has shown that a multiple-sensor system with voting mechanism has higher reliability than a single-sensor system.

Voting algorithms and rules have been studied in Lorzczak et al. [67], Blough and Sullivan [15], Parhami [82, 83], Latif-Shabgahi and Bennett [38], Zuev and Ivanov [124], Hillinger [31] and Alcantud and Laruelle [1]. In Lorzczak et al. [67], several commonly used voting techniques have been generalized to arbitrary N -version systems with arbitrary output types using a metric space framework. Blough and Sullivan [15] have proposed a new and more complex voting algorithm by introducing four distinct probability distributions to the model. These probability distributions can represent the behavior of both faulty and non-faulty modules as well as other underlying distributions. Latif-Shabgahi and Bennett [38] have designed an adaptive majority voting algorithm, in which the history of modules is considered and the result from the most reliable module from the majority set is selected as the output. Alcantud and Laruelle [1] have proposed a characterization of Dis & approval voting, in which each voting unit is allowed to have three levels of results: -1 , 0 , and 1 .

Instead of perfect voters, the study of imperfect voters can be found in Mine and Hatayama [75], Pham [86], Kshirsagar and Patrikar [37], Ban and de Barros Naviner [8, 9] and Balasubramanian and Prasad [7]. Mine and Hatayama [75] have considered different types of imperfect voters in majority voting systems, including output-imperfect voters and semi-imperfect voters. The reliability has been calculated for each system and an iterative structure for voters has been introduced to improve the reliability. Pham [86] has studied imperfect voters in threshold voting systems with two failure modes with application to communication systems. In this work, the reliability of the system can be expressed as:

$$R(k, n) = \sum_{i=0}^{k-1} \binom{n}{i} (\beta_0 r_0 p_{10}^i p_{00}^{n-i} + \beta_1 (1 - r_0) p_{11}^i p_{01}^{n-i}) \\ + \sum_{i=k}^n \binom{n}{i} (\beta_0 (1 - r_1) p_{10}^i p_{00}^{n-i} + \beta_1 r_1 p_{11}^i p_{01}^{n-i})$$

where the input dependent reliability for both voting units and the voter has been considered. Ban and de Barros Naviner [8] have considered the failure of voters and proposed the design of a novel majority voter used in triple modular redundancy (TMR) by considering only one fault at given time. Balasubramanian and Prasad [7] have proposed a more robust fault-tolerant majority voter with consideration of both single fault and multiple faults at given time.

In some papers, authors have used “ k -out-of- n systems with multiple failure modes” to represent the idea of voting systems. Such work can be found in Bendov [12], Jenney and Sherwin [32], Page and Perry [81], Sah and Stiglitz [97], Pham and Pham [89], Pham and Malon [88], Pham [87] and Torres-Echeverría et al. [102, 103]. Sah and Stiglitz [97] have studied a k -out-of- n system with binary inputs and two failure types. The paper mainly focuses on determining the optimal value of k to maximize mean profit of the system. The mean profit of the system considering the input dependent reliability functions can be expressed as [97]:

$$\alpha(B^1 h_1(k) + B^2 \overline{h_1}(k)) + (1 - \alpha)(B^3 \overline{h_2}(k) + B^4 h_2(k))$$

in which α is the probability that the system is in mode 1, $h_1(k)$ is the probability that the system succeeds in mode 1, $h_2(k)$ is the probability that the system fails in mode 2 and B^1, B^2, B^3, B^4 are cost-benefit factors. Pham [87] has studied the systems with units having three failure modes and dynamic configurations. Reliability functions have been derived for four different systems with dynamic configurations, and the methods to optimize system reliability and expected cost have been proposed. Torres-Echeverría et al. [102] have developed models for safety instrumented systems with M -out-of- N (MoonN) voting architectures with consideration of system reconfiguration during tests and repairs.

In some other papers, authors have used N modular redundant (NMR) systems or N -version programming to represent the similar idea of voting systems. Some work for NMR systems can be found in Mathur [71], Mathur and de Sousa [72, 73], Biernat [14], Radu et al. [92] and Namazi and Nourani [76]. TMR (Triple Modular Redundancy), which is the use of 2-out-of-3 voting concept [68], is the simplest version of NMR. It is very typical and classic with wide applications such as computers [21] and hardware [92]. For example, a TMR system has been used in the digital computer for Saturn V spacecraft launch vehicle [21], which is a very important application. Other research about TMR can be found Neumann [77], Avizienis [3], Mathur [70, 71], Pham [85], Kshirsagar and Patrikar [37], Ban and de Barros Naviner [8, 9] and Balasubramanian and Prasad [7]. Some work for N -version programming can be found in Chen and Avizienis [18], McAllister et al. [74], Goševa-Popstojanova and Grnarov [27], Xing et al. [111], Levitin et al. [54] and Levitin et al. [53]. In Xing et al. [111] and Levitin et al. [54], the authors have considered that some service component versions (SCV) may be corrupted and provide wrong outputs due to malicious actions from attackers. Minmax game problems with consideration of strategic attackers have been formulated to find the optimal numbers of SCVs, in which the users of the NVP system want to minimize the loss (or maximize the reward of success) while the attackers want to maximize the utility of attacks. In Levitin et al. [53], the authors have considered the security of data on the servers and a cancellation policy on task solver versions is implemented to reduce the chance of data theft.

Some other works for general voting systems include Srihari [99, 100], Stroud [101], Berg [13], Yacoub [112], Levitin and Hausken [49], Levitin et al. [50] and Li et al. [55]. Srihari [99, 100] has analyzed the reliability of majority voting systems with binary outputs, considering the conditional reliability given different correct outputs. The results have shown that the number of modules in the majority voting systems should not always be as large as possible to maximize the reliability. Yacoub [112] has presented a synthetic experimental procedure to study various performance indexes of majority voting systems. An enumerated simulation approach has been introduced and the results have been compared with existing mathematical models. Li et al. [55] have studied the reliability of an online majority voting system subject to multiple types of cyber-attacks, including targeted attacks, random attacks and

dynamic attacks. In targeted attacks, where the attacker selects the same set of components to attack multiple times, the reliability of the system decreases with the number of attacked components; in random attacks, where the attacker randomly selects a certain number of components to attack every time, the reliability of the system increases with the number of attacked components; in dynamic attacks, where the attacks arrive following a stochastic process, the reliability of the system is impacted by the attack rate and proposition time of the system.

2.2 Weighted Voting Systems

The reliability evaluation and optimization for weighted voting systems (WVS) is a very complex problem and a lot of research has been done in this particular area in the last decades. An algorithm has been proposed by Wu and Chen [108] to calculate the reliability of weighted k -out-of- n : G systems, with time complexity $O(kn)$. In addition, they have developed an algorithm to calculate the reliability of consecutive weighted k -out-of- n : F systems, based on the same assumptions, with time complexity $O(n)$ [109]. However, they have assumed that all the units can only be good or failed and all weights are positive integers. For reliability evaluation of WVS, the general problem formulation, which considers three failure modes for each voting unit, has been introduced by Nordmann and Pham [78], for both majority voting and threshold voting. They have also derived the formulas to evaluate the reliability of WVS and proved that the problem is NP hard. A simplified approach has been proposed by Nordmann and Pham [79] for reliability evaluation of WVS based on two restrictions: weights are integers and the threshold is a rational number. The general formula for reliability evaluation can be expressed as [79]:

$$R = \sum_{S_x \subseteq I} \left(\prod_{i \in S_x} q_x^{(i)} \right) \left(\prod_{i \notin S_x} (1 - q_x^{(i)}) \right) R(S_x)$$

where $R(S_x)$ is the reliability of the system given the set of indices of the units failing to stuck-at- x . In addition, a cost model has been developed. Their formulas can be used in most real situations. Xie and Pham [110] have proposed a recursive method to calculate the reliability of WVS without the two restrictions in [79]. In addition, they have developed a time dependent reliability function for WVS and a method to find the optimal stopping time, assuming that the durations of the decision time of all units are equal. To the best of our knowledge, this is one of the earliest papers that time issues are involved in reliability evaluation of voting systems.

Levitin and Lisnianski [51] have proposed a method to evaluate the reliability of WVS without the restrictions used by Nordmann and Pham [79]. Their method is based on a universal generating function (UGF) technique, which has been proved to be effective in numerical implementations [52]. The universal moment generating function for a discrete random variable X can be expressed as [51]:

$$U(z) = \sum_{k=1}^K s_k z^{x_k}, s_k = P(X = x_k)$$

In addition, they have developed an optimization tool using a genetic algorithm to find the optimal weights and voting threshold to maximize the system reliability. A series of research has been done by Levitin [40, 41, 43, 44] on WVS reliability evaluation for different extensions, including systems with units having limited availability [40], asymmetric weights [41], failures due to external impacts [44], different durations of decision time [45]. The methods to find the optimal weights and threshold have been developed for each situation respectively. Reliability evaluation for weighted voting wireless sensor networks with unreliable links using a UGF technique can be found in Liu et al. [65] and Liu and Zhang [63]. In their work, they have considered different levels of UGFs and used composition operators to combine them.

Some other interesting work on WVS can be found in Barbara and Garcia-Molina [10], Ebrahimipour et al. [23], Chen et al. [19], Alturki and Ali Rushdi [2], González-Pachón et al. [26] and Zhang and Pham [116, 117]. Barbara and Garcia-Molina [10] have proposed three heuristic methods to assign votes (weights) to maximize the reliability of network systems. Ebrahimipour et al. [23] have used the principal component analysis (PCA) to determine the weight of each performance index in the cost model, and applied a genetic algorithm to determine the optimal operation status for each unit, in the application of offshore pumps. Chen et al. [19] have studied the adjustment of weights based on the difference between individual preference and group preference. Zhang and Pham [116, 117] have introduced an indecisive parameter to WVS which can incorporate the weights of absent voting units during the final voting step. In another recent work, Liu and Zhang [64] have evaluated the reliability of WVS using Dempster-Shafer (D-S) evidence theory. In their work, an improved WVS has been introduced to fully utilize the information, in which the output of each voting unit is represented by a basic probability assignment (BPA) function and the voting rule of the system is replaced by Dempster combination rule.

Instead of binary inputs and outputs, the weighted voting system with discrete multi-state inputs and outputs, which is called weighted voting classifier (WVC), has been studied. WVCs have been used in many applications, such as facial expression recognition [123] and brain tumors classification [94]. For example, Zia et al. [123] have introduced a dynamic weight majority voting (DWMV) mechanism for spontaneous facial expression recognition; Rezaei et al. [94] have developed an integrated approach to segment three types of brain tumors by combining the results from three classifiers where the weights are determined by a multi-objective differential evolution (MODE) algorithm, which achieves an accuracy greater than 92%. Levitin [42, 43] has proposed a modified universal generating function technique to evaluate the reliability of WVC for both plurality and threshold systems. An optimization method has been proposed to minimize the cost of damage caused by classifier failure. Other work related to WVC can be found in Kim et al. [34], Goldberg and Eckstein [25], Zhang et al. [119] and Rojarath and Songpan [95]. Furthermore, Long et al. [66] have extended the work from discrete-state inputs to continuous-state inputs. In such

situation, the output of the system is the weighted average of the outputs for all the units and the reliability of the system is defined as the probability that the output of the system is within a certain range. The analytical expression has been derived under normality assumption and a Monte Carlo algorithm has been developed for general numerical problems.

3 Inspection Models

Inspections and maintenance are effective to improve system performance such as reliability, availability, performability and cost. In many recent studies, maintenance is performed during or after inspection and maintenance actions can be different for different inspection results. According to the time intervals between successive inspections, inspection policies can be categorized into periodic (equal intervals) inspection policies and sequential (unequal intervals) inspection policies, given the case of multiple inspections. Periodic inspection policy has been studied in many works such as Klutke and Yang [36], Van and Bérenguer [104], Dai et al. [20], Zhao et al. [120], Yang et al. [113] and Zhang et al. [118]. Klutke and Yang [36] have considered a system with continuous degradation and random shocks. The failures of the system are hidden and cannot be detected until periodic inspections, when the replacement can be performed. Van and Bérenguer [104] have studied a production system subject to periodic inspections and condition-based maintenance. The number of preventive maintenance actions are limited and the gain after preventive maintenance is considered to be deterministic or random (two models), but the maintenance may not make the system as good as new. Yang et al. [113] have studied a system with a periodic inspection and age replacement policy. The control limit for preventive replacement during inspection is shifting over time.

On the other hand, sequential inspection policy is another popular field of study. Numerous works include Chelbi and Ait-Kadi [17], Li and Pham [56], Ma et al. [69], Wang et al. [105], Yang et al. [114], Levitin et al. [46] and Levitin et al. [48]. In Li and Pham [56] and Ma et al. [69], the time intervals between successive inspections follow a geometric series and the time for n th inspection can be expressed as: $I_n = \sum_{j=1}^n \alpha^{j-1} I_1$ ($0 < \alpha \leq 1$), where I_1 is the time for the first inspection. Wang et al. [105] have proposed a two-phase inspection model which includes a semi-periodic inspection policy and a delayed preventive replacement time window. Specifically, inspections are performed at $T_1 + kT_2$ ($0 \leq k \leq n$) as well as $T_{\text{age}} - \tau_{\text{PR}}$ if there is no defect before, but there is no inspection after $T_{\text{age}} - \tau_{\text{PR}}$, where T_{age} is the time for renewal and τ_{PR} is the duration of delayed preventive replacement window. Yang et al. [114] have designed an inspection policy for a production system with regular monitoring and opportunistic monitoring. The regular monitoring is pre-determined at a fixed time while the opportunistic monitoring is performed randomly. Different preventive replacement limits are used for different types of monitoring and the regular monitoring will be rescheduled upon each replacement. In Levitin et al. [46] and Levitin et al. [48], an inspection and abort policy (IAP) is designed for

a system with a primary mission, and inspections will be implemented after each random shock to determine whether to abort the current mission and start the rescue procedure.

The inspection policies discussed above, whether deterministic or random, are “static”, which means that the time points of inspections in each cycle are independent from the historical inspection schedules and results within the same cycle. However, it is reasonable that in sequential inspections, the next inspection can be scheduled “dynamically” with consideration of current and/or previous inspection schedule(s) and/or result(s), which is the idea of state dependent or condition-based inspection policy.

Grall et al. [28] have proposed a condition-based inspection-maintenance policy. In their work, the state of the system is divided into $N + 1$ levels, from 0 to N , and the next inspection will be scheduled $N - l$ periods later given the level of system state for current inspection is $l(0 \leq l < N)$. Particularly, a replacement will be conducted if $l = N$ or failure. Their work has been extended to a two-unit system in Castanier et al. [16], where each component has a preventive replacement threshold and an opportunistic replacement threshold. An opportunistic replacement for a component will be triggered if the current state of this component reaches the opportunistic threshold and the other component is to be replaced after the inspection. Furthermore, an inspection scheduling function to determine the time for the next inspection is proposed by Grall et al. [29], which is a nonlinear decreasing function of system state during current inspection. The inspection scheduling function is also studied by Guo et al. [30]. In their work, three different types of functions (linear, convex and concave) have been compared, and the optimal scheduling function as well as the preventive replacement threshold have been determined by minimizing the expected cost rate with/without safety constraints respectively. Wang et al. [106] have designed a two-level inspection policy for a system with a three-stage (good, minor defective and severe defective) failure process, where minor and major inspections with different powers to detect defects, are scheduled periodically at the beginning. However, if a minor defect is found, the intervals between subsequent minor inspections will be shortened to half. Zhao et al. [121] have studied the optimization problem of an inspection and spare ordering policy, in which subsequent inspection intervals are halved if minor defects are detected during some inspection points.

Pham and Xie [90] have proposed a surveillance model by considering two mutually dependent stochastic processes for a system consisting of multiple subsystems. The first process is the state transition process of each subsystem, which the subsystem starts from a favorable state but may transfer to an unfavorable state after some time. The subsystem will remain in an unfavorable state until an inspector comes to visit this subsystem and fix the problem. The second process is the random surveillance process for the inspector. The surveillance process for a subsystem is a nonhomogeneous Poisson process (NHPP) with an intensity function depending on its surveillance history, which can be expressed as [90]:

$$\lambda_I(t) = \lambda_0 e^{G_k(t; \gamma)}, G_k(t; \gamma) = \gamma_1 w_k(t) + \gamma_2 I(t - t_{k-1} > T_2), t_{k-1} < t < t_k$$

where $w_k(t)$ is the number of unfavorable inspection results in the past and $I(t - t_{k-1} > T_2)$ is an indicator function that whether there is no visit for the subsystem in the past T_2 time. To the best of our knowledge, this is the first surveillance (inspection) model with consideration of state dependent random inspection process.

Instead of perfect inspections discussed above, which can identify system states correctly, some study of imperfect inspections can be found in Guo et al. [30], Peng et al. [84] and Levitin et al. [47]. In Guo et al. [30], a normally distributed random error in the measurement of system state during inspection has been considered. Peng et al. [84] have considered that the inspection can detect the defective state of the system with a probability. The expected long-run cost rate (ELRCR) is used as the metric to optimize the inspection interval. Levitin et al. [47] have considered both false positive and false negative errors for inspections when revealing the defective state of the system. A genetic algorithm is used to determine the optimal inspection schedule by maximizing mission success probability (MSP) given the desired level of failure avoidance probability (FAP).

4 Recent Studies

4.1 Motivation

As discussed in Sect. 2, many methods have been proposed for reliability evaluation and optimization for voting systems with applications in many different areas. However, we find that the consideration of time issues in voting systems is somehow limited. The time dependent reliability function for voting system was first proposed by Xie and Pham [110]. However, they have assumed that all the voting units have to make decisions at the same time. Levitin [45] has considered that the decision time of each unit in the system can be different. However, the decision time of each unit is still predetermined in that work. To the best of our knowledge, the study of voting systems with units having random decision time are very limited, but it is very common in human decision systems. In addition, the consideration of control and supervising mechanisms in voting systems is also very limited. From Sect. 3, many papers about periodic and sequential inspection policies have been published, but the study of state dependent or condition-based inspection policies is limited, especially for the applications in voting and decision-making systems.

4.2 Intervened Decision-Making Systems

Intervened Decision-making Systems are voting systems with decision units having random decision time, and supervising mechanisms during the decision process.

Supervising mechanism varies case by case, which can be as simple as some check points during the decision process and as complicated as an inspection and surveillance process.

The framework of a two-stage intervened decision-making system was first proposed by Lin and Pham [58, 59]. In their work, a triple-unit decision-making system with binary inputs and outputs for financial investment has been studied, which is shown in Fig. 1. In that system, each decision unit should provide a result within a time interval but the exact time is a random variable, and the final decision is made based on the results from all the units. Due to this randomness, it is possible that the final decision can be made earlier than expected. In this way, a check point is added to the decision process to check whether all the results are available to make the final decision, since it is not desirable to review the system continuously. At the check point, if all the results are available, the final decision will be made and the decision process will be terminated, which improves the efficiency of the system; otherwise, the units failing to provide the results by the check point will be monitored to make sure that all the results will be available within certain time limit. The two-stage intervened decision process of each unit is shown in Fig. 2. Based on this idea, a Two-stage Weighted Intervened Decision-making System has been studied in Lin and Pham [58], and a Two-stage Intervened Decision-making System with Interdependent Decision Units has been studied in Lin and Pham [59].

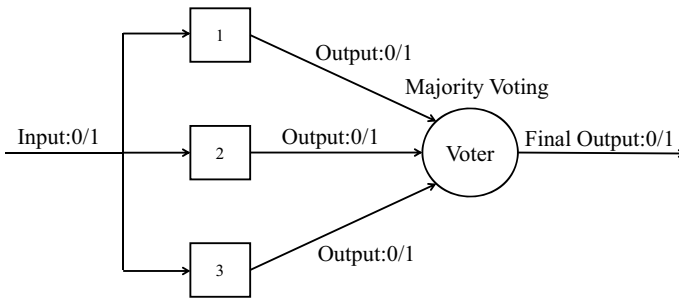


Fig. 1 Triple-unit decision-making system with binary input/output [59]

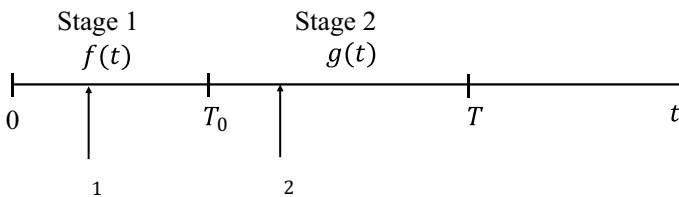


Fig. 2 Two-stage intervened decision process [59]

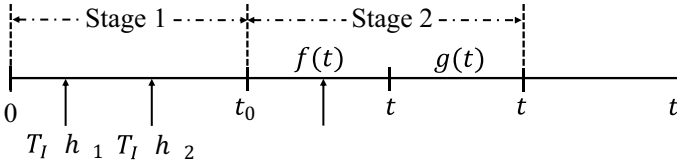


Fig. 3 Two-stage intervened decision process with random inspections [60]

Moreover, the supervising mechanism has been enhanced by adding one more stage for inspections at the beginning of the decision process (Fig. 3). This new system is called Two-stage Intervened Decision-making System with State Dependent Random Inspection Mechanisms, which has been studied in Lin and Pham [60]. Two dependent stochastic processes have been considered in this decision process: (1) state transition process of each decision unit, and (2) inspection process of the supervisor. The inspection of the supervisor may change the state of each decision unit, while the state of the decision unit during the inspection will also influence the subsequent behavior of the supervisor, which is the idea of state dependent inspection policy. Furthermore, a continued study on the system involving multi-state decision units and dynamic configuration can be found in Lin and Pham [61]. In that paper, some decision units may be removed from the system due to unfavorable inspection results, which leads to dynamic configuration of the system.

4.3 Numerical Example

For the system and the decision process described in Figs. 1 and 2, suppose that $f(t) = \lambda^2 t e^{-\lambda t} (t > 0)$ which is an Erlang distribution, $g(t) = \frac{1}{T-T_0}$ which is a uniform distribution. In addition, we assume that the time dependent probability function for each decision unit to provide a correct result is a linear function, which is $p(t) = a + bt (a + bT \leq 1)$. According to the total probability equation, the performability of each decision unit, which is the probability to give a correct result can be expressed as:

$$\begin{aligned}
 R_0(T_0, T) &= \int_0^{T_0} f(t)p(t)dt + (1 - F(T_0)) \int_{T_0}^T g(t)p(t)dt \\
 &= a(1 - e^{-\lambda T_0} - \lambda T_0 e^{-\lambda T_0}) + b \left(\frac{2}{\lambda} - \frac{2}{\lambda} e^{-\lambda T_0} - 2T_0 e^{-\lambda T_0} - \lambda T_0^2 e^{-\lambda T_0} \right) \\
 &\quad + \left(a + \frac{1}{2} b(T + T_0) \right) (e^{-\lambda T_0} + \lambda T_0 e^{-\lambda T_0}).
 \end{aligned}$$

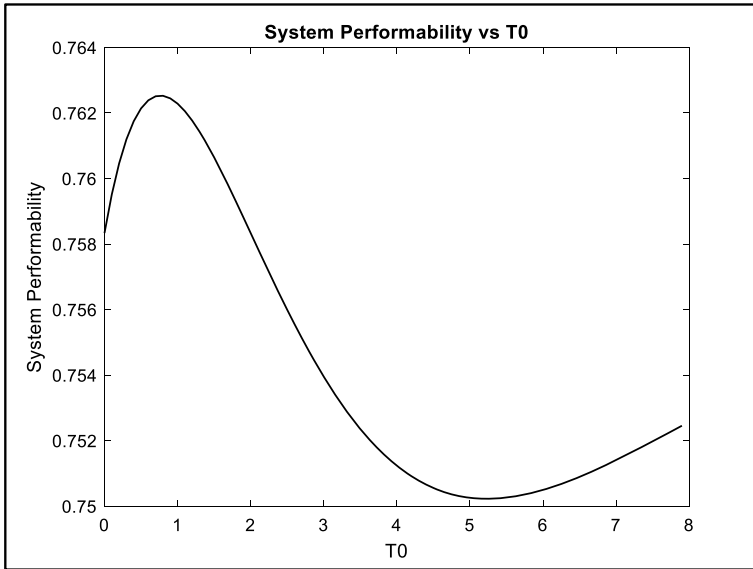


Fig. 4 System performability versus T_0 given $T = 8$

The performability of the system, which is the probability that the system will make a correct decision, can be expressed as:

$$R(T_0, T) = 3R_0(T_0, T)^2 - 2R_0(T_0, T)^3.$$

Given $\lambda = 0.5, a = 0.6, b = 0.02, T = 8$ and $T_0 = 4$, the performability of each decision unit and the system can be calculated as 0.6746 and 0.7512. If we wish to determine the optimal check point (T_0) by fixing $T = 8$ to maximize the system performability, the optimal value of T_0 should be $T_0^* = 0.76$ and the corresponding system performability is 0.7625. The detailed optimization procedure can be found in Lin [57] or Lin and Pham [58]. The behavior of the system performability function versus T_0 is shown in Fig. 4.

5 Conclusion and Future Research

In this chapter, a lot of previous works on voting systems and inspection models are discussed. Possible limitation and gaps have been figured out. Some recent studies about intervened decision making systems have been presented, which have considered random decision time of each unit and various supervising mechanisms during the decision process. This type of systems has many applications such as human organization, financial investment, banking systems, paper submission systems, proposal

review processes, and product recalls. From current work, many aspects can be considered in future research. For example, the input and output of system can be extended from binary to multiple discrete values and continuous values. In addition, time duration of inspections, multiple inspections & follow-ups, and state transitions during the whole decision process can be considered.

References

1. Alcantud JCR, Laruelle A (2014) Disapproval voting: a characterization. *Soc Choice Welfare* 43(1):1–10
2. Alturki AM, Ali Rushdi AM (2016) Weighted voting systems: a threshold-Boolean perspective. *J Eng Res* 4(1):1–19
3. Avizienis A (1967) Design fault-tolerant computers. In: *Proceedings of the November 14–16, 1967, fall joint computer of conference*, pp 733–743
4. Avizienis A, Gilley GC, Mathur FP, Rennels DA, Rohr JA, Rubin DK (1971) The STAR (self-testing and repairing) computer: an investigation of the theory and practice of fault-tolerant computer design. *IEEE Trans Comput* 100(11):1312–1321
5. Bai YW, Teng H, Shen LS (2009) Enhancement of the reliability of an embedded surveillance system by multiple sensors using a majority voting mechanism. In: *2009 IEEE instrumentation and measurement technology conference*. IEEE, pp 1157–1162
6. Bai YW, Cheng CC, Xie ZL (2013) Use of ultrasonic signal coding and PIR sensors to enhance the sensing reliability of an embedded surveillance system. In: *2013 IEEE international systems conference (SysCon)*. IEEE, pp 287–291
7. Balasubramanian P, Prasad K (2016) A fault tolerance improved majority voter for TMR system architectures. [arXiv:1605.03771](https://arxiv.org/abs/1605.03771)
8. Ban T, de Barros Naviner LA (2010) A simple fault-tolerant digital voter circuit in TMR nano architectures. In: *Proceedings of the 8th IEEE international NEWCAS conference 2010*. IEEE, pp 269–272
9. Ban T, de Barros Naviner LA (2012) Majority voter: signal probability, reliability and error bound characteristics. In: *2012 IEEE 55th international midwest symposium on circuits and systems (MWSCAS)*. IEEE, pp 538–541
10. Barbara D, Garcia-Molina H (1987) The reliability of voting mechanisms. *IEEE Trans Comput* 36(10):1197–1208
11. Bashir S, Qamar U, Khan FH (2015) Heterogeneous classifiers fusion for dynamic breast cancer diagnosis using weighted vote based ensemble. *Qual Q* 49(5):2061–2076
12. Ben-Dov Y (1980) Optimal reliability design of k-out-of-n-systems subject to two kinds of failure. *J Oper Res Soc* 31(8):743–748
13. Berg S (1997) Indirect voting systems: Banzhaf numbers, majority functions and collective competence. *Eur J Polit Econ* 13(3):557–573
14. Biernat J (1994) The effect of compensating faults models on NMR system reliability. *IEEE Trans Reliab* 43(2):294–300
15. Blough DM, Sullivan GF (1994) Voting using predispositions. *IEEE Trans Reliab* 43(4):604–616
16. Castanier B, Grall A, Bérenguer C (2005) A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliab Eng Syst Saf* 87(1):109–120
17. Chelbi A, Ait-Kadi D (1999) An optimal inspection strategy for randomly failing equipment. *Reliab Eng Syst Saf* 63(2):127–131
18. Chen L, Avizienis A (1978) N-version programming: a fault-tolerance approach to reliability of software operation. In: *Proceedings of the 8th IEEE international symposium on fault-tolerant computing (FTCS-8)*, vol 1, pp 3–9

19. Chen YK, Zou Y, Chen Z (2014) Preference integration and optimization of multistage weighted voting system based on ordinal preference. *Abstr Appl Anal* 2014 (Hindawi)
20. Dai Y, Levitin G, Xing L (2015) Optimal periodic inspections and activation sequencing policy in standby systems with condition-based mode transfer. *IEEE Trans Reliab* 66(1):189–201
21. Dickinson MM, Jackson JB, Randa GC (1964) Saturn V launch vehicle digital computer and data adapter. In: *Proceedings of the October 27–29, 1964, fall joint computer conference, part I*, pp 501–516
22. Dimitriadou E, Weingessel A, Hornik K (1999) Voting in clustering and finding the number of clusters
23. Ebrahimipour V, Azadeh A, Roohi SF, Shojaei E, Aalaei A (2008) A PCA-GA approach for weighted voting system optimization based on reliability, cost and system output analyses. In: *2008 IEEE international conference on industrial engineering and engineering management*. IEEE, pp 566–570
24. Gogiashvili JG, Namicheishvili OM, Shonia G (2000) Optimization of weights for threshold redundancy of binary channels by the method of (Mahalanobis') generalized distance. In: *Second international conference on mathematical methods in reliability: methodology, practice and interference*, pp 4–7
25. Goldberg N, Eckstein J (2012) Sparse weighted voting classifier selection and its linear programming relaxations. *Inf Process Lett* 112(12):481–486
26. González-Pachón J, Diaz-Balteiro L, Romero C (2019) A multi-criteria approach for assigning weights in voting systems. *Soft Comput* 23(17):8181–8186
27. Goševa-Popstojanova K, Grnarov A (1993) N version programming with majority voting decision: dependability modeling and evaluation. *Microprocess Microprogramm* 38(1–5):811–818
28. Grall A, Bérenguer C, Dieulle L (2002) A condition-based maintenance policy for stochastically deteriorating systems. *Reliab Eng Syst Saf* 76(2):167–180
29. Grall A, Dieulle L, Bérenguer C, Roussignol M (2002) Continuous-time predictive-maintenance scheduling for a deteriorating system. *IEEE Trans Reliab* 51(2):141–150
30. Guo C, Bai Y, Peng R (2019) Condition-based maintenance optimization with safety constraints under imperfect inspection. In: *2019 prognostics and system health management conference (PHM-Qingdao)*. IEEE, pp 1–6
31. Hillinger C (2005) The case for utilitarian voting. *Homo Oeconomicus* 22(3)
32. Jenney BW, Sherwin DJ (1986) Open & short circuit reliability of systems of identical items. *IEEE Trans Reliab* 35(5):532–538
33. Kim H, Jeon HJ, Lee K, Lee H (2002) The design and evaluation of all voting triple modular redundancy system. In *Annual reliability and maintainability symposium. 2002 proceedings* (Cat. No. 02CH37318). IEEE, pp 439–444
34. Kim H, Kim H, Moon H, Ahn H (2011) A weight-adjusted voting algorithm for ensembles of classifiers. *J Korean Stat Soc* 40(4):437–449
35. Kim H, Lee H, Lee K (2005) The design and analysis of AVTMR (all voting triple modular redundancy) and dual-duplex system. *Reliab Eng Syst Saf* 88(3):291–300
36. Klutke GA, Yang Y (2002) The availability of inspected systems subject to shocks and graceful degradation. *IEEE Trans Reliab* 51(3):371–374
37. Kshirsagar RV, Patrikar RM (2009) Design of a novel fault-tolerant voter circuit for TMR implementation to improve reliability in digital circuits. *Microelectron Reliab* 49(12):1573–1577
38. Latif-Shabgahi G, Bennett S (1999) Adaptive majority voter: a novel voting algorithm for real-time fault-tolerant control systems. In: *Proceedings 25th EUROMICRO conference. Informatics: theory and practice for the new millennium, vol 2*, pp 113–120. IEEE
39. Lee M, Pham H, Zhang X (1999) A methodology for priority setting with application to software development process. *Eur J Oper Res* 118(2):375–389
40. Levitin G (2001) Analysis and optimization of weighted voting systems consisting of voting units with limited availability. *Reliab Eng Syst Saf* 73(1):91–100
41. Levitin G (2002) Asymmetric weighted voting systems. *Reliab Eng Syst Saf* 76(2):205–212

42. Levitin G (2002) Evaluating correct classification probability for weighted voting classifiers with plurality voting. *Eur J Oper Res* 141(3):596–607
43. Levitin G (2003) Threshold optimization for weighted voting classifiers. *Naval Res Logist (NRL)* 50(4):322–344
44. Levitin G (2004) Maximizing survivability of vulnerable weighted voting systems. *Reliab Eng Syst Saf* 83(1):17–26
45. Levitin G (2005) Weighted voting systems: reliability versus rapidity. *Reliab Eng Syst Saf* 89(2):177–184
46. Levitin G, Finkelstein M, Dai Y (2020) State-based mission abort policies for multistate systems. *Reliab Eng Syst Saf* 204:107122
47. Levitin G, Finkelstein M, Huang HZ (2019) Scheduling of imperfect inspections for reliability critical systems with shock-driven defects and delayed failures. *Reliab Eng Syst Saf* 189:89–98
48. Levitin G, Finkelstein M, Xiang Y (2021) Optimal inspections and mission abort policies for multistate systems. *Reliab Eng Syst Saf* 214:107700
49. Levitin G, Hausken K (2013) Defending threshold voting systems with identical voting units. *IEEE Trans Reliab* 62(2):466–477
50. Levitin G, Hausken K, Haim HB (2013) Defending majority voting systems against a strategic attacker. *Reliab Eng Syst Saf* 111:37–44
51. Levitin G, Lisnianski A (2001) Reliability optimization for weighted voting system. *Reliab Eng Syst Saf* 71(2):131–138
52. Levitin G, Xie M, Zhang T (2007) Reliability of fault-tolerant systems with parallel task processing. *Eur J Oper Res* 177(1):420–430
53. Levitin G, Xing L, Dai Y (2021) Security and reliability of N-version cloud-based task solvers with individual version cancellation under data theft attacks. *Reliab Eng Syst Saf* 216:107920
54. Levitin G, Xing L, Xiang Y (2020) Optimization of time constrained N-version programming service components with competing task execution and version corruption processes. *Reliab Eng Syst Saf* 193:106666
55. Li Y, Hu X, Zhao P (2021) On the reliability of a voting system under cyber attacks. *Reliab Eng Syst Saf* 107996
56. Li W, Pham H (2005) An inspection-maintenance model for systems with multiple competing processes. *IEEE Trans Reliab* 54(2):318–327
57. Lin T (2022) Performability modeling and cost-benefit analysis for decision-making systems with supervising mechanisms (Doctoral dissertation, Rutgers University-School of Graduate Studies)
58. Lin T, Pham H (2017) Two-stage weighted intervened decision systems. *Life Cycl Reliab Saf Eng* 6(1):69–77
59. Lin T, Pham H (2019) Reliability and cost-benefit analysis for two-stage intervened decision-making systems with interdependent decision units. *Int J Math Eng Manag Sci* 4(3):531–541
60. Lin T, Pham H (2019) A two-stage intervened decision system with state-dependent random inspection mechanisms. *IEEE Trans Comput Soc Syst* 6(2):365–376
61. Lin T, Pham H (2019) A two-stage intervened decision system with multi-state decision units and dynamic system configuration. *Ann Oper Res*. <https://doi.org/10.1007/s10479-019-03334-8>
62. Liu Y, Lei H, Zhang D, Wu Z (2018) Robust optimization for relief logistics planning under uncertainties in demand and transportation time. *Appl Math Model* 55:262–280
63. Liu Q, Zhang H (2017) Weighted voting system with unreliable links. *IEEE Trans Reliab* 66(2):339–350
64. Liu Q, Zhang H (2021) Reliability evaluation of weighted voting system based on D–S evidence theory. *Reliab Eng Syst Saf* 108079
65. Liu Q, Zhang H, Ma Y (2016) Reliability evaluation for wireless sensor network based on weighted voting system with unreliable links. In 2016 3rd international conference on information science and control engineering (ICISCE). IEEE, pp 1384–1388
66. Long Q, Xie M, Ng SH, Levitin G (2008) Reliability analysis and optimization of weighted voting systems with continuous states input. *Eur J Oper Res* 191(1):240–252

67. Lorczac PR, Caglayan AK, Eckhardt DE (1989) A theoretical investigation of generalized voters for redundant systems. In: [1989] The nineteenth international symposium on fault-tolerant computing. Digest of papers. IEEE, pp 444–451
68. Lyons RE, Vanderkulk W (1962) The use of triple-modular redundancy to improve computer reliability. *IBM J Res Dev* 6(2):200–209
69. Ma X, Wang W, Liu X, Peng R (2017) Optimal inspection and replacement strategy for systems subject to two types of failures with adjustable inspection intervals. *J Shanghai Jiaotong Univ (Science)* 22(6):752–755
70. Mathur FP (1969) Reliability modeling and analysis of a dynamic TMR system utilizing standby spares. In *Proceedings of the 7th annual allerton conference on circuit and system theory*, pp 243–252
71. Mathur FP (1971) On reliability modeling and analysis of ultrareliable fault-tolerant digital systems. *IEEE Trans Comput* 100(11):1376–1382
72. Mathur FP, de Sousa PT (1975) Reliability models of NMR systems. *IEEE Trans Reliab* 24(2):108–113
73. Mathur FP, de Sousa PT (1975) Reliability modeling and analysis of general modular redundant systems. *IEEE Trans Reliab* 24(5):296–299
74. McAllister DF, Sun CE, Vouk MA (1990) Reliability of voting in fault-tolerant software systems for small output-spaces. *IEEE Trans Reliab* 39(5):524–534
75. Mine H, Hatayama K (1981) Reliability analysis and optimal redundancy for majority-voted logic circuits. *IEEE Trans Reliab* 30(2):189–191
76. Namazi A, Nourani M (2007) Reliability analysis and distributed voting for NMR nanoscale systems. In: *2007 2nd international design and test workshop*. IEEE, pp 130–135
77. Neumann JV (1956) Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Autom Stud* 34:43–98
78. Nordmann L, Pham H (1997) Reliability of decision making in human-organizations. *IEEE Trans Syst Man Cybern-Part A: Syst Hum* 27(4):543–549
79. Nordmann L, Pham H (1999) Weighted voting systems. *IEEE Trans Reliab* 48(1):42–49
80. Onan A, Korukoğlu S, Bulut H (2016) A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. *Expert Syst Appl* 62:1–16
81. Page LB, Perry JE (1988) Optimal ‘series-parallel’ networks of 3-state devices. *IEEE Trans Reliab* 37(4):388–394
82. Parhami B (1994) Threshold voting fundamentally simpler than plurality voting. *Int J Reliab Qual Saf Eng* 1(01):95–102
83. Parhami B (1994) Voting algorithms. *IEEE Trans Reliab* 43(4):617–629
84. Peng R, Liu B, Zhai Q, Wang W (2019) Optimal maintenance strategy for systems with two failure modes. *Reliab Eng Syst Saf* 188:624–632
85. Pham H (1993) Optimal cost-effective design of triple-modular-redundancy-with-spares systems. *IEEE Trans Reliab* 42(3):369–374
86. Pham H (1997) Reliability analysis of digital communication systems with imperfect voters. *Math Comput Model* 26(4):103–112
87. Pham H (1999) Reliability analysis for dynamic configurations of systems with three failure modes. *Reliab Eng Syst Saf* 63(1):13–23
88. Pham H, Malon DM (1994) Optimal design of systems with competing failure modes. *IEEE Trans Reliab* 43(2):251–254
89. Pham H, Pham M (1991) Optimal designs of $(k, n-k+1)$ -out-of- n : F systems (subject to 2 failure modes). *IEEE Trans Reliab* 40(5):559–562
90. Pham H, Xie M (2002) A generalized surveillance model with applications to systems safety. *IEEE Trans Syst Man Cybern Part C (Appl Rev)* 32(4):485–492
91. Qiu Q, Cui L, Gao H, Yi H (2018) Optimal allocation of units in sequential probability series systems. *Reliab Eng Syst Saf* 169:351–363
92. Radu M, Pitica D, Posteuca C (2000) Reliability and failure analysis of voting circuits in hardware redundant design. In *International symposium on electronic materials and packaging (EMAP2000)* (Cat. No. 00EX458). IEEE, pp 421–423

93. Randhawa K, Loo CK, Seera M, Lim CP, Nandi AK (2018) Credit card fraud detection using AdaBoost and majority voting. *IEEE access* 6:14277–14284
94. Rezaei K, Agahi H, Mahmoodzadeh A (2020) A weighted voting classifiers ensemble for the brain tumors classification in MR images. *IETE J Res* 1–14
95. Rojarath A, Songpan W (2021) Cost-sensitive probability for weighted voting in an ensemble model for multi-class classification problems. *Appl Intell* 1–25
96. Sah RK, Stiglitz JE (1984) The architecture of economic systems: hierarchies and polyarchies (No. w1334). National Bureau of Economic Research
97. Sah RK, Stiglitz JE (1988) Qualitative properties of profit-making k-out-of-n systems subject to two kinds of failures. *IEEE Trans Reliab* 37(5):515–520
98. Soroudi A, Amraee T (2013) Decision making under uncertainty in energy systems: state of the art. *Renew Sustain Energy Rev* 28:376–384
99. Srihari SN (1982) Reliability analysis of biased majority-vote systems. *IEEE Trans Reliab* 31(1):117–118
100. Srihari SN (1982) Reliability analysis of majority vote systems. *Inf Sci* 26(3):243–256
101. Stroud CE (1994) Reliability of majority voting based VLSI fault-tolerant circuits. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 2(4):516–521
102. Torres-Echeverría AC, Martorell S, Thompson HA (2011) Modeling safety instrumented systems with MooN voting architectures addressing system reconfiguration for testing. *Reliab Eng Syst Saf* 96(5):545–563
103. Torres-Echeverría AC, Martorell S, Thompson HA (2012) Multi-objective optimization of design and testing of safety instrumented systems with MooN voting architectures using a genetic algorithm. *Reliab Eng Syst Saf* 106:45–60
104. Van PD, Bérenguer C (2012) Condition-based maintenance with imperfect preventive repairs for a deteriorating production system. *Qual Reliab Eng Int* 28(6):624–633
105. Wang H, Wang W, Peng R (2017) A two-phase inspection model for a single component system with three-stage degradation. *Reliab Eng Syst Saf* 158:31–40
106. Wang W, Zhao F, Peng R (2014) A preventive maintenance model with a two-level inspection policy based on a three-stage failure process. *Reliab Eng Syst Saf* 121:207–220
107. Willsky AS (1976) A survey of design methods for failure detection in dynamic systems. *Automatica* 12(6):601–611
108. Wu JS, Chen RJ (1994) An algorithm for computing the reliability of weighted-k-out-of-n systems. *IEEE Trans Reliab* 43(2):327–328
109. Wu JS, Chen RJ (1994) Efficient algorithms for k-out-of-n and consecutive-weighted-k-out-of-n: F system. *IEEE Trans Reliab* 43(4):650–655
110. Xie M, Pham H (2005) Modeling the reliability of threshold weighted voting systems. *Reliab Eng Syst Saf* 87(1):53–63
111. Xing L, Levitin G, Xiang Y (2019) Defending N-version programming service components against co-resident attacks in IoT cloud systems. *IEEE Trans Serv Comput*
112. Yacoub S (2003) Analyzing the behavior and reliability of voting systems comprising tri-state units using enumerated simulation. *Reliab Eng Syst Saf* 81(2):133–145
113. Yang L, Zhao Y, Peng R, Ma X (2018) Hybrid preventive maintenance of competing failures under random environment. *Reliab Eng Syst Saf* 174:130–140
114. Yang L, Zhao Y, Peng R, Ma X (2018) Opportunistic maintenance of production systems subject to random wait time and multiple control limits. *J Manuf Syst* 47:12–24
115. Zhang Y (2014) Modeling the effects of the two stochastic-processes on the reliability and maintenance of k-out-of-n surveillance systems. Rutgers The State University of New Jersey-New Brunswick
116. Zhang H, Pham H (2019) Modeling reliability of threshold weighted indecisive voting systems. *IEEE Trans Comput Soc Sys* 7(1):35–41
117. Zhang H, Pham H (2021) Dynamic process in threshold weighted indecisive-voting systems. *IEEE Trans Comput Soc Syst*
118. Zhang A, Zhang T, Barros A, Liu Y (2020) Optimization of maintenances following proof tests for the final element of a safety-instrumented system. *Reliab Eng Syst Saf* 196:106779

119. Zhang Y, Zhang H, Cai J, Yang B (2014) A weighted voting classifier based on differential evolution. *Abstr Appl Anal* 2014 (Hindawi)
120. Zhao F, Liu X, Peng R (2018) Inspection-based policy considering human errors for three-stage delay time degradation systems. *J Shanghai Jiaotong Univ (Science)* 23(5):702–706
121. Zhao F, Liu X, Peng R, Kang J (2020) Joint optimization of inspection and spare ordering policy with multi-level defect information. *Comput Ind Eng* 139:106205
122. Zhou J, Huang N, Sun X, Xing L, Zhang S (2015) Network resource reallocation strategy based on an improved capacity-load model. *EksplatacjaiNiezawodność* 17(4):487–495
123. Zia MS, Hussain M, Jaffar MA (2018) A novel spontaneous facial expression recognition using dynamically weighted majority voting based ensemble classifier. *Multimed Tools Appl* 77(19):25537–25567
124. Zuev YA, Ivanov SK (1999) The voting as a way to increase the decision reliability. *J Franklin Inst* 336(2):361–378

Tingnan Lin Rutgers University, New Jersey, USA

Dr. Tingnan Lin received B.S. degree in Industrial Engineering from Tsinghua University, Beijing, China in 2013, M.S. in Industrial and Systems Engineering in 2016, M.S. in Statistics in 2017, and Ph.D. in Industrial and Systems Engineering in 2022 from Rutgers University, Piscataway, USA. His research interests include reliability and performability modeling and optimization for decision-making systems.

Assessing the Severity of COVID-19 in the United States



Kehan Gao, Sarah Tasneem, and Taghi Khoshgoftaar

Abstract The paper proposes a unique set of relatively simple-to-construct-and-use metrics to track the spread and severity of COVID-19 cases in the United States with respect to state-wise distribution. To our knowledge, such metrics have not been proposed by others. The proposed metrics and case study presented in the paper can serve as a template for public health officials to update the disease's severity and spread across states. Consequently, effective resources can be directed to the hard-hit states. Since the number of infections and deaths due to COVID-19 is dynamic, the proposed metrics along with the case study's approach can be easily updated to match new available data on the disease. The primary message of our work is to provide a unique perspective into investigating the current severity and spread of the pandemic in different states in the country and highlight states with a relatively high severity score.

Keywords COVID-19 · Pandemic · Ratio-based metrics · Severity score

1 Introduction

The world has been experiencing an unprecedented public health crisis for more than two years due to the ongoing COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. The United States is one of the countries severely affected by this disease, with 83,949,036 confirmed cases

K. Gao (✉) · S. Tasneem

Department of Computer Science, Eastern Connecticut State University, 83 Windham Street, Willimantic 06226, CT, USA

e-mail: gaok@easternct.edu

S. Tasneem

e-mail: tasneems@easternct.edu

T. Khoshgoftaar

Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, 777 Glades Road, Boca Raton 33431, FL, USA

e-mail: khoshgof@fau.edu

and 1,002,067 deaths reported by the Centers for Disease Control and Prevention (CDC) as of May 31, 2022 [2]. The interest of this research is to investigate the data obtained from the Johns Hopkins Coronavirus Resource Center [3] and to determine the severity of COVID-19 in U.S. states over the past 2-plus years.

To assess the severity (or impact) of COVID-19 in an area, many common epidemiological metrics are used, including the number of positive cases and deaths in the area, incidence, prevalence, and mortality rate [4]. Each metric plays a unique role in evaluating the effects of COVID-19. When comparing the epidemic situation between two regions, one must be careful because a single metric can sometimes mislead on the state of the epidemic. For example, California's death toll accounted for 9.12% of U.S. COVID-19 deaths, higher than New York's 6.88% (as of May 31, 2022). However, taking into account the two states' 12.04% to 5.93% population ratio, New York's adjusted death rate was 355 deaths per 100K, higher than California's 231 deaths per 100K. This shows that the population proportion is an important factor that cannot be ignored.

In this paper, we present common epidemiological metrics, such as positive rate and death rate, and propose four new ratio-based metrics, namely "positive percent/population percent", "death percent/population percent", "mortality ratio" and "severity score." The severity score metric is an aggregation of the other three metrics and one that can determine the severity level of COVID-19. To the best of our knowledge, such metrics have not been proposed by others. We applied these metrics to analyze COVID-19 data for all U.S. states and provided a severity score for each state accordingly. The proposed metrics have two main advantages: (1) These new metrics inherit all the properties and characteristics of the conventional metrics in a relatively simple format, so that it can provide public health authorities with direct and accurate judgments on the epidemic's status. Practitioners can refer to these metrics and take effective measures to adequately monitor and control the spread of COVID-19; and (2) These new metrics can be easily constructed and adjusted according to different requirements, including changing data. Our data analysis and results show that the five states most affected by COVID-19 in 2020 are New Jersey, New York, North Dakota, Rhode Island, and Connecticut, and in 2021 they are Oklahoma, Alabama, West Virginia, Kentucky, and Arizona, while the five states most affected by the Omicron variant this year as of May 31 are Oklahoma, New Mexico, West Virginia, Kentucky, and Tennessee.

The rest of the paper is organized as follows. Section 2 presents the related work. Section 3 discusses methods used in this study, including the conventional metrics and newly proposed metrics (see Sect. 3.3). The data analysis and results are presented in Sect. 4. Section 5 covers the discussion. Finally, the conclusion and future work are summarized in Sect. 6.

2 Related Work

At present, there is a large amount of research on COVID-19 from different perspectives, including effectiveness of COVID-19 vaccines [5, 6], COVID-19 trend prediction [7, 8], SARS-CoV-2 variants [9–11], and COVID-19 therapeutics [12]. Our study focuses on evaluating the severity of COVID-19. The Pandemic Severity Assessment Framework (PSAF) [13] is one of two assessment tools developed by CDC to guide and coordinate actions among federal, state, local, and tribal entities involved in pandemic response. Public health officials use the PSAF to determine the impact of the pandemic. There are two main factors that can be used to determine the impact of a pandemic: (1) clinical severity, that is, how serious is the illness associated with infection, and (2) transmissibility, that is, how easily the pandemic virus spreads from person to person. The combination of these two factors is used to guide the decision on what actions CDC recommends at specific times during the pandemic [14]. In addition, the World Health Organization (WHO) believes that severity assessments should be carried out in the early stage of a pandemic and regularly thereafter as the pandemic evolves. Since the World Health Assembly highlighted this need especially after H1N1 in 2009, WHO has made great progress in developing a framework on pandemic influenza severity assessment (PISA) [15]. The framework defines the severity of influenza based on three indicators: transmission, seriousness of disease, and impact.

In this paper, we propose four new ratio-based metrics, one of which called “severity score” integrates the other three metrics. These metrics reflect clinical severity and transmissibility of infectious diseases. The severity of COVID-19 in a state can be assessed by comparing the new indicators with a simple uniform benchmark (for example, a state’s status relative to the U.S. national average).

3 Methodology

This study investigates the COVID-19 data obtained from the Center for Systems Science and Engineering at Johns Hopkins University (JHU CSSE) [3], as of May 31, 2022. JHU CSSE provides daily time series summary tables, including confirmed positives and deaths. The time series tables are subject to updates if inaccuracies are identified in the historical data. For population data, we use the estimates of the resident population for the states provided by the U.S. Census Bureau, Population Division [16], as of July 1, 2019. Table 1 lists the metrics of data related to our case study.

Table 1 COVID-19 metrics with descriptions

Metric	Description
Confirmed	Aggregated case count for the state [3]
Deaths	Aggregated death toll for the state [3]
Incident_Rate	Cases per 100,000 (100K) persons [3]
Case_Fatality_Rate	Number recorded deaths * 100/Number confirmed cases [3]
Population	Estimates of the resident population for the states July 1, 2019 [16]

3.1 Basic Metrics

Laboratory data elements may be reported in different ways [17]: (1) Directly sent to state or local public health organizations according to state/or local law or policy; (2) Submitted to state and local public health organizations through a centralized platform (such as the Association of Public Health Laboratories' AIMS platform [18]), where the data will then be routed to the appropriate state and local authorities and routed to CDC after removal of Personally Identifiable Information (PII) according to applicable rules and regulations; (3) Submitted through a state or regional Health Information Exchange (HIE) to the appropriate state or local public health organizations and then to the CDC as directed by the state.

Basic metrics refer to metrics that provide a brief summary (or information) of laboratory data related to epidemics. These metrics usually appear in dashboards at various levels [2, 19]. Commonly used basic metrics include:

1. The number of confirmed positive cases in a specific time period (e.g., reported within the last 24 hours)
2. The number of deaths in a specific time period
3. The number of people tested in a specific time period
4. The number of recovered patients in a specific time period
5. The number of hospital admissions in a specific time period
6. The population of a region (city, county, state, region, country, continent, world)
7. Location of the region (e.g., longitude and latitude).

The last two metrics are very important for epidemiology, because public health professionals want to know the geographic distribution of epidemics and their relationship to population distribution. The basic metrics used in this study include metrics 1, 2, 6 and 7 as shown above.

3.2 *Derived Metrics*

In addition to the above basic metrics, there are some derived metrics which are computed from basic metrics. We list the derived metrics used in this study as follows:

1. **Positive%:** Positive proportion (e.g., the number of positives reported in NY (5,437,811) accounted for 6.52% of the total number of positives reported in the U.S. (83,453,922))
2. **Death%:** Death proportion (e.g., the number of deaths reported in NY (68,989) accounted for 6.88% of the total number of deaths reported in the U.S. (1,002,125))
3. **Population%:** Population proportion (e.g., NY population (19,453,561) accounted for 5.93% of the U.S. population (328,239,523))
4. **Positives per 100K:** The number of confirmed positives per 100,000 persons (e.g., in NY, the number of positives (5,437,811) divided by the NY population (19,453,561) and then multiplied by 100K, to get 27,953)
5. **Deaths per 100K:** The number of deaths per 100,000 persons (e.g., in NY, the number of deaths (68,989) divided by the NY population (19,453,561) and then multiplied by 100K, to get 355)
6. **Case Fatality Rate (CFR):** The number of death cases in 100 confirmed positive cases (e.g., in NY, the number of deaths (68,989) divided by the number of positives (5,437,811), to get 1.27%).

Please note that the NY COVID-19 data provided above is as of May 31, 2022.

3.3 *Four New Proposed Metrics*

For infectious diseases, there are two key indicators to determine the severity of the disease. One is to measure how fast the virus is growing (spreading). It is the average number of people who become infected by an infectious person, usually denoted by R_t [20]. If R_t is above 1.0, the virus will spread quickly. When R_t is below 1.0, the virus will stop spreading. Since the R_t index is a dynamically changing indicator, it is difficult to obtain an accurate R_t value at a specific point in time. Usually, the average value of the index can be obtained in a specific time interval. An alternative metric is the incidence rate over a period of time in the past (e.g., the number of confirmed positives per 100K last week). A high incidence rate implies that the virus has grown or spread rapidly in the past period of time. Another measure of disease severity is the case fatality rate (CFR), which is the ratio of the number of deaths from a specific disease to the total number of people diagnosed with the disease during a specific period [4]. Although both R_t and CFR are important, they cannot accurately reflect the overall severity of an epidemic in a region when used individually. In this study, we proposed four ratio-based metrics as shown below to help assess the severity of COVID-19.

- **Positive%/Population%:** The ratio of Positive% to Population% (e.g., in NY, $6.52\%/5.93\% = 1.10$)
- **Death%/Population%:** The ratio of Death% to Population% (e.g., in NY, $6.88\%/5.93\% = 1.16$)
- **Mortality Ratio:** The ratio of CFR of a region to the average CFR of a larger population that includes the region (e.g., in NY, the NY CFR (1.27%) divided by the U.S. CFR (1.20%), to get 1.06)
- **Severity Score:** The average of the above three metrics, namely, Positive%/Population%, Death%/Population%, and Mortality Ratio metrics (e.g., in NY, the severity score is $(1.10 + 1.16 + 1.06)/3 = 1.11$).

The first two metrics are equivalent to the frequently used “Positives per 100K” and “Deaths per 100K”, respectively, but provide more information and show the severity of the epidemic in terms of the proportion of positives and deaths relative to the population. For example, in NY, Positive%/Population% = 1.10 and Death%/Population% = 1.16, which means that in NY, the number of positive cases and the death toll are 1.1 times and 1.16 times, respectively, the state’s share of the U.S. population. This suggests that New York has higher positive cases and deaths than the U.S. average. As another example, in California (CA), these two values are 0.96 and 0.76, which means that the number of positive cases and deaths from COVID-19 in CA is lower than the national average.

Next, we would like to point out why these two ratio-based metrics are equivalent to “Positives per 100K” and “Deaths per 100K”, respectively. We use the “Positive%/Population%” metric with respect to “Positives per 100K” as an example to indicate why they are equivalent. According to the definition of Positive%/Population%, it is

$$\frac{\text{Positive}\%}{\text{Population}\%} = \frac{\text{Positives}}{\text{Total Positives}} \bigg/ \frac{\text{Population}}{\text{Total Population}} \quad (1)$$

which is equivalent to

$$\frac{\text{Positive}\%}{\text{Population}\%} \times \frac{\text{Total Positives}}{\text{Total Population}} = \frac{\text{Positives}}{\text{Population}} \quad (2)$$

namely,

$$\frac{\text{Positive}\%}{\text{Population}\%} \times \text{Total positives per 100K} = \text{Positives per 100K} \quad (3)$$

As seen from Eq. 3, “Positive%/Population%” is equivalent to “Positives per 100K”. The only difference is the additional term “Total positives per 100K” on the left side of the equation, which is constant at a time point or time period. This also provides another way to calculate “Positives per 100K”, by multiplying “Positive%/Population%” with “Total positives per 100K”. For example, in NY, Positive%/Population% = 1.10 and Death%/Population% = 1.16, which means that

the incidence of COVID-19 in NY (27,953 per 100K) is 1.10 times the U.S. national average (25,425 per 100K), and the death rate (355 per 100K) is 1.16 times the national average (305 per 100K).

The same approach is applied to the mortality ratio. We define Mortality Ratio as the ratio of the CFR to the overall CFR. As mentioned above, the mortality ratio of NY is 1.06, which means that the CFR in NY (1.27%) is 1.06 times of the national CFR (1.20%).

Finally, we introduce the fourth new metric, called Severity Score, which combines the positive rate, mortality, and case fatality rate (CFR). The higher the severity score, the more severe the consequences of an epidemic in a given area. The strategy used here is to combine several useful metrics to form a unique comprehensive metric, thus providing more constructive information for public health practice. Further analysis will be provided in the next section. In addition, this strategy is very flexible. For example, when evaluating severity score, practitioners can select metrics that they are interested in or wish to consider, and then aggregate them together. Furthermore, when a certain metric is more critical than others, a weighted average can be used to calculate the final severity score. Finally, healthcare management can classify instances (such as states) into different categories based on their severity scores and take effective disease preventive and mitigation measures accordingly.

4 Results

Since the proposed metrics are ratio-based, a benchmark must be given or selected first. In this study, we aim to investigate the severity of COVID-19 in all U.S. states. Therefore, we chose the U.S. national average as the benchmark, and then compared each state with this benchmark to determine the severity of the epidemic in that state.

We first provide the overall performance of the United States on COVID-19. Table 2 and Fig. 1 show incidence (cases per 100K), mortality (deaths per 100K) and CFR in the U.S. Each includes cumulative and monthly results. For example, the cumulative cases per 100K in Apr-20 refers to the total number of positive cases per 100K from January 2020 to the end of April 2020, while the monthly cases per 100K in Apr-20 represents the number of positive cases detected only in April 2020. Note that reported cases (deaths) per 100K are rounded to the nearest whole number and percent and ratio-based values are rounded to two decimal places. As can be seen from the monthly data, in the spring and summer of 2020, the incidence rate reached its highest in July 2020 (577 cases per 100K) before falling back to a relatively low level of 362 cases per 100K in September. It then spiked rapidly, peaking at 1,992 cases per 100K in December, before beginning to decline again in January 2021 and reaching a low of 121 per 100K in June. However, due to the emergence of the Delta variant, the incidence rate increased again in August, reaching 1,296 cases per 100K. The most contagious Omicron variant then emerged in November 2021, with infections surging to 6,111 per 100K in January this year, and falling rapidly to 317 per 100K in March, and rising slight since. The mortality rates showed a similar

Table 2 Incidence (cases per 100K), mortality (deaths per 100K) and CFR (%) of COVID-19 in the U.S

Date	Cumulative			Monthly		
	Cases per 100K	Deaths per 100K	CFR %	Cases per 100K	Deaths per 100K	CFR %
Jan-20	0	0	12.50	0	0	12.50
Feb-20	0	0	7.69	0	0	5.56
Mar-20	58	2	2.79	58	2	2.79
Apr-20	328	20	6.14	269	18	6.87
May-20	544	33	6.02	217	13	5.83
Jun-20	805	39	4.81	261	6	2.30
Jul-20	1,382	47	3.39	577	8	1.40
Aug-20	1,835	56	3.03	453	9	1.92
Sep-20	2,197	63	2.84	362	7	1.93
Oct-20	2,778	70	2.51	580	7	1.26
Nov-20	4,133	82	1.98	1,356	12	0.87
Dec-20	6,125	106	1.73	1,992	24	1.23
Jan-21	7,992	135	1.69	1,867	29	1.57
Feb-21	8,724	155	1.78	732	20	2.71
Mar-21	9,275	166	1.79	551	11	2.02
Apr-21	9,843	174	1.76	568	7	1.26
May-21	10,122	179	1.77	279	5	1.96
Jun-21	10,244	182	1.78	121	3	2.62
Jul-21	10,645	185	1.74	401	3	0.66
Aug-21	11,941	193	1.62	1,296	8	0.65
Sep-21	13,199	211	1.60	1,257	18	1.42
Oct-21	13,961	226	1.62	762	14	1.90
Nov-21	14,736	236	1.60	775	10	1.35
Dec-21	16,613	250	1.50	1,877	14	0.75
Jan-22	22,724	269	1.18	6,111	19	0.31
Feb-22	23,919	287	1.20	1,195	19	1.56
Mar-22	24,236	297	1.23	317	10	3.15
Apr-22	24,599	301	1.23	363	4	1.09
May-22	25,425	305	1.20	826	4	0.47

Note Cases (deaths) per 100K are rounded to the nearest whole Number and the percentage to two decimal places

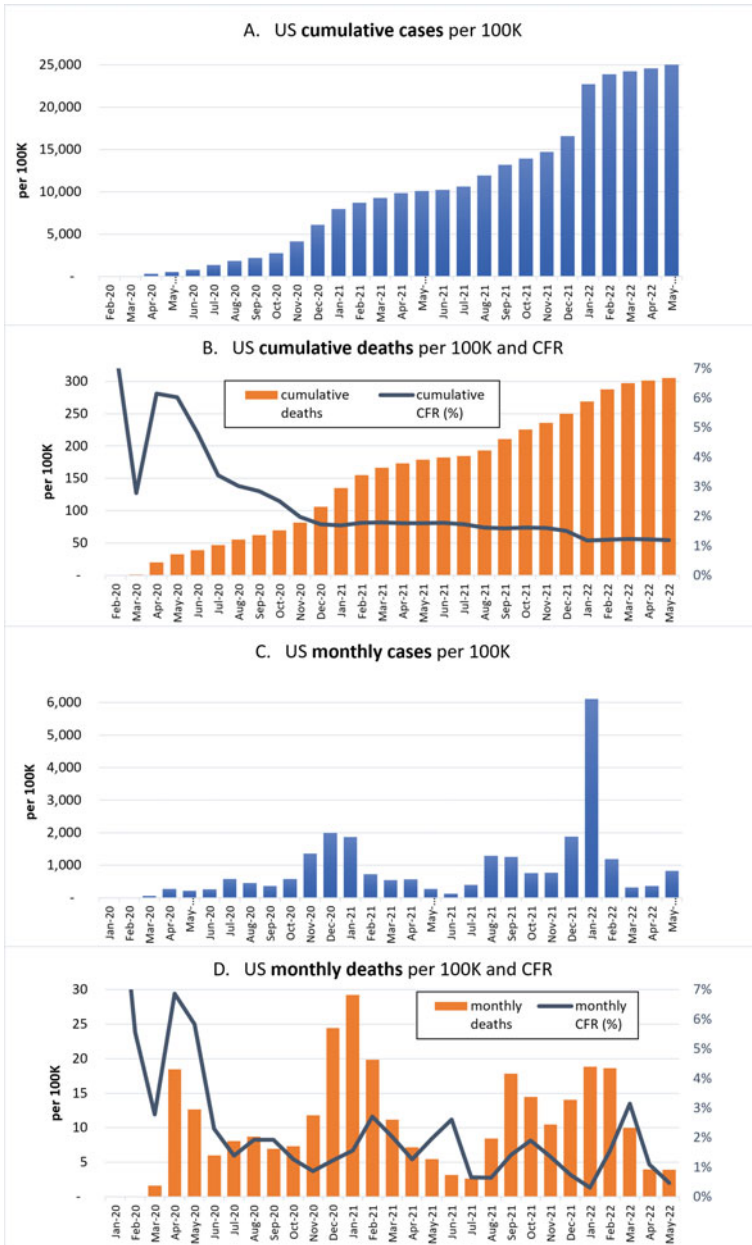


Fig. 1 Incidence, mortality and CFR of COVID-19 in the U.S.; **a** Cumulative cases per 100K, **b** Cumulative deaths per 100K and CFR, **c** Monthly cases per 100K, and **d** Monthly deaths per 100K and CFR

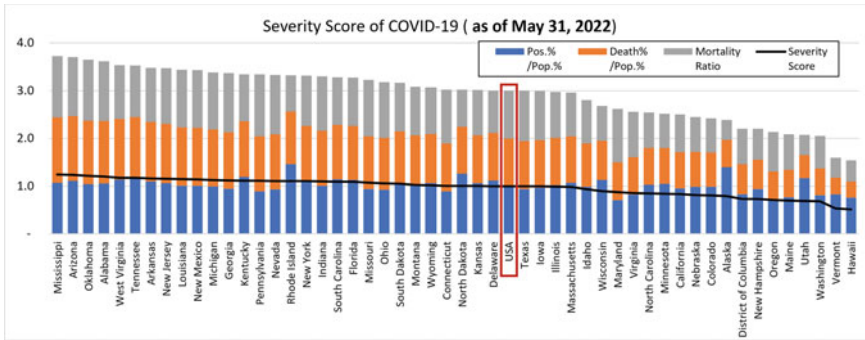


Fig. 2 As of May 31, 2022, the incidence, mortality, CFR and severity score of each state in the U.S. compared with the national averages, where the U.S. incidence rate is 25,425 per 100K, death rate is 305 per 100K, CFR is 1.20%, and all 4 ratio-based metrics, including severity score, are 1

trend, but were generally delayed by a month. Figure 1 includes 4 subplots, where A and B show cumulative results of incidence (in A) and mortality and CFR (in B); C and D show the monthly results of incidence (in C) and mortality and CFR (in D). Lines in subplots B and D represent the cumulative and monthly changes in case fatality rate (CFR) over the past 2 years (see right axis).

Using the national incidence, mortality, and CFR as benchmarks, each state’s data can be compared with these benchmarks, and then their severity can be assessed. Figure 2 displays the ratio-based metrics we proposed for all U.S. states, including their severity scores. This figure shows the cumulative results as of May 31, 2022. Please note that although we only provide a snapshot on May 31, one can take snapshots at any point in time according to their needs. In the chart, each vertically stacked bar represents metrics for each state, including Positive%/Population%, Death%/Population, and Mortality Ratio (with different color), while the black line represents the severity score. The graph is displayed in descending order of the state’s severity score from left to right. The U.S. national average is designated as the benchmark in the middle of the graph, highlighted by a red box. There are 29 states to the left of the benchmark having a severity score higher than the national average, while 22 states (including D.C.) to the right have a severity score lower than the national average. The results demonstrate that as of May 31, 2022, the five states severely affected by COVID-19 were Mississippi, Arizona, Oklahoma, Alabama, and West Virginia, and the five states least affected were Hawaii, Vermont, Washington, Utah and Maine. In addition, 29 states had higher incidence rates than the national average (25,425 cases per 100K), 27 states had higher death rates than the national average (305 deaths per 100K), and 25 states had a higher CFR than the national average (1.20%).

In this study, we chose to observe and compare changes in the severity of COVID-19 for all U.S. states from 2020 to 2022 (as of May 31, 2022). Figure 3 shows the COVID-19 severity in three time intervals: 2020, 2021 and 2022. It can be observed that in 2020, 21 states including D.C., had severity scores higher than the national

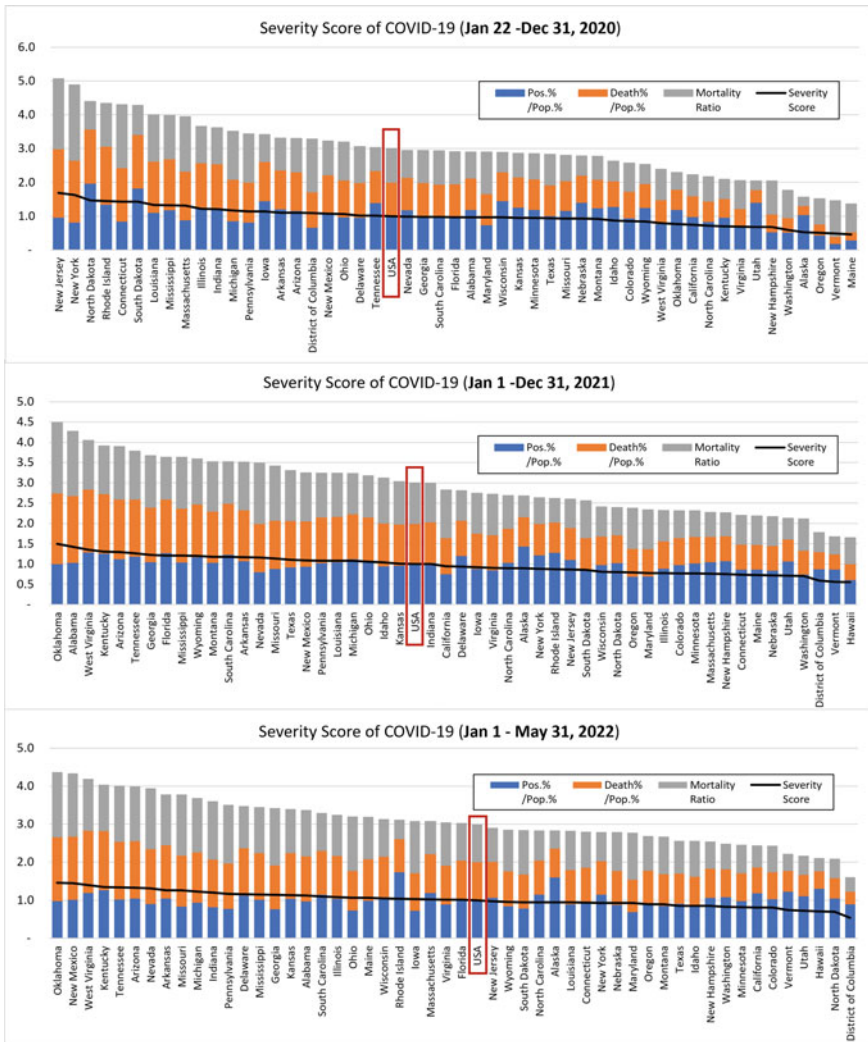


Fig. 3 Incidence, mortality, CFR and severity score of all states compared with US national average in the three time intervals: 2020 (upper), 2021 (middle), and 2022 (lower); the incidence rate of the U.S. is 6,075 per 100K, death rate is 105 per 100K, and CFR is 1.73% in 2020, 10,538, 145 and 1.38% in 2021, and 8,812, 55 and 0.63% in 2022 as of May 31

average, and the remaining 30 states had severity scores lower than the national average. In 2021, 23 states had severity scores higher than the national average, and 28 states including D.C. lower than the average. In 2022, 27 states had severity scores above the national average and 24 states, including D.C., below the average. In addition, it is observed that the severity score range narrowed from [0.40, 1.69] as of December 31, 2020 to [0.51, 1.24] as of May 31, 2022, implying that the disease has experienced extensive and intensive spread in all states over the past two-plus years.

Furthermore, Fig. 4 shows the geographic distribution in terms of the severity score (SS) of each state. States with SS greater than 1.05 (1.05 times the national average) are marked as orange; states with SS less than 0.95 are marked as blue; states with SS between 0.95 and 1.05 are marked as gray. Please note that the national average (benchmark) for each metric is constantly changing. In 2020, the positive rate, death rate and CFR were 6,125 per 100K, 106 per 100K and 1.73%, respectively, in 2021 they were 10,488 per 100K, 144 per 100K and 1.37%, while in 2022 as of May 31, the three values were 8,812, 55, and 0.63%. The following facts can be observed from the figure.

- Nine of the 19 orange states in 2020 that turned blue or gray in 2021 and 2022 showed better (less) COVID-19 severity scores. The 8 states are NJ, NY, ND, RI, CT, SD, MA, IA, plus the District of Columbia.
- Nine of the 32 blue or gray states turned orange, meaning their COVID-19 severity scores got worse in 2021 and 2022. The 9 states are TN, NV, GA, SC, AL, MO, WV, OK, and KY.
- Seven of the 19 orange states in 2020 remained the same color in 2021 and 2022, implying their COVID-19 severity scores consistently remained above the U.S. national average. The seven states are MS, MI, PA, AR, AZ, NM, and OH.
- Sixteen of the 32 blue or gray states remained the same colors in 2021 and 2022, demonstrating that their COVID-19 severity scores kept staying below the U.S. national average. These 16 states are MD, WI, MN, NE, ID, CO, CA, NC, VI, UT, NH, WA, AK, OR, VT, and HI.
- The remaining ten states have changed colors back and forth, meaning their COVID-19 severity scores have fluctuated over the past two-plus years. These ten states are LA, IL, IN, DE, FL, KS, TX, MT, WY, and ME.

Moreover, we used four proposed ratio-based metrics to describe the severity of COVID-19 in all U.S. states over time. Here, we use New York State as an example. Table 3 provides monthly results for the four metrics in NY. Values greater than 1 are highlighted in bold to indicate that they are above the U.S. average. Figure 5 shows the four ratio-based metrics of NY starting from May 2020. The baseline is the dashed line with a value of 1, representing the U.S. national average. It is worth mentioning that the metrics of the U.S. national average are dynamic, as shown in subgraphs C and D of Fig. 1 and the monthly columns in Table 2. The four colored curves represent four ratio-based metrics. When the line is above the horizontal dashed line, it means that the state's COVID-19 is more severe than the national average; when it is below that horizontal line, it means that the state's COVID-19 severity is lower than the

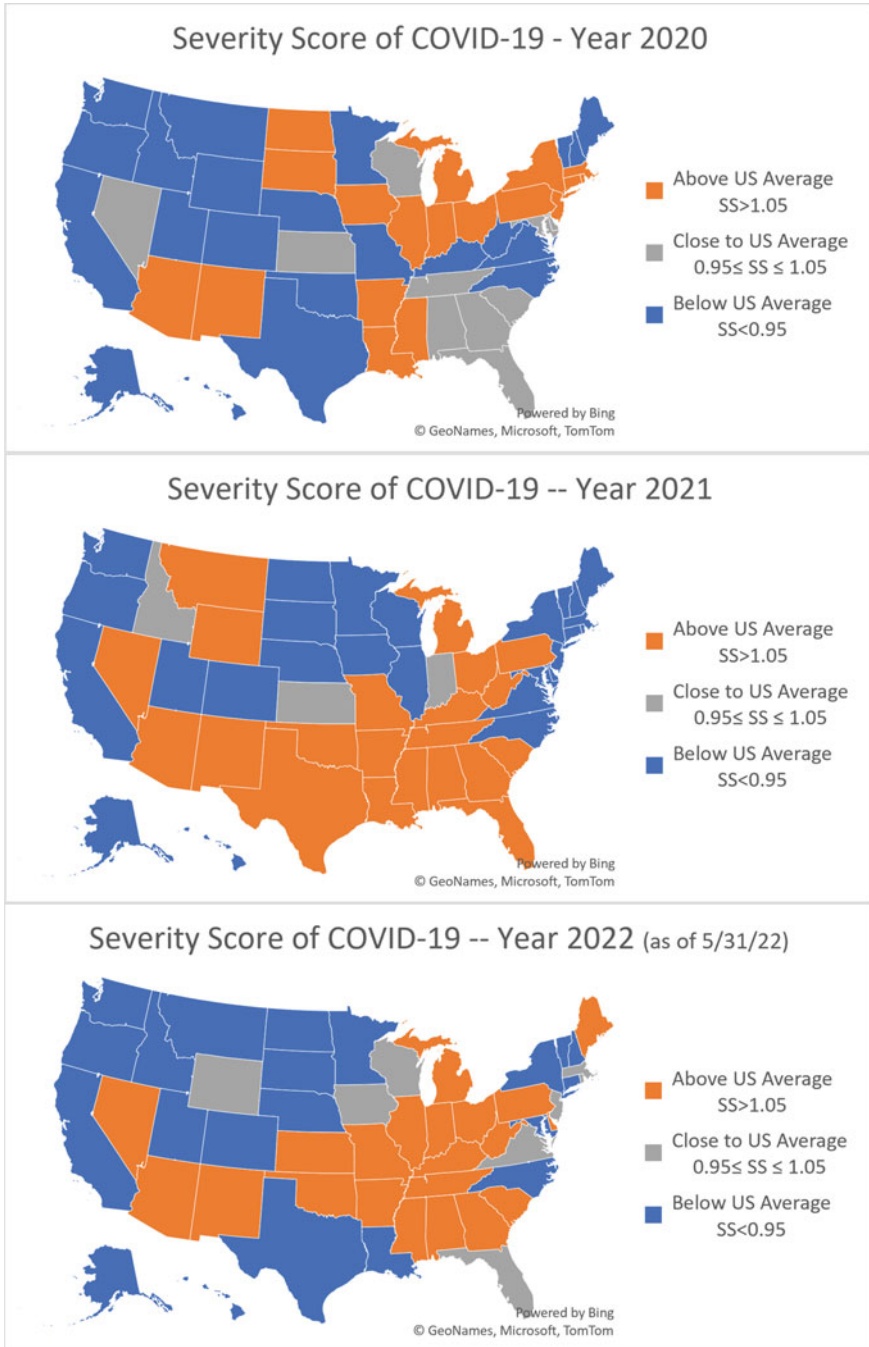


Fig. 4 The geographic distribution in terms of the severity score (SS) in the U.S. states for 2020 (upper), 2021 (middle) and 2022 (lower)

Table 3 Four ratio-based metrics described severity of COVID-19 in NY State

Date Mon-yy	Pos.% /Pop.%	Death% /Pop.%	Mortality ratio	Severity score
Mar-20	6.68	8.74	1.31	5.58
Apr-20	4.44	6.02	1.36	3.94
May-20	1.55	2.49	1.61	1.88
Jun-20	0.44	1.15	2.63	1.41
Jul-20	0.20	0.38	1.93	0.84
Aug-20	0.22	0.17	0.77	0.39
Sep-20	0.35	0.27	0.78	0.47
Oct-20	0.44	0.28	0.63	0.45
Nov-20	0.55	0.39	0.71	0.55
Dec-20	0.83	0.70	0.84	0.79
Jan-21	1.21	0.99	0.82	1.01
Feb-21	1.57	1.03	0.65	1.08
Mar-21	2.14	1.24	0.58	1.32
Apr-21	1.58	1.40	0.89	1.29
May-21	1.00	0.98	0.98	0.99
Jun-21	0.53	0.60	1.13	0.75
Jul-21	0.45	0.33	0.73	0.51
Aug-21	0.51	0.37	0.74	0.54
Sep-21	0.61	0.32	0.53	0.48
Oct-21	0.90	0.38	0.43	0.57
Nov-21	1.13	0.54	0.48	0.71
Dec-21	2.06	0.72	0.35	1.04
Jan-22	1.11	1.46	1.31	1.29
Feb-22	0.52	0.69	1.35	0.85
Mar-22	1.14	0.31	0.27	0.58
Apr-22	2.44	0.49	0.20	1.04
May-22	1.68	0.85	0.51	1.01

national average. As you can see from the chart, COVID-19 in NY was very severe in the spring of 2020. The situation has improved significantly since then, with the severity scores below the national average most of the time.

5 Discussion

The main advantage of our proposed metrics over the conventional metrics is that the new metrics can directly provide practitioners with a good judgment on the severity of the epidemic, because they are in a ratio-based format. Once a benchmark is given

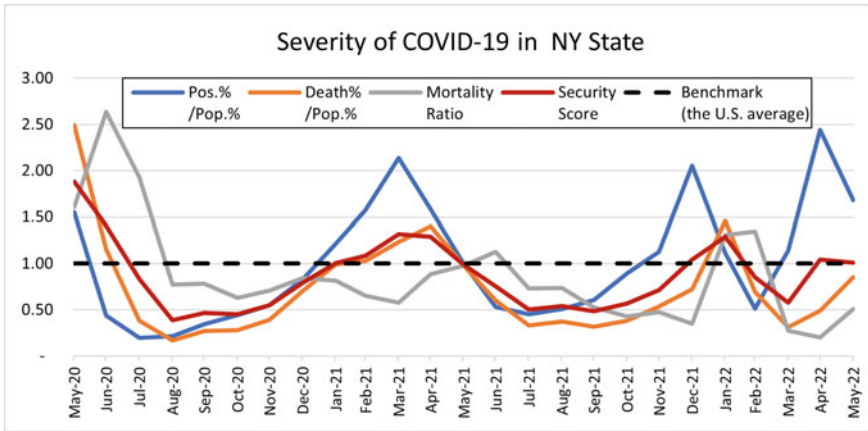


Fig. 5 Incidence, mortality, CFR and severity score of COVID-19 in NY State compared to the US national average

Table 4 Four ratio-based metrics describing severity of COVID-19 in NY State

Date Mon-yy	NY cases	NY deaths	NY CFR	US cases	US deaths	US CFR	Ratio-based metrics			
							Pos.% /Pop.%	Death% /Pop.%	Mortality ratio	Severity score
Dec-21	3,859	10	0.26	1,877	14	0.75	2.06	0.72	0.35	1.04
Jan-22	6,812	27	0.40	6,111	19	0.31	1.11	1.46	1.31	1.29
Feb-22	616	13	2.10	1,195	19	1.56	0.52	0.69	1.35	0.85
Mar-22	362	3	0.86	317	10	3.15	1.14	0.31	0.27	0.58
Apr-22	886	2	0.22	363	4	1.09	2.44	0.49	0.20	1.04

or selected (as in this study, the national average is used as the benchmark), these ratio-based metrics can help quickly and accurately determine the severity of a given area relative to that benchmark. For example, a simple classification method can be implemented by comparing the ratio value with 1 (which represents the benchmark). When the ratio is greater than 1, it is classified as severe; when the ratio is equal to (or close to) 1, it is classified as comparable; when the ratio is less than 1, it is classified as mild.

Another issue we would like to point out is that when calculating the CFR metric, we used the number of recorded deaths divided by the number of confirmed cases in the same time period, and did not take into account changes in mortality (deaths per 100K) which usually lags changes in incidence (cases per 100K). The CFR provided by JHU CSSE [3] uses the same approach. This may cause the calculated CFR to not perfectly reflect the true CFR. For example, during the Omicron outbreak earlier this year (see the monthly data in Table 2), in Dec. 2021, Jan. 2022 and Feb. 2022, the cases were 1,887, 6,111, 1,195 per 100K, respectively and dropped to 317 in March. However, the change in mortality came about a month later than the change

in incidence. The mortality in March is 10 deaths per 100K, resulting in a 3.15% CFR, which is not the case. Due to the limitations of the tracking data obtained, it is difficult to get accurate CFRs, especially for a time point or a short time interval. However, this does not affect our judgment of the disease evolution, trends, and severity, particularly when using our proposed ratio-based measures. Table 4 reports related metrics for the NY State and the U.S. average during the omicron variant outbreak, i.e., Dec, 2021 to Apr. 2022. Some data also appears in the previous tables and figures. As can be seen, the Omicron variant outbreak in NY State preceded the nationwide outbreak in the U.S. by several weeks. The severity scores indicate that NY State was more severe than the national average in December and January, and then less than the average for the following two months, especially in March, where cases and deaths were both at very low levels. After that, there was a slight upward trend.

One thing we can be sure of about SARS-CoV-2, the virus that causes COVID-19, is that it is constantly evolving. Since the start of the pandemic, we have seen many prominent variants, including Alpha, Beta, Delta, and Omicron. The surges of COVID-19 are closely linked to these SARS-CoV-2 variants. The link between the severity of COVID-19 and these SARS-CoV-2 variants will be investigated in the future.

6 Conclusion

In this paper, we discussed what would be a good indicator to describe the severity of COVID-19 in an area. We recommend using the incidence/death rate that takes into account the population (such as, the number of positives (deaths) per 100K or the ratio of number of positive cases (or deaths) to the population) rather than the absolute number of positives or deaths in an area. We proposed four new ratio-based metrics, one of which is called severity score, which integrates the other three metrics to describe the severity of COVID-19. These scores reflect the different levels of some key metrics. As a result, public health authorities can take necessary actions to effectively surveil and control the spread of the disease. Furthermore, we would like to emphasize the two key factors that affect the severity of COVID-19, namely the incidence rate and the case fatality rate. Health professionals and experts hope to reduce both in order to decrease the overall mortality of the disease. Effective testing and surveillance for suspected cases, protection of susceptible people, reduction of socializing activity in outbreak areas, and increased use of vaccines and therapeutic agents are good measures that are now being addressed or implemented on a global scale.

Future work will involve creating the severity score with more metrics, such as hospitalization rate, recovered counts, etc. In addition, severity score by sex, age, and race/ethnicity is also an interesting topic that requires further research. Finally, a weighted average to obtain the severity score may provide additional insight into the severity spread of the disease across the different states in the U.S.

References

1. Zhu N et al (2020) A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382(8):727–733. <https://doi.org/10.1007/s10796-013-9430-0>
2. CDC COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/#cases_casesinlast7days
3. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (2020). <https://github.com/CSSEGISandData/COVID-19>
4. Principles of Epidemiology in Public Health Practice, 3rd edn. U.S. Department Of Health And Human Services and Centers for Disease Control and Prevention (CDC) (2011). <https://www.cdc.gov/csels/dsepd/ss1978/index.html>
5. Fowlkes A et al (2021) Effectiveness of COVID-19 vaccines in preventing SARS-CoV-2 infection among frontline workers before and during B.1.617.2 (Delta) variant predominance—Eight U.S. Locations, December 2020–August 2021. *Morb Mortal Weekly Rep (MMWR)* 70(34):1167–1169. <https://doi.org/10.15585/mmwr.mm7034e4>
6. Andrews N et al (2022) COVID-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *N Engl J Med* 70(34):1532–1546. <https://doi.org/10.15585/mmwr.mm7034e4>
7. Pham H (2020) Estimating the COVID-19 death toll by considering the time-dependent effects of various pandemic restrictions. *Mathematics* 8(9):1628. <https://doi.org/10.3390/math8091628>
8. Friedman J et al (2021) Predictive performance of international COVID-19 mortality forecasting models. *Nat Commun* 12:2609. <https://doi.org/10.1038/s41467-021-22457-w>
9. Sander A-L et al (2021) Mutations Associated with SARS-CoV-2 variants of concern, Benin, early 2021. *Emerg Infect Dis (EID)* 27(11). <https://doi.org/10.3201/eid2711.211353>
10. SARS-CoV-2 Variants of Concern as of 12 May 2022 (2022). <https://www.ecdc.europa.eu/en/covid-19/variants-concern>
11. Delta Variant: what We Know About the Science (2022). <https://www.cdc.gov/coronavirus/2019-ncov/variants/delta-variant.html>
12. Grobler J et al (2020) Accelerated preclinical paths to support rapid development of COVID-19 therapeutics. *Cell Host and Microbe* 28(5):636–645. <https://doi.org/10.1016/j.chom.2020.09.017>
13. Reed C et al (2013) Novel framework for assessing epidemiologic effects of influenza epidemics and pandemics. *Emerg Infect Dis* 19(1):85–91. <https://doi.org/10.3201/eid1901.120124>
14. Pandemic Severity Assessment Framework (PSAF). <https://www.cdc.gov/flu/pandemic-resources/national-strategy/severity-assessment-framework.html>
15. Pandemic Influenza Severity Assessment (PISA) A WHO guide to assess the severity of influenza in seasonal epidemics & pandemics (2017). <https://apps.who.int/iris/bitstream/handle/10665/259392/WHO-WHE-IHM-GIP-2017.2-eng.pdf>
16. 2019 National and State Population Estimates (2019). <https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html>
17. How to Report COVID-19 Laboratory Data. <https://www.cdc.gov/coronavirus/2019-ncov/lab/reporting-lab-data.html>
18. How to Report COVID-19 Laboratory Data, APHL Informatics Messaging Services (AIMS). <https://www.cdc.gov/coronavirus/2019-ncov/lab/reporting-lab-data.html>
19. WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int/>
20. Rothman KJ, Lash TL, Greenland S (2013) *Modern epidemiology*, 3rd edn. Lippincott Williams & Wilkins, Philadelphia, PA

Kehan Gao Eastern Connecticut State University, Connecticut, USA

Dr. Kehan Gao is a Professor in the Department of Computer Science at Eastern Connecticut State University. Her research interests include data mining and machine learning, software engineering, software reliability & quality engineering, and computational intelligence. She has published more than 80 refereed journal and conference papers in these fields. She has served on

organizing and technical program committees of various national and international conferences. See her Google Scholar link: https://scholar.google.com/citations?user=_SJIoHwAAAAJ.

Sarah Tasneem Eastern Connecticut State University, Connecticut, USA

Dr. Sarah Tasneem is a Professor in the Department of Computer Science at Eastern Connecticut State University. Her research interests include performance modeling of operating systems scheduling, queuing theory, modeling and simulations, and undergraduate computer education.

Taghi Khoshgoftaar Florida Atlantic University, Florida, USA

Dr. Taghi M. Khoshgoftaar is Motorola Endowed Chair Professor of the Department of Electrical Engineering and Computer Science, Florida Atlantic University, and the Director of NSF Big Data Training and Research Laboratory. His research interests are in big data analytics, data mining and machine learning, health informatics and bioinformatics, social network mining, and software engineering. He has published more than 850 refereed journal and conference papers in these areas. He was the conference chair of the IEEE International Conference on Machine Learning and Applications (ICMLA 2019). He is the Co-Editor-in-Chief of the journal of Big Data. He has served on organizing and technical program committees of various international conferences, symposia, and workshops. Also, he has served as the North American Editor of the Software Quality Journal and was on the editorial boards of the journals Multimedia Tools and Applications, Knowledge and Information Systems, and Empirical Software Engineering and is on the editorial boards of the journals Software Quality, Software Engineering and Knowledge Engineering, and Social Network Analysis and Mining.

Promoting Expert Knowledge for Comprehensive Human Risk Management in Industrial Environments



Ilyas Mzougui, Silvia Carpitella, and Joaquín Izquierdo

Abstract This research develops a methodological approach aimed at promoting expert knowledge and, as a consequence, exploiting the strength of practical experience in industry in a scientifically rigorous way. In particular, we make use of Fuzzy Cognitive Maps (FCMs) to organise in a flexible way human knowledge about decision-making problems focused on the topic of human risks management. FCMs enable the representation of causal relations among relevant decision-making elements on the basis of spontaneous human brainstorming stimulated from a single expert or a decision-making team. After leading a comprehensive literature review, aimed at identifying industrial human risks in the most exhaustive way, a FCM will be built to define relations among risks. Specifically, the calculation of the total effect will be supported by means of the use of the Nonlinear Hebbian Learning (NHL) algorithm, something that will be useful to make the approach even more reliable. Risks will be later prioritised by means of a modified Failure Modes, Effects and Criticality Analysis (FMECA): the total effect previously calculated will be considered as an additional criterion for risk prioritisation, apart from the ones traditionally considered by FMECA. Moreover, an alternative calculation way for the Risk Priority Number (RPN) will be proposed: the Multi-Criteria Decision-Making (MCDM) approach making use of the TODIM (an acronym in Portuguese standing for interactive and multi-criteria decision-making) method under Z-environment will be used to perform the final ranking. This technique is indeed useful to deal with the psycho-

I. Mzougui

Faculty of Sciences and Technologies, Abdelmalek Essaadi University, Tangier, Morocco

S. Carpitella (✉)

Department of Manufacturing Systems Engineering and Management, California State University, Northridge, USA

e-mail: silvia.carpitella@csun.edu

J. Izquierdo

Institute for Multidisciplinary Mathematics, Universitat Politècnica de València, Valencia, Spain

e-mail: jizquier@upv.es

logical behaviours of decision-makers. A real case study is eventually implemented to provide useful implications for risk management.

Keywords FCM · Human risk · Nonlinear Hebbian learning algorithm · FMECA · Z-TODIM

1 Introduction

The continuous improvement of industrial risk management processes is widely considered as a key factor for optimising organisational efficacy, especially under uncertain conditions and challenging competitiveness at a global level [15]. In this context, developing proper decision-making models aimed at implementing effective solutions can be strategic. This process should meet rigorous methodological requirements as well as keep adherence to the practical nature of the specific industrial context of reference. To such an aim, experts familiar with practical risk management problems are involved and often asked to provide judgments of preference by comparing pairs of decision-making elements. Decision models implemented on the basis of the preliminary collection of these types of data are grounded on preference transitivity characterized by ordinal consistency as a fundamental principle [89]. Transitivity represents a cornerstone of normative decision theory [24], being largely considered as an essential requirement for deriving reliable and coherent solutions. This principle has also been criticised by several authors, who underline as it may force experts to align their opinions to requirements of the consistency property [1]. This may lead to artificial decisions aimed at responding to mathematical principles rather than reflecting the often non-transitive nature of human reasoning.

The present research aims to develop a previous conference contribution [18] that offered an approach independent on preference transitivity to model and solve such complex problems as making effective decisions on industrial risk management. By making use of Fuzzy Cognitive Maps (FCMs) integrated within the framework of Failure Modes, Effects and Criticality Analysis (FMECA), we demonstrated how to stimulate a spontaneous brainstorming for evaluating decision-making elements relevant for risk management purposes. Such evaluations can be led in the most natural way, by removing the need of eventually checking mathematical consistency within the body of input judgments formulated by the interviewed decision-makers. On the basis of the obtained results, the current work aims to develop a novel structured approach for risk assessment. Specifically, FCMs previously used to model interactions among human risks potentially involving a system will allow us to consider again the causal relationships for risk prioritization but, this time, calculations will be further strengthened through the Nonlinear Hebbian Learning (NHL) algorithm. Furthermore, the TODIM (an acronym in Portuguese standing for interactive and multi-criteria decision-making) method extended under Z-number theory (Z-TODIM) will be used to perform the final risk prioritisation and increase the reliability of the performed analyses.

The rest of this study is organized as follows. The literature review is developed in Sect. 2, by discussing the use of FCMs supporting decision-making processes, traditional methods for risk assessment and related Multi-Criteria Decision-Making (MCDM) applications. The stage of risk identification is formalised as a set of input elements derived from the study of the existing literature. Section 3 describes the proposed theoretical approach and Sect. 4 presents a real case study in the sector of the automotive industry, applied to show the effectiveness of the proposed approach. Section 5 lastly provides the conclusions and potential future lines of research.

2 Literature Review

2.1 FCMs Easing Decision-Making

Decision support systems frequently rely on judgments of pairwise comparisons attributed by selected stakeholders that are experts in the business field of interest. Such an approach usually represents a useful way to increase the probability that the obtained solutions are actually effective for pursuing process optimisation and not randomly made. This is what often happens in practice, where subjects in charge of making decisions, despite occupying management positions, are not properly familiar with practical problems at a technical level. In any case, even when it comes to expert subjects, their capability of being consistent when pairwise comparing elements may flaw, especially in complex situations [10, 11, 26] and when vagueness plagues phenomena [19], something that normally occurs in the business world.

With this regard, as highlighted in [1], transitivity of preference judgments is a useful mathematical assumption in decision making given to the link between consistency of pairwise comparisons and reliability of the obtained priorities, the last ones being important for establishing the mutual importance of the main elements taken into account within the analysed problem. However, transitive reasoning may be the result of memory phenomena rather than deduction processes. The theory of transitive reasoning is discussed in [88], by stating that it explains transitivity as information-processing constructs such as encoding, mentally-scanning and retrieval cuing. In this context, the author underlines as an important part of human transitive reasoning process responds to genuinely deductive paradigms, which do not merely tend to access an associative mode capturing transitive generalisations as soon as information is acquired.

On the one hand, the ability to understand and manipulate logical transitive relations by making inferences is certainly fundamental for acquiring and developing many mathematical concepts [78]. On the other hand, preferences may be intransitive for a huge amount of practical problems [83] and, in such cases, distorting original assessments provided by experts may not be a desirable solution. This is even truer when one has to deal with decision making problems in which many diverse elements to be pairwise compared are considered [9]. This may be likely the case of risk man-

agement problems in industry, where several risks can be highlighted depending on the specific industrial sector under analysis. Focusing on this kind of practical problems, the present research further develops the risk assessment approach presented in [18], making use of FCMs to stimulate experts towards a spontaneous brainstorming for effective risk analysis. Relationships among the identified risks would be captured and represented without requiring the formal check of consistency for preferences expressed between pairs of elements.

Cognitive maps (CMs) are directed graphs explicitly representing a set of variables and their causal relationships by means of causal weights, also developed in fuzzy version [41]. The CM tool has been formalised and integrated within the mathematics discipline to flexibly describe a general behaviour representation by encoding relationships among elements [8]. The CM topic has been extensively discussed in the literature as an approach capable to provide structural frameworks for managing complex information [68]. Having underlined the possibility of applying CMs to a wide spectrum of scientific fields, Mourhir [54] develops a review on FCM applications in the sector of environmental assessment promoting integration models assessment. Chen et al. [21] propose a hybrid soft computing approach where a FCM is built to evaluate risks associated to projects on public-private partnership. Azar and Mostafae Dolatabad [4] integrate FCM and Bayesian Networks to evaluate operational risk and implement risk management strategies in financial institutions. Being the topic of CMs and FCMs particularly lively in the current literature, we propose a FCM-based application for industrial risk management and, particularly, for human risk modeling and potential hazard propagation identification on the basis of relations highlighted among risks. This approach is integrated within a modified FMECA framework for eventually proceeding with risk prioritisation. With respect to the previous work of research [18], a wider list of human risks is formalised and the approach is further enriched by the use of the NHL algorithm and the Z-TODIM, useful to analyse a number of significant aspects.

2.2 Formal Human Risks Identification

The field of human factor risk is the scientific discipline concerning the understanding of interaction amongst the core elements of a system. It is a complex domain [13] where diverse aspects are bonded with each other in terms of multiple characteristics and behaviors that could describe the functionality of a given system under specific environmental conditions. These relationships may interfere and potentially impact on systems performance [16], increasing complexity within the risk management process. Roles and effects of human activity within the system introduce additional strengths, weaknesses and uncertainties [87]. Managing human factors on the basis of risks is a difficult task for companies because of the complexity of the problem that combines several criteria and alternatives. [13]. Aiming at supporting this process, a list synthesising significant human risks in industry has been first drawn up from the literature and provided in Table 1. As already mentioned, this list extends the

Table 1 Main human risks identified in the existing literature

Category	ID code	Description
Environment	E_1	Noise beyond the normal level [12]
	E_2	Exposition to intense vibration [12]
	E_3	Improper lighting and microclimate [75]
Work space	W_1	Unsuitable workplaces organisation [12]
	W_2	Dangerous distribution of work spaces [12]
	W_3	Overlapping individual work spaces [12]
	W_4	Incorrect ergonomics conditions [13]
	W_5	Personal protective equipment (PPE) damaged or improperly used [13]
Tools, machinery	T_1	Inappropriate use and/or management of machines and plants [90]
	T_2	High variation of tools to be used [12]
	T_3	Incorrect use of work equipment [13]
	T_4	Improper work scheduling and deadlines which cause decrease of production activities [57, 75]
	T_5	Excessive trust in global network contents[40]
	T_6	Lack of habit of systematic updating and self-control mechanisms [40]
	T_7	High obsolete inventory costs with no strategic direction [25]
Mental load	M_1	Negative psychological state [90]
	M_2	Existence of several distraction factors [12]
	M_3	Bad life habits influencing work [13]
	M_4	Fatigue, poor response to abnormal situation and lack of awareness [57, 75]
	M_5	Memory deterioration [57, 75]
	M_6	Toxic relationship with internal/external stakeholders [75]
	M_7	Lack of vision, discontinuous evaluation and obsolete policies/human resources practices [25]

set of human risks considered in our previous contribution [18] with the objective of increasing the effectiveness of the risk identification stage.

2.3 FMECA Supporting Risk Management

FMECA was first developed in the US Defense Department in 1949 and later applied by NASA to the Apollo project in the 1960s [14] to enhance system reliability. The technique was then adopted and upgraded by Ford Motors in 1977 and it is nowadays

considered as a valuable and effective reliability analysis method, capable to deal with a variety of industries in a flexible way [27, 62]. FMECA-based analyses are indeed among the most widely methods used for risk assessment in the existing literature. Specifically, FMECA is a structured and proactive reliability management technique [79] based on team members' cooperation towards safety and quality of products and services. This analysis can be particularly beneficial for quantitative risk evaluation purposes by contributing to the general improvement of the performance level for systems and processes [58]. FMECA is used to identify failures, causes, and potential effects with the general goal of evaluating and prioritising all the potential failure modes by supporting the identification of proper preventive/corrective actions [76].

Despite its various advantages, the literature also pinpoints several shortcomings for FMECA, something that may potentially limit the effectiveness of the method itself along with the whole level of reliability of analyses. First of all, the risk prioritisation method performed by FMECA has been highly criticised as not having a robust scientific basis [32]. It is based on the calculation of the Risk Priority Number (RPN), an index that can be referred to disregard uncertainties in the values of the related risk factors [5]. Moreover, the risk assessment carried out by means of FMECA is not capable to capture the risk effect deriving from the causal relationships bonding failures with each other [74]. This evidence reduces the global effectiveness of analysis as it is not realistic to not take into account the existence of these kinds of relationships. As a consequence, without considering this important aspect, the final risk ranking would be biased [61]. To deal with this particular shortcoming, the integration between FCMs and FMECA is encouraged [69]. For instance, this integration has been used under a fuzzy environment to consider causal relationships for an IT case study [23]. Again, a hybrid multi-attribute decision-making model combining fuzzy preference programming, FCMs and fuzzy graph-theoretical matrix approach has been suggested to rank FMECA risks [7]. In any case, the discussed weaknesses pose clear challenges for increasing reliability of risk analyses.

To such an aim, the question of how to effectively manage risks has been encouraging researchers. Several approaches enhancing traditional FMECA have been developed in literature [33, 44, 46]. Because of the involvement of multiple risk factors within the evaluation process and failure prioritisation, the risk assessment process carried out by FMECA could be described as an MCDM problem [38, 45, 61]. According to that, plenty of studies using MCDM methods have been developed to properly evaluate the risk associated to the potential occurrence of the identified failure modes for complex systems [45].

2.4 MCDM Approaches in the Analysed Field

The existing literature proposes many examples of industrial risk management problems effectively supported by means of MCDM methods integrated with mathematical tools capable to deal with uncertainty and vagueness of input evaluations.

An improved approach for risk assessment is proposed by Kutlu and Ekmekçioğlu [42], who combine the Fuzzy Analytic Hierarchy Process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). The extended VIKOR (an acronym in Serbian standing for multi-criteria optimization and compromise solution) method under fuzzy environment has been used by Liu et al. [47] to analyse and rank failure modes of general anesthesia processes under uncertain conditions. An integrated MCDM approach making use of fuzzy Decision Making Trial and Evaluation Laboratory (DEMATEL), Analytic Network Process (ANP) and VIKOR has been implemented on FMECA analyses to improve risk assessment for city logistic concept selection [81]. FMECA has been combined with the Grey Relational Analysis (GRA) method under fuzzy environment to effectively identify root causes of tank-ship equipment failures [92]. Wang et al. [86] have proposed an improved FMECA for a feed system case integrated with DEMATEL for analysing relations of dependence. Mohsen and Fereshteh [53] have suggested an extension of the VIKOR method to be used on Z-numbers risk to analyze and identify critical causes of failures for geothermal power-plants. The use of Z-numbers has been introduced to improve the reliability of analysis' results. Li et al. [43] have proposed a FMECA-based improved approach for risk assessment by prioritising risks through the TOPSIS method under the rough set and cloud model theories. New risk factors have also been introduced to improve the stage of risk assessment [59].

Experts' subjectivity has been treated by using a combination of support vector machine and fuzzy inference system [51]. Huang et al. [34] have combined the normal projection and TOPSIS methods under Z-environment to assess risks related to an aircraft landing system. The power of Z-numbers has been applied to reduce the vagueness of experts' opinions [22]. An approach based on fuzzy DEMATEL has been developed to assess risks of automotive supply chains [58], as risks impacting supply chains represent a topic widely studied in the literature. With this regard, Liu and Li [48] have used a K-means clustering method to improve reliability of experts' opinions and proposed a combination between the Regret Theory (RT) and Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) II to rank risks.

Given the proven usefulness of MCDM methods as a structured methodological tool to treat the research field of reference, we propose integrating FCMs within FMECA according to a MCDM-based perspective, the last one exploiting the advantages of Z-numbers. The proposed theoretical approach is described in the next section.

3 Integrated Methodological Approach

In this section, we will present the proposed approach to analyse human factor risks. The approach is detailed and explained step by step by means of the flowchart of Fig. 1, showing as we are going to combine different methods and concepts to increase effectiveness of the risk management practice. With relation to the human risks that

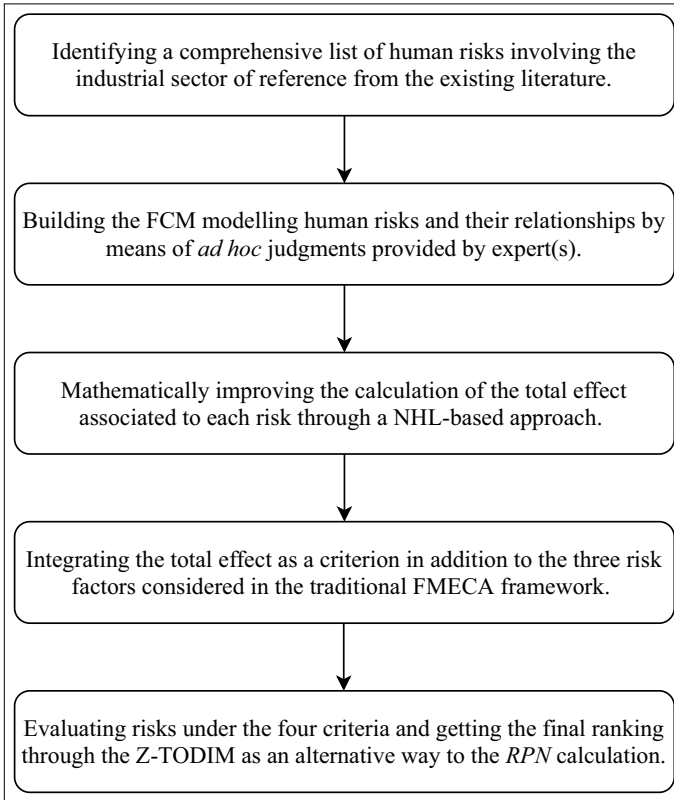


Fig. 1 Flowchart representing the proposed methodological approach for human risk assessment

have been identified on the basis of the study of the existing literature and formalised in Sect. 2.2, we make use of FCMs to identify and model failure interdependence as well as to calculate the total effect for each risk. These calculations will be supported by the NHL algorithm as it allows to obtain more effective and rigorous results. Moreover, we are going to use an expanded version of the TODIM method under Z-number theory to provide an alternative way to traditional FMECA-based risks prioritization, with the objective of further increasing the quality of the proposed approach. Among the plethora of techniques in the literature, we specify that the TODIM method has been chosen because of its peculiar capability of considering the psychological behaviors of decision makers, being directly derived from prospect theory [82]. Specifically, risks will be treated as alternatives of a decision-making problem. The prioritisation will be carried out through Z-TODIM by considering as criteria the three risk factors traditionally taken into account by FMECA plus the total effect calculated via FCM-NHL, as a criterion expressing dependence and mutual relations among the potential occurrence of human risks. In this way, we

aim to improve FMECA calculations by also integrating the aspect of dependence, something that traditional analyses do not take into account.

Methodological details about the used techniques are provided in the following subsections.

3.1 FCMs Modelling Human Risks

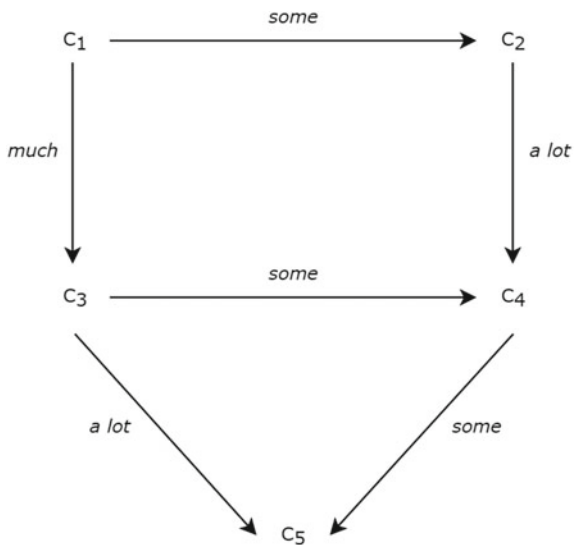
Initially developed by Axelrod [3] to study social scientific knowledge in decision-making activities focused on international politics, CMs were later extended to FCMs by [41]. FCMs allow to solve two specific problems that respectively consists in modeling complex systems and displaying relationships among elements [6]. Concepts and relations are linguistically represented and treated with the support of fuzzy sets.

Indirect effects (*IE*) and total effects (*TE*) from C_i to C_j can be formalised by means of such linguistic evaluations as e_{ij} as *much*, *some* and *a lot*. These evaluations will be translated to fuzzy numbers. Figure 2 exemplifies the FCM used by Kosko [41] and also referenced in a previous work [17] in order to explain the next equations.

We can observe as three possible casual paths connect C_1 to C_5 , namely, are $P_1(1 - 2 - 4 - 5)$, $P_2(1 - 3 - 5)$ and $P_3(1 - 3 - 4 - 5)$. The tree indirect effects between C_1 and C_5 associated to these paths (IE_1 , IE_2 and IE_3) can be assessed as follows:

$$IE_1(C_1, C_5) = \min\{e_{12}, e_{24}, e_{45}\} = \min\{\text{some}, \text{a lot}, \text{some}\} = \text{some}; \quad (1)$$

Fig. 2 Example of FCM developed by Kosko [41]



$$IE_2(C_1, C_5) = \min\{e_{13}, e_{15}\} = \min\{\text{much, a lot}\} = \text{much}; \quad (2)$$

$$IE_3(C_1, C_5) = \min\{e_{13}, e_{34}, e_{45}\} = \min\{\text{much, some, some}\} = \text{some}. \quad (3)$$

Once evaluated indirect effects IE , the total effect TE of element C_1 over element C_5 will correspond to the maximum evaluation among them, as follows:

$$\begin{aligned} TE(C_1, C_5) &= \max\{IE_1(C_1, C_5), IE_2(C_1, C_5), IE_3(C_1, C_5)\} \\ &= \max\{\text{some, much, some}\} = \text{much}. \end{aligned} \quad (4)$$

We can derive as, broadly speaking, element C_1 imparts *much* causality to element C_5 . To be quantitatively manipulated, linguistic evaluations will be translated into fuzzy numbers (for instance triangular or trapezoidal) according to scales defined *ad hoc* by the analyst.

As already highlighted, FCM-based applications are very useful to understand the nature of complex relationships bounding decision-making elements, by modeling them as an interconnected network. Literature offers various examples, e.g. in the fields of risk assessment [17], environmental management [55], hazardous industrial operations [50], maintenance management [52], Internet of Things strategy development [20], and so on.

3.2 Nonlinear Hebbian Algorithms on FCMs

Companies usually build their future and reputation around complex systems integrating human factors with new technologies to make decisions in the presence of uncertainty. Human knowledge is certainly the primary basis to develop design processes of different systems. However, most of the times, human reasoning is plagued with vagueness and experts' opinions are subjective and imprecise [77]. Inaccuracy of judgments is also caused by the presence of different variables and components [2]. To overcome these limitations, learning algorithms have been developed to increase weight accuracy, decrease dependence on experts' opinions and improve maps' performance [64, 72]. In this context, the NHL algorithm [31, 80] as described in literature as the best-known unsupervised learning approach to deal with connections problems [66]. Since FCMs are considered as having a non-linear structure, this algorithm is used in the literature to train FCMs by increasing precision of predictions and network classification [39, 63]. Considering these aspects, we use this unsupervised learning algorithm to improve experts' opinion and to build a more reliable FCM to model human risks. The proposed algorithm is based on NHL rule for Artificial Neural Network (ANN) learning [29, 60] under an adaptation to better fit the FCM case [63].

Examples of applications showing the effectiveness of an integrated FCM-NHL approach refer to a deeper analysis of hidden concepts in such relevant scientific

fields as, for instance, soft computing [67], energy management systems [56] and maritime safety [85], among others.

The value of each concept per each simulation step can be obtained by evaluating the influence of the interconnected concepts to the specific concept, as follows:

$$A_i^{(k)} = f\left(A_i^{(k-1)} + \sum_{j \neq 0, j=1}^n A_j^{(k-1)} \cdot w_{ji}\right); \quad (5)$$

where $A_i^{(k)}$ is the value of concept C_i at the iteration step (k), $A_j^{(k-1)}$ is the value of concept C_j at the iteration step ($k-1$), w_{ji} is the weight of the interconnection from concept C_j to concept C_i , and f is the sigmoid function. The last one belongs to the family of squeezing functions, for which the following function is generally used:

$$f(x) = \frac{1}{1 + e^{-\lambda x}}. \quad (6)$$

The following mathematical equation incorporates a learning rule along with a learning rate parameter [65]:

$$\Delta w_{ji} = \eta_k A_i^{(k-1)} (A_j^{(k-1)} - w_{ji}^{(k-1)} A_i^{(k-1)}); \quad (7)$$

the coefficient η being the so-called learning parameter, a small positive scalar factor that can be determined by means of the experimental trial and error method.

The following equation is adapted as a nonlinear weight learning rule to train FCMs:

$$w_{ji}^{(k)} = \gamma w_{ji}^{(k-1)} + \eta A_i^{(k-1)} \left(A_j^{(k-1)} - \text{sgn}(w_{ji}) w_{ji}^{(k-1)} A_i^{(k-1)} \right); \quad (8)$$

where η is the learning rate parameter, γ is the weight decay parameter and the sgn function is used to keep the sign of the corresponding weight as well as to further keep the physical meaning of relationships for the problem under analysis.

Two termination conditions are proposed for the algorithm to converge to an ideal solution [65]. In the first termination condition, function F_1 is introduced aiming at analysing the values of desired output concepts (DOCs) of interest:

$$F_1 = ||\text{DOC}_i - T_i||; \quad (9)$$

where DOC_i takes its value in the range $[T_i^{\min}, T_i^{\max}]$, and $T_i = \frac{T_i^{\min} + T_i^{\max}}{2}$, as the mean target value of the interesting concept DOC_i . Assuming to have m DOCs in a FCM model, in order to converge to a desired region, the equation for F_1 becomes:

$$F_1 = \sqrt{\sum_{j=1}^m (\text{DOC}_i - T_i)^2}; \quad (10)$$

The goal consists in achieving a minimum function F_1 for a set of weights by getting the sought equilibrium point of the FCM when the termination function is minimized.

When it comes to the second termination condition, an additional function F_2 has been introduced for NHL [65], aiming at finalising the iterative process of the learning algorithm after a certain number of steps. The function can be expressed as:

$$F_2 = |\text{DOC}_i^{(k+1)} - \text{DOC}_i^{(k)}| < e = 0.002; \quad (11)$$

DOC_i being the value of i -th concept.

3.3 Modified FMECA Integrating the Total Effect

The above presented FCM-based approach, integrated with the NHL algorithm, is now embedded within the framework of the FMECA technique. Specifically, we are going to include the total effect calculated by FCM-NHL as a fourth criterion for risk assessment, along with the three risk parameters considered by FMECA for the RPN calculation. This will be done in order to overcome one of the FMECA drawbacks: not to consider causal relationships bounding the elements of analysis.

As already expressed, FMECA is a systematic procedure aimed at analysing all the potential failure modes that may impact on a system by identifying the related causes and effects. The output coming from the application of this technique consists in calculating the RPN as a priority index associated to each failure mode. The traditional calculation is based on the following multiplication:

$$\text{RPN} = S \times O \times D; \quad (12)$$

where parameters S , O and D respectively indicate severity (intensity of impact of a given failure mode), occurrence (frequency of occurrence of a given failure mode) and detection (probability of correct failure detection). These parameters are generally ranged within discrete intervals. In the present research, we are going to use the scale proposed in the risk assessment matrix formalised in [13] to assess the severity parameter, whereas scales proposed in [35] will be used to evaluate the occurrence and detection parameters for each human risk. As already said, the total effect for each human risk will be taken into account for risk analysis together with these three factors, and RPN calculation will be updated for the sake of precision. This last objective will be achieved by prioritising risks through the MCDM approach based on Z-TODIM, as we are going to specify next.

3.4 The Z-TODIM Method for Risk Prioritisation

Z-TODIM is an effective procedure for solving prioritisation decision-making problems by considering the psychological behavior of the involved decision makers as well as the simultaneous interactions among criteria which, in our case, are risk indicators [84]. Furthermore, the TODIM method has been recommended in the literature as a valuable way of increasing the reliability of risk ranking on FMECA analyses [30]. TODIM is going to be herein extended under the umbrella of Z-number theory which is capable to treat uncertainty of human experts and allocates their reliability to estimate suitable fuzzy parameters [91].

Before introducing the method, we are going to succinctly give an overview on the Z-number theory. The concept of Z-numbers has been introduced by Zadeh [91] as an ordered pair of fuzzy numbers (\tilde{A}, \tilde{R}) , where the first component \tilde{A} represents a constraint on the values that a variable can assume and the second component \tilde{R} provides a measure of the degree of certainty about the value assumed by \tilde{A} . The degree of confidence may be expressed as a probability density function or fuzzy set for which $\tilde{Z} = x \in \tilde{A}$ with a certain degree equal to \tilde{R} . While declaring problems described by Z-numbers can be easy, their actual solution is quite complicated at a computational level. In this study, Z-numbers have been converted to fuzzy numbers as shown in [37]. Let us consider an opinion expressed by an expert as a fuzzy number $\tilde{A} = (x, \mu_{\tilde{A}})$ with reliability $\tilde{R} = (x, \mu_{\tilde{R}})$. Expert's knowledge can be first expressed as a Z-number as $\tilde{Z} = (\tilde{A}, \tilde{R})$. Secondly, expert's reliability can be converted into a crisp value by:

$$\alpha = \frac{\int x \mu_{\tilde{R}}(x) dx}{\int \mu_{\tilde{R}}(x) dx}. \tag{13}$$

For trapezoidal fuzzy numbers, the centroid α can be alternatively calculated by:

$$\alpha = \frac{a_1 + a_2 + a_3 + a_4}{4}. \tag{14}$$

Crisp values of reliability are applied to the constraint. This means that the weight of the second part (reliability) has to be added to the first part (restriction). The weighted Z-number is expressed as:

$$\tilde{Z}^\alpha = \{(x, \mu_{\tilde{A}^\alpha}) \mid \mu_{\tilde{A}^\alpha}(x) = \alpha \mu_{\tilde{A}}(x), x \in [0, 1]\}. \tag{15}$$

Lastly, the irregular fuzzy number (weighted restriction) has to be converted to regular fuzzy number \tilde{Z}' , which can be assumed as having the same Fuzzy Expectation of \tilde{Z}^α . If \tilde{A} is a trapezoidal fuzzy number, then \tilde{Z}' is calculated by:

$$\tilde{Z}' = \left(\sqrt{\alpha} \times a_1, \sqrt{\alpha} \times a_2, \sqrt{\alpha} \times a_3, \sqrt{\alpha} \times a_4 \right). \tag{16}$$

Once formalised these preliminaries, we are going to provide in the following an exhaustive description of TODIM. The method was developed by Gomes and Lima [28] to handle MCDM problems on the basis of the prospect theory [36], which enables to model the psychological behavior of decision makers [71].

The traditional version of TODIM shows several shortcomings due to inconsistency of experts' judgment. However, attempts to overcome these limitations have been made, and a generalised TODIM method has been proposed [49]. We will herein expand this generalised method under the Z-numbers environment. The method can be implemented by developing the following steps.

- Defining the decision matrix in terms of Z-numbers $\tilde{X}(\tilde{A}, \tilde{R})$.
- Converting the decision matrix $\tilde{X}(\tilde{A}, \tilde{R})$ into the corresponding matrix of regular fuzzy numbers \tilde{Z} . This matrix collects m alternatives and a set of n criteria.
- Normalising matrix \tilde{Z} using the normalization method regarding different types of criteria such as cost or benefit.
- Calculating the relative weight of criterion C_k to a reference criterion C_r for each $k \in M$, by using the quotient:

$$w_{kr} = \frac{w_k}{w_r}. \tag{17}$$

The reference criterion C_r should be ideally established by the decision maker, even if the criterion with the highest weight is usually considered to such an aim. In this case, we would have $w_r = \max_{l \in M} w_l$.

- Calculate the dominance degree of alternative A_i over alternative A_j under criterion C_k for each $i, j \in N$ and $k \in M$, through the following rules:

$$\Phi_k(A_i, A_j) = \begin{cases} \sqrt{\frac{w_{kr}(z_{jk} - z_{jk})}{\sum_{l=1}^n w_{lr}}} & \text{if } z_{ik} > z_{jk} \\ -\frac{1}{\theta} \sqrt{\frac{(\sum_{l=1}^m w_{lr})(z_{jk} - z_{ik})}{w_{kr}}} & \text{if } z_{ik} < z_{jk}; \\ 0 & \text{if } z_{ik} = z_{jk} \end{cases} \tag{18}$$

where $\theta > 0$ is a parameter determining effects of losses, then allowing decision makers to properly rank alternatives. Specifically, if $\theta < 1$, losses are amplified while, if $\theta > 1$, losses are attenuated. As we will use Z-numbers that will be successively converted to fuzzy numbers, the following equation will be adapted for this use, as suggested in [73]:

$$\Phi_k(A_i, A_j) = \begin{cases} \sqrt{\frac{w_{kr}d(l_{ik} - l_{jk})}{\sum_{k=1}^n w_{kr}}} & \text{if } l_{ik} > l_{jk} \\ -\frac{1}{\theta} \sqrt{\frac{(\sum_{k=1}^n w_{kr})d(l_{ik} - l_{jk})}{w_{kr}}} & \text{if } l_{ik} < l_{jk}; \\ 0 & \text{if } l_{ik} = l_{jk} \end{cases} \tag{19}$$

where the parameter θ indicates the attenuation factor of losses, and $d(l_{ik}, l_{jk})$, measuring the distance between l_{ik} and l_{jk} , is calculated as follows:

$$d(l_{ik}, l_{jk}) = \sqrt{\frac{1}{4} [(\mu_{ik} - \mu_{jk})^2 + (\beta_{ik} - \beta_{jk})^2 + (\sigma_{ik} - \sigma_{jk})^2 + (\tau_{ik} - \tau_{jk})^2]}. \quad (20)$$

If $l_{ik} > l_{jk}$, then $\Phi_k(A_i, A_j)$ represents a gain. If $l_{ik} < l_{jk}$, then $\Phi_k(A_i, A_j)$ represents a loss. The value of $\theta = 2.5$ has been confirmed to give better results [73].

- Calculating the overall performance of alternative A_i for each $i \in N$:

$$\delta(A_i) = \sum_{k=1}^n \Phi_k(A_i, A_j). \quad (21)$$

- Calculating the normalised overall performance of alternative A_i for each $i \in N$:

$$\xi_i = \frac{\delta(A_i, A_j) - \min_{j \in N} \delta(A_i, A_j)}{\max_{j \in N} \delta(A_i, A_j) - \min_{j \in N} \delta(A_i, A_j)}. \quad (22)$$

- Rank risks according to their overall performance values. The higher the overall performance value, the more critical the risk to the system.

4 Case Study

Being aware that in industrial contexts humans resources are exposed to significant risks, whose occurrence could easily lead to catastrophic results, we are going to propose a practical application in the sector of the automotive industry. This choice is motivated by the fact that impacts of human risks dramatically increase in the heavy industry so that they have to continuously be monitored and kept at a minimum level. Connections among human risks identified in the previous section have been evaluated for an automotive company operating in the North of Morocco by involving the responsible of the safety and security system, with a wide proven experience in the field of analysis.

4.1 FCM Based on NHL Algorithm to Derive Total Effects

Linguistic evaluations have been provided by the interviewed expert and collected in the matrix shown in Table 2. Evaluations reflect the existence and the intensity of mutual influence between pairs of those elements that are relevant to our analysis (that are those reported in Table 1). These evaluations have been formulated according to the following linguistic scale: VL (very low), L (low), M (medium), H (high), VH (very high). They have been then translated into trapezoidal fuzzy numbers (TrFN) and successively defuzzified as done in [70].

Table 2 Connection matrix formalising indirect effects for human risks

Risk	E_1	E_2	E_3	W_1	W_2	W_3	W_4	W_5	T_1	T_2	T_3	T_4	T_5	T_6	T_7	M_1	M_2	M_3	M_4	M_5	M_6	M_7	IE	
E_1	0	0	0	0	VL	0	0	0	0	0	0	0	0	0	0	0	VH	L	M	VL	0	0	VL	
E_2	0	0	0	0	L	0	0	0	0	0	0	0	0	0	0	0	VH	0	VL	0	0	0	VL	
E_3	0	0	0	0	H	0	M	0	0	0	0	0	0	0	0	L	H	VL	L	0	0	0	VL	
W_1	0	0	0	0	M	0	L	VL	0	0	0	H	0	0	VH	VL	0	0	0	VL	0	VH	0	VL
W_2	0	0	0	0	0	M	0	0	0	0	0	0	0	0	0	0	0	L	0	0	VH	0	0	L
W_3	M	0	0	0	M	0	0	0	0	0	0	M	0	0	0	0	VH	0	0	0	H	VH	0	M
W_4	0	0	0	L	H	L	0	L	0	0	0	0	0	0	0	M	0	0	VH	VH	0	0	L	
W_5	VH	VH	0	0	VH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	VH	0	0	VH
T_1	VL	VL	L	0	0	M	0	0	0	L	H	L	0	0	0	0	0	0	0	0	VH	M	0	VL
T_2	0	0	0	M	0	0	0	0	0	0	0	0	0	0	VL	L	VH	0	0	0	VH	VH	0	VL
T_3	L	L	VL	0	L	M	0	0	H	0	0	VL	0	0	0	0	0	0	0	0	VH	VH	0	VL
T_4	0	0	0	0	0	M	0	0	M	0	0	0	0	0	0	VH	L	0	VH	0	VH	0	L	
T_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	M	VL	0	VL	
T_6	0	0	0	0	M	0	0	0	0	0	0	0	0	0	0	0	0	H	0	0	M	0	M	
T_7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	M	0	L	0	0	VH	VH	L	
M_1	0	0	0	0	VH	0	0	VH	0	0	0	M	0	0	0	0	0	0	0	0	VH	0	M	
M_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	L	0	0	H	M	0	VH	L	
M_3	0	0	M	VH	0	VH	0	0	0	0	VH	0	L	M	0	L	0	0	0	0	H	0	L	
M_4	0	0	0	0	VH	M	0	0	M	0	H	0	0	0	0	VH	0	VH	0	0	VH	H	M	
M_5	0	0	0	0	VH	0	H	M	0	0	VH	0	0	0	0	M	0	0	0	0	VH	VH	M	
M_6	0	0	0	0	VH	0	0	VH	0	0	0	0	VH	H	H	VH	VH	0	0	0	0	0	M	
M_7	0	0	0	0	VH	L	0	0	H	H	VH	VH	0	0	VH	M	0	VL	L	0	M	0	VL	
IE	VL	VL	VL	L	VL	L	L	L	H	L	H	L	L	L	M	VL	VL	VL	VL	VL	VL	M	M	

The corresponding defuzzified matrix is given in Table 3, while Table 4 shows the matrix resulting from the NHL algorithm iteration. Specifically, IE values of Table 3 reflecting indirect effects between pairs of elements have been obtained through formulas (1), (2) and (3). As previously specified, the corresponding defuzzified IE values reported in Table 4 correspond to those suggested in [70].

As it is possible to observe, applying the NHL algorithm enables to discriminate in a more precise way in terms of total effects' evaluation. Let us consider, for example, the human risk M_6 , which, without the algorithm iteration, had associated a total effect $TE = \max\{0.06, 0.23\} = 0.23$, corresponding to the linguistic evaluation "medium" (M) [70]. After integrating the NHL in the FMC calculations, we can notice as the same human risk has now associated a total effect $TE = \max\{0.06, 0.03\} = 0.06$, corresponding to the linguistic evaluation "very low" (VL) [70]. This is a clear example about how performing the FCM without integrating the NHL would have overestimated the evaluation of human risk M_6 under the total effect criterion, with potential consequent impact on the whole risk assessment procedure. For the sake of precision, Table 5 presents a comparison between the values of TE obtained before (TE_B) and after (TE_A) the NHL algorithm along with the related linguistic evaluations (LE).

Figure 3 has been produced by the Mental Modeler software. The map graphically shows the 252 connections, identified for the 22 elements, what corresponds to some 11.45 connections per element.

4.2 Z-TODIM Prioritising Risks and Discussion of Results

We are going to proceed with the application of Z-TODIM for risk prioritisation. As already explained, we first rank risks to find critical aspects and establish priorities of intervention. The application will be carried out by considering as evaluation criteria TE (obtained by means of FCM-NHL during the previous stage) and the three traditional FMECA parameters, e.g. S , O and D . The last ones have been first directly evaluated by the involved expert using the scales specified in Sect. 3.3), then translated to trapezoidal fuzzy numbers and associated to proper degrees of certainty to be treated as Z-numbers. Table 6 provides the set of input data for the Z-TODIM application. Values under S , O and D have been performed via formula (16), while evaluations under TE correspond to the fuzzy translations of linguistic judgments given in Table 5 (after NHL).

Table 7 shows the final results of the Z-TODIM application along with the ranking of risks, each one of them associated with the related normalised overall performance value ξ_i . These values have been calculated by applying formula (22). Particularly, the final ranking positions of Table 7 have been obtained for risks by ordering the corresponding ξ_i values in a decreasing way, as explained in the last step of the Z-TODIM procedure, detailed in Sect. 3.4. Indeed, as already specified, risks should be ranked according to their overall performance values: the higher the overall performance value, the more critical the risk to the system. Therefore, representing the

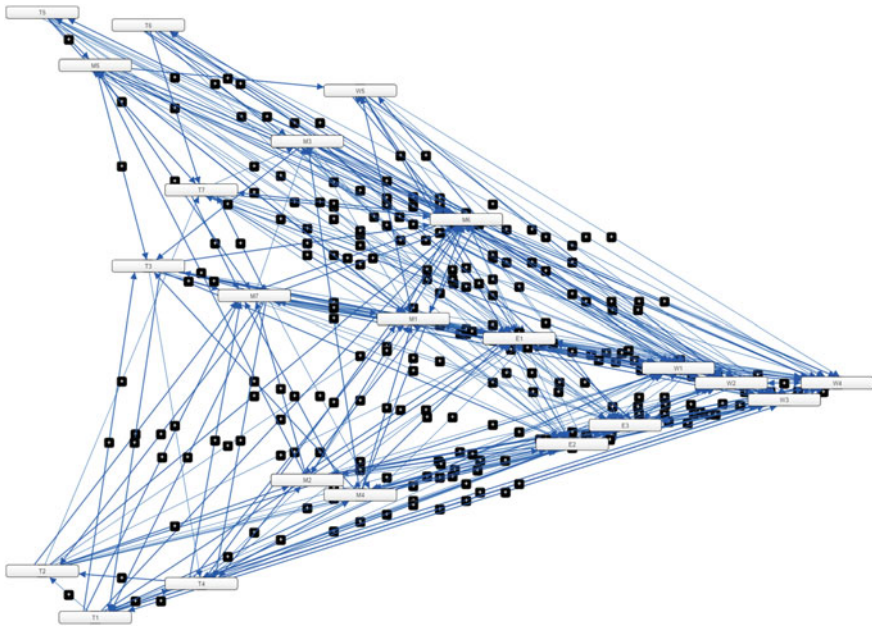


Fig. 3 FCM based on NHL showing relationships among the identified human risks

most critical aspects, those risks with associated higher values of overall performance should be managed with priority to increase such fundamental aspects as safety and security within the context of reference.

Under a practical point of view, it is possible to observe as the human risk T_1 , which is inappropriate use and/or management of machines and plants, belonging to the category of tools and machinery, appears to be the most critical. Proper management actions should be addressed towards the prevention and/or minimisation of this kind of risk.

Workers' training processes should be improved with priority and a safer use of plants should be promoted. Similarly, we can point out as such human risks as T_3 (incorrect use of work equipment) and W_5 (Personal Protective Equipment damaged or improperly used), respectively belonging to the categories of tools and machinery and work space, are considered quite critical, following the risk T_1 in the final ranking. The less critical risks are W_2 (dangerous distribution of work spaces), W_3 (overlapping individual work spaces) and $M2$ (existence of several distraction factors), the first two belonging to the category of work space and the last one belonging to the category of mental load. Organisation of work spaces seems to be quite effective, with consequent higher possibility of avoiding potential distraction sources within the considered industrial context. Related interventions of risk management can certainly be postponed for these kinds of human risks. It is clear as the obtained

Table 5 Comparison of total effects before (TE_B) and after (TE_A) the NHL algorithm iteration

Risk	TE_B	LE	TE_A	LE
E_1	0.06	VL	0.15	L
E_2	0.06	VL	0.06	VL
E_3	0.06	VL	0.06	VL
W_1	0.14	L	0.05	VL
W_2	0.14	L	0.05	VL
W_3	0.23	M	0.03	VL
W_4	0.14	L	0.05	VL
W_5	0.40	VH	0.06	VL
T_1	0.23	M	0.23	M
T_2	0.14	L	0.14	L
T_3	0.23	M	0.23	M
T_4	0.14	L	0.06	VL
T_5	0.14	L	0.14	L
T_6	0.23	M	0.23	M
T_7	0.14	L	0.06	VL
M_1	0.23	M	0.06	VL
M_2	0.14	L	0.14	L
M_3	0.14	L	0.06	VL
M_4	0.23	M	0.06	VL
M_5	0.23	M	0.06	VL
M_6	0.23	M	0.06	VL
M_7	0.23	M	0.23	M

risk prioritisation constitutes a useful driver as well as an effective tool supporting management.

5 Conclusions

This paper applies a hybrid methodological approach to the field of industrial human risk assessment. We first make use of FCMs as effective tools to comprehensively organise knowledge and opinions provided by the involved decision maker(s), expert(s) in the industrial sector of interest. The FCM-based approach is proposed to model relations among relevant human risks by generating spontaneous brainstorming activities, and is herein integrated with the iteration of the NHL algorithm to increase the precision of final results. Specifically, such an approach aims to include the criterion of total effect as a further parameter for risk prioritisation, along with the three risk factors considered in FMECA. In such a way, we enrich risk analyses

Table 6 Input matrix for the TODIM application

Risk	<i>S</i>					<i>O</i>				<i>D</i>				<i>TE</i>		
<i>E</i> ₁	0.00	0.00	0.84	1.67	2.83	3.54	3.54	4.24	3.35	4.18	4.18	5.02	2.00	3.00	4.00	5.00
<i>E</i> ₂	0.84	1.67	1.67	2.51	2.83	3.54	3.54	4.24	3.35	4.18	4.18	5.02	1.00	1.00	2.00	3.00
<i>E</i> ₃	0.84	1.67	1.67	2.51	2.83	3.54	3.54	4.24	3.35	4.18	4.18	5.02	1.00	1.00	2.00	3.00
<i>W</i> ₁	5.86	6.69	6.69	7.53	0.00	0.00	0.84	1.67	0.84	1.67	1.67	2.51	1.00	1.00	2.00	3.00
<i>W</i> ₂	0.00	0.00	0.71	1.41	0.00	0.00	0.84	1.67	0.84	1.67	1.67	2.51	0.00	0.00	1.00	2.00
<i>W</i> ₃	0.84	1.67	1.67	2.51	0.84	1.67	1.67	2.51	4.95	5.66	5.66	6.36	0.00	0.00	1.00	2.00
<i>W</i> ₄	2.83	3.54	3.54	4.24	4.95	5.66	5.66	6.36	4.95	5.66	5.66	6.36	1.00	1.00	2.00	3.00
<i>W</i> ₅	5.86	6.69	6.69	7.53	5.86	6.69	6.69	7.53	5.86	6.69	6.69	7.53	1.00	1.00	2.00	3.00
<i>T</i> ₁	6.52	7.45	7.45	8.38	6.52	7.45	7.45	8.38	3.72	4.65	4.65	5.59	4.00	5.00	5.00	6.00
<i>T</i> ₂	0.00	0.00	0.71	1.41	3.35	4.18	4.18	5.02	0.84	1.67	1.67	2.51	2.00	3.00	4.00	5.00
<i>T</i> ₃	3.35	4.18	4.18	5.02	5.86	6.69	6.69	7.53	5.86	6.69	6.69	7.53	4.00	5.00	5.00	6.00
<i>T</i> ₄	3.35	4.18	4.18	5.02	5.86	6.69	6.69	7.53	0.84	1.67	1.67	2.51	1.00	1.00	2.00	3.00
<i>T</i> ₅	4.95	5.66	5.66	6.36	0.00	0.00	0.71	1.41	5.66	6.36	7.07	7.07	2.00	3.00	4.00	5.00
<i>T</i> ₆	2.83	3.54	3.54	4.24	0.00	0.00	0.71	1.41	5.66	6.36	7.07	7.07	4.00	5.00	5.00	6.00
<i>T</i> ₇	5.86	6.69	6.69	7.53	3.35	4.18	4.18	5.02	5.86	6.69	6.69	7.53	1.00	1.00	2.00	3.00
<i>M</i> ₁	3.35	4.18	4.18	5.02	5.86	6.69	6.69	7.53	6.69	7.53	8.37	8.37	1.00	1.00	2.00	3.00
<i>M</i> ₂	3.35	4.18	4.18	5.02	0.00	0.00	0.84	1.67	0.00	0.00	0.84	1.67	2.00	3.00	4.00	5.00
<i>M</i> ₃	3.35	4.18	4.18	5.02	4.18	5.02	5.86	6.69	3.35	4.18	4.18	5.02	1.00	1.00	2.00	3.00
<i>M</i> ₄	5.86	6.69	6.69	7.53	4.18	5.02	5.86	6.69	3.35	4.18	4.18	5.02	1.00	1.00	2.00	3.00
<i>M</i> ₅	5.86	6.69	6.69	7.53	5.86	6.69	6.69	7.53	4.18	5.02	5.86	6.69	1.00	1.00	2.00	3.00
<i>M</i> ₆	3.35	4.18	4.18	5.02	3.35	4.18	4.18	5.02	4.18	5.02	5.86	6.69	1.00	1.00	2.00	3.00
<i>M</i> ₇	5.86	6.69	6.69	7.53	0.84	1.67	1.67	2.51	6.69	7.53	8.37	8.37	4.00	5.00	5.00	6.00

by taking into account the crucial issue of dependence among risks, something that it is not dealt with by current FMECA.

After evaluating relations among risks and calculating their total effects via the FCM-NHL approach, we collect evaluations of risks under the other FMECA parameters as a set of input data for the subsequent MCDM application. This application aims to propose an alternative way to the RPN calculation and it is based on the Z-TODIM method. The support offered by the Z-numbers theory is useful to further treat uncertainty of subjective evaluations. The output will be a final ranking of risks which will highlight those aspects to be improved with priority to globally enhance the quality of safety and security, which is a fundamental aspect in industry. A case study on the sector of automotive industry is implemented to show the applicability of the proposed approach to real contexts. We claim that our framework can be applied to other business contexts and, in general, to support in a structured and reliable way decision-making problems assisted by human experts.

Possible future developments of the present research may refer to the integration of approaches based on the use of Bayesian Networks for probability computations.

Table 7 Z-TODIM results

Risk	ξ_i	Ranking position
E_1	0.5014	15
E_2	0.4652	16
E_3	0.4650	17
W_1	0.4364	18
W_2	0.0000	22
W_3	0.2678	21
W_4	0.6265	10
W_5	0.8498	3
T_1	1.0000	1
T_2	0.4030	19
T_3	0.8756	2
T_4	0.5518	14
T_5	0.6088	11
T_6	0.5929	12
T_7	0.7617	5
M_1	0.5921	13
M_2	0.3320	20
M_3	0.6486	8
M_4	0.7382	6
M_5	0.8013	4
M_6	0.6285	9
M_7	0.6552	7

Causation linking several critical aspects could be modelled by representing conditional dependence as edges in a directed graph. This approach may be valuable to further consider the human factor within analyses aimed at managing human risk and improving safety and security conditions in industrial environments.

References

1. Amenta P, Lucadamo A, Marcarelli G (2020) On the transitivity and consistency approximated thresholds of some consistency indices for pairwise comparison matrices. *Inform Sci* 507:274–287
2. Amirkhani A, Papageorgiou EI, Mohseni A, Mosavi MR (2017) A review of fuzzy cognitive maps in medicine: taxonomy, methods, and applications. *Comput Methods Program Biomed* 142:129–145
3. Axelrod R (1976) *Structure of decision*. Princeton University Press
4. Azar A, Dolatabad KM (2019) A method for modelling operational risk with fuzzy cognitive maps and bayesian belief networks. *Exp Syst Appl* 115:607–617

5. Bagheri M, Yousefi S, Rezaee MJ (2018) Risk measurement and prioritization of auto parts manufacturing processes based on process failure analysis, interval data envelopment analysis and grey relational analysis. *J Intell Manuf* 29(8):1803–1825
6. Bakhtavar E, Valipour M, Yousefi S, Sadiq R, Hewage K (2020) Fuzzy cognitive maps in systems risk analysis: a comprehensive review. *Complex Intell Syst* 1–18
7. Baykasoğlu A, Gölcük İ (2020) Comprehensive fuzzy FMEA model: a case study of ERP implementation risks. *Oper Res* 20(2):795–826
8. Behrens TE, Muller TH, Whittington JC, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018) What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100(2):490–509
9. Benítez J, Carpitella S, Certa A, Ilaya-Ayza AE, Izquierdo J (2018) Consistent clustering of entries in large pairwise comparison matrices. *J Comput Appl Math* 343:98–112
10. Benítez J, Carpitella S, Certa A, Izquierdo J (2019) Characterization of the consistent completion of analytic hierarchy process comparison matrices using graph theory. *J Multi-Criteria Decis Anal* 26(1–2):3–15
11. Benítez J, Carpitella S, Certa A, Izquierdo J (2020) Constrained consistency enforcement in AHP. *Appl Math Comput* 380:125–273
12. Bertolini M (2007) Assessment of human reliability factors: a fuzzy cognitive maps approach. *Int J Ind Ergon* 37(5):405–413
13. Bevilacqua M, Ciarapica FE (2018) Human factor risk management in the process industry: a case study. *Reliab Eng Syst Saf* 169:149–159
14. Bowles JB, Peláez CE (1995) Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis. *Reliab Eng Syst Saf* 50(2):203–213
15. Carlson C (2012) Effective FMEAs: achieving safe, reliable, and economical products and processes using failure mode and effects analysis (Vol 1). Wiley
16. Carpitella S, Carpitella F, Certa A, Benítez J, Izquierdo J (2018) Managing human factors to reduce organisational risk in industry. *Math Comput Appl* 23(4):67
17. Carpitella S, Izquierdo J (2022) Preference-based assessment of organisational risk in complex environments. In *International symposium on integrated uncertainty in knowledge modelling and decision making*, pp 40–52
18. Carpitella S, Mzougui I, Izquierdo J (2021) Fuzzy cognitive maps for knowledge-oriented human risk management in industry. In: *26th ISSAT international conference on reliability and quality in design*, pp 134–140
19. Carpitella S, Ocaña-Levario SJ, Benítez J, Certa A, Izquierdo J (2018) A hybrid multi-criteria approach to GPR image mining applied to water supply system maintenance. *J Appl Geophys* 159:754–764
20. Chen C-T, Chiu Y-T (2021) A study of dynamic fuzzy cognitive map model with group consensus based on linguistic variables. *Technol Forecast Soc Change* 171120–171948
21. Chen H, Zhang L, Wu X (2020) Performance risk assessment in public-private partnership projects based on adaptive fuzzy cognitive map. *Appl Soft Comput* 93:106–413
22. Das S, Dhalmahapatra K, Maiti J (2020) Z-number integrated weighted VIKOR technique for hazard prioritization and its application in virtual prototype based EOT crane operations. *Appl Soft Comput* 94:106–419
23. Erbay B, Özkan C (2018) Fuzzy FMEA application combined with fuzzy cognitive maps to manage the risks of a software project. *Eur J Eng Formal Sci* 2(2):7–22
24. Fishburn PC (1991) Nontransitive preferences in decision theory. *J Risk Uncertain* 4(2):113–134
25. Flouris T, Yilmaz AK (2010) The risk management framework to strategic human resource management. *Int Res J Financ Econ* 36(1):25–45
26. Francés-Chust J, Brentan BM, Carpitella S, Izquierdo J, Montalvo I (2020) Optimal placement of pressure sensors using fuzzy DEMATEL-based sensor influence. *Water* 12(2):493
27. Ghaleh S, Omidvari M, Nassiri P, Momeni M, Lavasani SMM (2019) Pattern of safety risk assessment in road fleet transportation of hazardous materials (oil materials). *Saf Sci* 116:1–12

28. Gomes L, Lima M (1992) From modeling individual preferences to multicriteria ranking of discrete alternatives: a look at prospect theory and the additive difference model. *Found Comput Decis Sci* 17(3):171–184
29. Hassoun MH et al (1995) *Fundamentals of artificial neural networks*. MIT press
30. He S-S, Wang, Y-T, Peng J-J, Wang J-Q (2020) Risk ranking of wind turbine systems through an improved fmea based on probabilistic linguistic information and the TODIM method. *J Oper Res Soc* 1–14
31. Hebb DO (1949) The first stage of perception: growth of the assembly. *Organiz Behav* 4:60–78
32. Hu Y-P, You X-Y, Wang L, Liu H-C (2019) An integrated approach for failure mode and effect analysis based on uncertain linguistic GRA-TOPSIS method. *Soft Comput* 23(18):8801–8814
33. Huang J, Liu H-C, Duan C-Y, Song M-S (2019) An improved reliability model for FMEA using probabilistic linguistic term sets and TODIM method. *Annals Oper Res* 1–24
34. Huang J, Xu D-H, Liu H-C, Song M-S (2019) A new model for failure mode and effect analysis integrating linguistic Z-numbers and projection method. *IEEE Trans Fuzzy Syst* 29(3):530–538
35. IEC-60812-Technical-Committee (2006) *Analysis techniques for system reliability: procedure for failure mode and effects analysis (FMEA)*. IEC 60812
36. Kahneman D, Tversky A (1980). *Prospect theory*. *Econometrica* 12
37. Kang B, Wei D, Li Y, Deng Y (2012) A method of converting z-number to classical fuzzy number. *J Inf Comput Sci* 9(3):703–709
38. Karunathilake H, Bakhtavar E, Chhipi-Shrestha G, Mian HR, Hewage K, Sadiq R (2020) *Decision making for risk management: a multi-criteria perspective*
39. Kenter JO, O'Brien L, Hockley N, Ravenscroft N, Fazey I, Irvine KN et al (2015) What are shared and social values of ecosystems? *Ecol Econ* 111:86–99
40. Kobis P (2021) Human factor aspects in information security management in the traditional it and cloud computing models. *Oper Res Decis* 1:61–76
41. Kosko B (1986) Fuzzy cognitive maps. *Int J Man-Mach Stud* 24:1,65–75
42. Kutlu AC, Ekmekçiöğlü M (2012) Fuzzy failure modes and effects analysis by using fuzzy TOPSIS-based fuzzy AHP. *Exp Syst Appl* 39(1):61–67
43. Li J, Fang H, Song W (2019) Modified failure mode and effects analysis under uncertainty: a rough cloud theory-based approach modified failure mode and effects analysis under uncertainty: a rough cloud theory-based approach. *Appl Soft Comput* 78:195–208
44. Liu H-C (2016) FMEA using uncertainty theories and MCDM methods. In: *FME using uncertainty theories and MCDM methods*. Springer, pp 13–27
45. Liu, H-C, Chen X-Q, Duan C-Y, Wang Y-M (2019) Failure mode and effect analysis using multi-criteria decision making methods: a systematic literature review. *Comput Ind Eng* 135:881–897
46. Liu H-C, Liu L, Liu N (2013) Risk evaluation approaches in failure mode and effects analysis: a literature review. *Exp Syst Appl* 40:2,828–838
47. Liu H-C, Liu L, Liu N, Mao L-X (2012) Risk evaluation in failure mode and effects analysis with extended VIKOR method under fuzzy environment. *Exp Syst Appl* 39(17):12926–12934
48. Liu P, Li Y (2021) An improved failure mode and effect analysis method for multi-criteria group decision-making in green logistics risk assessment. *Reliab Eng Syst Saf* 107826
49. Llamazares B (2018) An analysis of the generalized TODIM method. *Eur J Oper Res* 269(3):1041–1049
50. Longo F, Padovano A, Nicoletti L, Fusto C, Elbasheer M, Diaz R (2021) Fuzzy cognitive map-based knowledge representation of hazardous industrial operations. *Procedia Comput Sci* 180:1042–1048
51. Mangeli M, Shahraki A, Saljooghi FH (2019) Improvement of risk assessment in the FMEA using nonlinear model, revised fuzzy TOPSIS, and support vector machine. *Int J Ind Ergon* 69:209–216
52. Mazzuto G, Antomarioni S, Ciarapica F, Bevilacqua M (2021) Health indicator for predictive maintenance based on fuzzy cognitive maps, grey wolf, and k-nearest neighbors algorithms. *Math Probl Eng* 2021
53. Mohsen O, Fereshteh N (2017) An extended VIKOR method based on entropy measure for the failure modes risk assessment-a case study of the geothermal power plant (GPP). *Saf Sci* 92:160–172

54. Mourhir A (2020) Scoping review of the potentials of fuzzy cognitive maps as a modeling approach for integrated environmental assessment and management. *Environ Modell Softw* 104891
55. Mourhir A (2021) Scoping review of the potentials of fuzzy cognitive maps as a modeling approach for integrated environmental assessment and management. *Environ Modell Softw* 135:104891
56. Mpelogianni V, Groumpos PP (2019) Building energy management system modelling via state fuzzy cognitive maps and learning algorithms. *IFAC-PapersOnLine* 52(25):513–518
57. Murray SL, Thimngan MS (2016) Human fatigue risk management: improving safety in the chemical processing industry. Academic Press
58. Mzougui I, Carpitella S, Certa A, Felsoufi ZE, Izquierdo J (2020) Assessing supply chain risks in the automotive industry through a modified MCDM-based FMECA. *Processes* 8(5):579
59. Mzougui I, Elfelsoufi Z (2019) Improvement of failure mode, effects, and criticality analysis by using fault tree analysis and analytical hierarchy process. *J Fail Anal Prevent* 19(4):942–949
60. Oja E (1991) Learning in non-linear constrained hebbian networks. In: *Proceedings of the ICANN'91*, pp 385–390
61. Onari MA, Yousefi S, Rezaee MJ (2021) Risk assessment in discrete production processes considering uncertainty and reliability: Z-number multi-stage fuzzy cognitive map with fuzzy learning algorithm. *Artif Intell Rev* 54(2):1349–1383
62. Panchal D, Singh AK, Chatterjee P, Zavadskas EK, Keshavarz-Ghorabae M (2019) A new fuzzy methodology-based structured framework for ram and risk analysis. *Appl Soft Comput* 74:242–254
63. Papageorgiou E, Stylios C, Groumpos P (2003) Fuzzy cognitive map learning based on non-linear hebbian rule. In: *Australasian joint conference on artificial intelligence*, pp 256–268
64. Papageorgiou EI (2011) Review study on fuzzy cognitive maps and their applications during the last decade. In: *2011 IEEE international conference on fuzzy systems (fuzz-IEEE)*, pp 828–835
65. Papageorgiou EI, Groumpos PP (2005) A weight adaptation method for fuzzy cognitive map learning. *Soft Comput* 9(11):846–857
66. Papageorgiou EI, Stylios C, Groumpos PP (2006) Unsupervised learning techniques for fine-tuning fuzzy cognitive map causal links. *Int J Hum-Comput Stud* 64(8):727–743
67. Papakostas GA, Koulouriotis DE, Polydoros AS, Tourassis VD (2012) Towards Hebbian learning of fuzzy cognitive maps in pattern classification problems. *Exp Syst Appl* 39(12):10620–10629
68. Peer M, Brunec IK, Newcombe NS, Epstein RA (2020) Structuring knowledge with cognitive maps and cognitive graphs. *Trends Cogn Sci*
69. Peláez CE, Bowles JB (1996) Using fuzzy cognitive maps as a system model for failure modes and effects analysis. *Inf Sci* 88(1–4):177–199
70. Poomagal S, Sujatha R, Kumar PS, Vo D-VN (2020) A fuzzy cognitive map approach to predict the hazardous effects of malathion to environment (air, water and soil). *Chemosphere* 263:127926
71. Qin J, Liu X, Pedrycz W (2017) An extended TODIM multi-criteria group decision making method for green supplier selection in interval type-2 fuzzy environment. *Eur J Oper Res* 258(2):626–638
72. Ravasan AZ, Mansouri T (2014) A FCM-based dynamic modeling of ERP implementation critical failure factors. *Int J Enterp Inf Syst (IJEIS)* 10(1):32–52
73. Ren P, Xu Z, Gou X (2016) Pythagorean fuzzy TODIM approach to multi-criteria decision making. *Appl Soft Comput* 42:246–259
74. Rezaee MJ, Yousefi S, Valipour M, Dehdar MM (2018) Risk analysis of sequential processes in food industry integrating multi-stage fuzzy cognitive map and process failure mode and effects analysis. *Comput Ind Eng* 123:325–337
75. Rybalkina A, Enikeev R (2021) Fatigue management methodology in aircraft maintenance as a way of reducing errors related to the human factor. In: *Matec web of conferences*, vol 341, p 00006

76. Sagnak M, Kazancoglu Y, Ozen YDO, Garza-Reyes JA (2020) Decision-making for risk evaluation: integration of prospect theory with failure modes and effects analysis (FMEA). *Int J Qual Reliab Manag*
77. Salmemon JL, Mansouri T, Moghadam MRS, Mardani A (2019) Learning fuzzy cognitive maps with modified asexual reproduction optimisation algorithm. *Knowl-Based Syst* 163:723–735
78. Schwartz F, Epinat-Duclos J, Léone J, Poisson A, Prado J (2020) Neural representations of transitive relations predict current and future math calculation skills in children. *Neuropsychologia*, 107410
79. Sharma RK, Kumar D, Kumar P (2005) Systematic failure mode effect analysis (FMEA) using fuzzy linguistic modelling. *Int J Qual Reliab Manag*
80. Sudjianto A, Hassoun MH (1995) Statistical basis of nonlinear Hebbian learning and application to clustering. *Neural Netw* 8(5):707–715
81. Tadić S, Zečević S, Krstić M (2014) A novel hybrid MCDM model based on fuzzy DEMATEL, fuzzy ANP and fuzzy VIKOR for city logistics concept selection. *Exp Syst Appl* 41(18):8112–8128
82. Tian X, Li W, Liu L, Kou G (2021) Development of TODIM with different types of fuzzy sets: a state-of-the-art survey. *Appl Soft Comput* 107661
83. Tversky A (1969) Intransitivity of preferences. *Psychol Rev* 76(1):31
84. Wang J, Wei G, Lu M (2018) TODIM method for multiple attribute group decision making under 2-tuple linguistic Neutrosophic environment. *Symmetry* 10(10):486
85. Wang L, Liu Q, Dong S, Soares CG (2019) Effectiveness assessment of ship navigation safety countermeasures using fuzzy cognitive maps. *Saf Sci* 117:352–364
86. Wang L-E, Liu H-C, Quan M-Y (2016) Evaluating the risk of failure modes with a hybrid MCDM model under interval-valued intuitionistic fuzzy environments. *Comput Ind Eng* 102:175–185
87. Wang Y (2008) On cognitive properties of human factors and error models in engineering and socialization. *Int J Cogn Inform Natural Intell (IJCINI)* 2(4):70–84
88. Wright BC (2012) The case for a dual-process theory of transitive reasoning. *Dev Rev* 32(2):89–124
89. Wu Z, Tu J (2021) Managing transitivity and consistency of preferences in AHP group decision making based on minimum modifications. *Inf Fusion* 67:125–135
90. Xie X, Guo D (2018) Human factors risk assessment and management: process safety in engineering. *Process Saf Environ Protect* 113:467–482
91. Zadeh LA (2011) A note on z-numbers. *Inf Sci* 181(14):2923–2932
92. Zhou Q, Thai VV (2016) Fuzzy and grey theories in failure mode and effect analysis for tanker equipment failure prediction. *Saf Sci* 83:74–79

Ilyas Mzougui Abdelmalek Essaadi University, Tangier, Morocco

Dr. IlyasMzougui received a Ph.D. in Sciences and Engineering Techniques by defending a thesis entitled “Study of the FMECA limitations and proposition of an improved method for risk assessment”. He has been working at the Department of Mechanical Engineering of the Abdelmalek Essaadi University in Morocco. Ilyas does research in Mechanical Engineering, Manufacturing Engineering, and Industrial Engineering. His current project refers to the proposition of a modified FMEA to improve the reliability of products.

Silvia Carpitella California State University, Northridge, CA

Dr. Silvia Carpitella holds a Ph.D. in Technological Innovation Engineering and a Ph.D. in Mathematics. She has been an Assistant Professor at the Department of Manufacturing Systems Engineering and Management of the California State University, Northridge, since August 2022. Silvia’s activity has been internationally recognised by means of two prizes awarded to her doctoral thesis, the first one in Spain and the second one in Italy, and through the Otto Wichterle award, granted from the Czech Academy of Science for the contributions to the advancement of scientific knowledge in her field.

Joaquín Izquierdo Universitat Politècnica de València, Valencia, Spain

Dr. Joaquín Izquierdo, Ph.D. in Mathematics and Full Professor (currently Emeritus Professor) of Applied Mathematics at the Universitat Politècnica de València (UPV). Educational career developed in the UPV and research activity as the Director of the research group FluIng, devoted to mathematical modelling, knowledge-based systems, and DSSs in Engineering. Author and editor of several books and author or co-author of more than 300 research papers, book chapters, and contributions to international events. Tutor of 16 Doctoral Dissertations, wide experience in consulting and R&D projects.

Data Quality Assessment for ML Decision-Making



Alexandra-Ștefania Moloiu, Grigore Albeanu, Henrik Madsen,
and Florin Popențiu-Vlădicescu

Abstract Data quality has a strong effect on the design, validation and testing of decision-making systems. New paradigms of future models in the knowledge society need to analyze clean, complete, consistent, and high-quality data. This paper presents three case studies from different fields in which models are constructed using machine learning strategies. Projects on text recognition, electrocardiogram-based identification and data analysis are described in relation to input data quality and system performance.

Keywords Quality factors · Time series · Neural networks · Text recognition · Electrocardiogram based identification · Data analysis

1 Introduction

Understanding the effect of data quality indicators on a system model behavior is an important task before designing explicable decision-making systems in context of big data analytics based on artificial intelligence methodologies [8, 14, 38].

According to Almeida et al. [2], data quality is one of the most challenging factor affecting the decision-making process, and the multivariate data should be analyzed to evaluate not only for classical quality dimension indicators, but also for the *index*

A.-Ș. Moloiu
R&D Department, TypingDNA, Bucharest, Romania
e-mail: alexandra_moloiu@yahoo.com

G. Albeanu
“Spiru Haret” University, Bucharest, Romania
e-mail: g.albeanu.mi@spiruharet.ro

H. Madsen
Danish Technical University, Lyngby, Denmark
e-mail: hmada@dtu.dk

F. Popențiu-Vlădicescu (✉)
University “Politehnica” of Bucharest & Academy of Romanian Scientists, Bucharest, Romania
e-mail: popentiu@imm.dtu.dk

value that determine the models' interpretability [5] and explicability [12], mainly for the case of *black-box methods*. Also, using an index to measure the data quality is important for improving the performance of *case-based reasoning systems* (CBRS), as shown by Hönigl and Küng [16].

Ensuring explicability, it calls for appropriate models of such methods that are able to describe the reasons for the model's prediction-accuracy, ensure user confidence, or provide performance indicators for decision-making chains. The completion of such mechanisms is necessary to provide the ability to defend end-users actions, to give relevant answers to various questions and the system to be audited [12]. Users have to consider a deep understanding of the machine and they should trust the results and output created by machine learning algorithms. An example of an information system that can be explained is IBM Watson Studio [46], which "empowers data scientists, developers and analysts to build, run and manage AI models, and optimize decisions".

Usually, the lack of data quality refers to data missing, incompleteness, inconsistency, and inaccurate items [30]. According to [14], there are many approaches to address the missing data challenge: *mean substitution* (using the mean of the observed values), *multiple imputation* (sampling from data distribution of existing values, etc.), *checking consistency* (mainly between variables collected from different sources), and *accuracy improvement* (using right data sources, efficient data capturing, efficient coding and reliable data entry, data reviewing, automated error reports generation, adopting pre-defined data standards).

In the machine learning projects, data quality is assessed before the model is built as well as after [14, 40, 42]. The triadic approach in data partitioning concerning training, validation and testing asks for high quality data. As Marcheggiani and Sebastiani mentioned in [26], there is a strong impact of the quality of training data on the accuracy of information extraction systems.

The quality of the surveyed data was discussed by Timmermann and Bronselaer [44]. They proposed a kind of expert system for assessing the quality of the data. Shome et al. [41] found that the performance of any speech-based system depends on the quality of the input voice signals. It will be asked for high quality signals when working with electrocardiograms.

This chapter is a revisited and extended version of [30] and gives a broader view on the effect of quality factors on some machine learning models and presents some appropriate case studies covering input quality in decision based systems working on text, electrocardiograms, and timeline sequences. The mentioned projects are based on [27–29]. The third case study is inspired by the recent efforts to understand timeline data (data items where the value can vary over time [13, 44, 48]).

Recent metrics are used for data quality assessment. In the next sections we give a new interpretation of the behavior and performance of these decision-making systems taking into account data quality.

2 Data Quality Assessment: Recent Metrics

When analyzing data, the scientist has to process three types of inputs [39]: *structured data* (following a predefined data model: arrays, relation table/SQL databases or equivalent), *unstructured data* (not organized in a predefined manner but given as a sequence of symbols to be interpreted according to its nature: texts, images, sounds, videos, three-dimensional data, geographic information system data, aerial photographs, weather satellite images, and cartographic satellite images, NoSQL databases), and *semi-structured data* (containing self descriptions like JSON and XML). Timeline data can be modeled as structured data when the data volume is small, or as NoSQL database (as Grid DB [13]) when using large collections of timeline data as appeared in Internet of Things (IoT) [48] and Big Data [14, 45]. In IoT context the data team experiences frequently changed data due to the velocity characteristic. Also the large volume of data may impose data reduction procedures. Such transformation should not decrease the output accuracy.

Pipino et al. [36], claim that data quality is a multi-dimensional concept and data quality assessment is a continuous effort that requires awareness of the fundamental principles underlying the development of data quality metrics. Also, data quality assessment depends on what a team wants to obtain as output. The following 16 dimensions were considered in [36]:

- *Accessibility* (available or quickly retrievable);
- *Appropriateness of the amount of data* (the required volume to provide a significant output);
- *Believability* (make use of true/credible data);
- *Completeness* (all attributes are available if multidimensional type);
- *Conciseness of the representation* (compactness);
- *The representation consistency* (based on some template);
- *Easiness in data manipulation* (or transformation);
- *Flawless and faultless* (free-of-error, accuracy, correctness and reliability);
- *Interpretability* (follow a clear definition based on a well established language);
- *Objectivity* (being unbiased, unprejudiced, and impartial);
- *Relevancy* (good for solving the task);
- *Reputability* (related to its source or content);
- *Security* (degree of protection to prevent unauthorized modification);
- *Timeliness* (being up-to-date);
- *Understandability* (easily comprehended);
- *Ability for value-added contribution* (it is important with a high degree to obtain the result).

Reviewing literature on data quality in IoT environments, Zhang et al. [48] identified 24 dimensions of data quality. The new dimensions are related to:

- *Traceability* (along the lifecycle of data from a unique/multi-source, cleansing operations and data under processing);

- *Portability* (data are collected in such a manner to be processed on many platforms and/or by many services);
- *Recoverability* (ability to restore lost or inaccessible data);
- *Speed of generation* (in the framework of the new WIFI Internet, the speed of acquired data is increased generating huge volume of data);
- *Latency* (related to the ability to store/retrieve data in short time);
- *Drop rate* (the probability of losing the connection to the data source);
- *Efficiency* (when referring to the process of preparing data for efficient storage, access, retrieval, processing, and general management. This dimension include the easiness of manipulation already mentioned in [35]);
- *Response time* (the elapsed time between launching a request to data reading and the response).

When reporting/reading data, by completeness it is asked for all reportable/readable items and all acquired attributes to be used to retrieve items when necessary (for instance, links to files/servers, links to sensors etc.). For a data record, completeness means that all attributes are available, while for a data report, completeness requires that all data records are reported. Any data quality assessment procedure/standard should check for timeliness, validity, integrity, precision, and reliability. Not only timeliness are required but also, availability, completeness (all necessary items are reported) and relevance for the decision-making process. Always validity is connected to the objectives. It is imported to provide results based on real and high quality data, or simulated data.

Data integrity can be compromised during collection and digital transfer. Any cybersecurity attacks working against a data server is able to deteriorate the initial collected data.

Many equivalent traditional data quality metrics, such as “free-of-error”, “relevancy”, “completeness”, “consistency”, “concise representation”, and “ease of manipulation” are based on simple *ratio principle* i.e. the number of undesirable outcomes divided by total outcomes subtracted from 1:

$$\text{quality metric} = 1 - \text{undesirable}/\text{total}.$$

Data interpretation refers to the implementation of the processes by which data is examined if it follows the correct specification.

The quality indicators for “believability” and “appropriate amount of data” are based on min operator, while “timeliness” and “accessibility” are based on max operator.

If there are n sources for similar data to be used for decision making but with different degrees of believability, the believability metric of entire data collection is the minimum of believability degrees. Similar rule can be applied for degree of appropriateness amount of data. Contrary, for other metrics, the most suited operator is based on maximum value of individual indicators.

Other aspects related to the dimensions of quality in context IoT can be found in [48], where frameworks and standards of IoT data quality are compared.

Heinrich et al. [15] were interested in using *distance based indicators* to design innovative metrics for the data quality dimensions “correctness” (free-of-error) and “timeliness”, in a goal-oriented manner.

Jang et al. [18] have investigated a quality index calculation model with structured and unstructured data. Two indexes are defined: the attribute value quality index (AVQI), and the structured data value quality index (SDVI). AVQI is used to discover the attribute with high error possibility. In order to compute AVQI of the i th attribute in the data table, a weighting procedure is proposed in [18] according to the following cases and criteria:

- A. In the case of missing data the weight α is derived according to the β criterion in the rule R_β :
 - R1. If no missing data then $\alpha = 0$.
 - R2. If Number of missing values $\in (0, 5\%]$ then $\alpha = 1.2$.
 - R3. If Number of missing values $\in (5, 15\%]$ then $\alpha = 1.5$.
 - R4. If Number of missing values $> 15\%$ then $\alpha = 2$.
- B. For the case of outliers, a decision is taken based on Z-score (measured in numbers of standard deviations from the average [47]). The weight α is derived according to the β criterion in the following R_β rules:
 - R5. If Z-score $\in [0, 2]$, then $\alpha = 0$.
 - R6. If Z-score $\in (2, 3]$, then $\alpha = 1.2$.
 - R7. If Z-score $\in (3, 4]$, then $\alpha = 1.5$.
 - R8. If Z-score > 4 , then $\alpha = 2$.

If the data table is given as an array $X[n][m]$, where n is the number of measured items, and every time a measurement consists of m attributes, for an attribute i , $1 \leq i \leq m$, $AVQI_i$ is given by.

$$AVQI_i = \left| 1 - \frac{u_i}{v_i} \right|,$$

where $u_i = \sum_{k=2..8; k \neq 5} U_k$, $v_i = \sum_{k=2..8; k \neq 5} V_k$ with $U_k =$ Number of data from column i satisfying the R_k -rule multiplied by the corresponding α , and $V_k =$ Number of data from column i satisfying the R_k -rule multiplied by 1.0.

For instance, we compute the following AVQI for $n = 1000$ items and $m = 5$ attributes A_i , having missing values/outliers as given in Table 1.

The AVQI formula can be generalized for more β criteria, following the above model. The calculation formula for the structured data value quality index (SDVI) becomes the following:

$$SDVI = \left| 1 - \frac{\sum_{k=1..m} u_k}{\sum_{k=1..m} v_k} \right|,$$

Table 1 Missing values/outliers extracted information for AVQI evaluation

β	α	A1	A2	A3	A4	A5
1	0	928	966	915	1000	948
2	1.2	72	0	85	0	0
3	1.5	0	34	0	0	52
4	2	0	0	0	0	0
5	0	843	865	993	934	995
6	1.2	150	0	7	65	0
7	1.5	6	135	0	0	4
8	2	1	0	0	1	1
AVQI		0.211	0.5	0.2	0.212	0.509

where u_k and v_k are given above. After computation, the above database SDVI is $\left|1 - \frac{807.3}{613}\right| = 0.317$. The lower the index value, the poor quality is indicated.

Hönigl and Küng [16] have defined three indices to be used to build an index for the quality of case base of CBRS [16]. The first one is given by the ratio of bad cases and the size of the case database subtracted from 1. The second is the ratio of the count of similar retained queries and total number of pairs used for similarity checking, also subtracted from 1. The third value is the ratio of the count of fields with missing data and the total number of fields, subtracted from 1. Finally, the main index is obtained by a weighted average of the three indices.

Another type of metric is based on analogy or distance. Depending on the type of data there are the well-known geometrical distances:

- *Euclidean* distance—the square root of the sum of square of differences;
- *Manhattan* distance—the sum of absolute values of differences;
- *Chebyshev* distance—the maximum of the absolute values of differences;
- The p -distance— $(\sum(x[i] - [y[i]]^p)^{1/p})$, with $p \geq 1$,

as well as:

- text distances Levenshtein and Damerau-Levenshtein [10], Hamming [43], Lee (see [17]), Jaro-Winkler similarity index [19],
- set-based distances (the most popular being Hausdorff-Pompeiu [9]) and
- statistical distances (the most popular being Mahalanobis [25]).

To give some computing examples, let us consider two objects a and b given by five numerical attributes $a = (1.2; 1.5; 1.0; 2.0; 1.0)$, and $b = (1.19; 1.51; 0.99; 2.01; 1.01)$. The geometrical distances between objects a and b are given in Table 2.

If the objects are of text type, like $a = \text{“Data Quality”}$, and $b = \text{“Datenqualität”}$, the Levenstein distance (i.e. the minimum number of operations like insertions, deletions or substitutions of a single character, required to change one text object into the other) between a and b is 4.

Table 2 Computing geometric distances

<i>Euclidean</i> distance	$D_2(a, b) = ((1.2-1.19)^2 + (1.5-1.51)^2 + (1.0-0.99)^2 + (2.0-2.01)^2 + (1.0-1.01)^2)^{0.5} = 0.0024$
<i>Manhattan</i> distance	$D_1(a, b) = 1.2-1.19 + 1.5-1.51 + 1.0-0.99 + 2.0-2.01 + 1.0-1.01 = 0.05$
<i>Chebyshev</i> distance	$D_\infty(a, b) = \max(1.2-1.19 , 1.5-1.51 , 1.0-0.99 , 2.0-2.01 , 1.0-1.01) = 0.01$
The 4-distance	$D_4(a, b) = ((1.2-1.19)^4 + (1.5-1.51)^4 + (1.0-0.99)^4 + (2.0-2.01)^4 + (1.0-1.01)^4)^{0.25} = 0.0150$

Hamming distance is used to compare two data objects having the same length and is the number of positions at which the corresponding values are different. Mainly is used for binary objects, but can also be used in bioinformatics to analyze multiple DNA or protein sequences [3]. For instance if $a = "1,112,998,899"$ and $b = "1,112,998,899"$, the Hamming distance between a and b is 3.

However, other indicators from statistics (the first order temporal correlation coefficient, the Pearson’s correlation factor, metrics based on autocorrelation or partial autocorrelation [21, 31, 32]) and information theory (likethe Kullback–Leibler divergence [22]) can be used.

In order to classify time series of data collected along some timeline in order to understand not only the structure but also to reduce the number of sensors used to monitor some phenomenon or process, specific dissimilarity measures for time series may be used. The most popular dissimilarity measures of two time series are [31–33]:

- Fréchet distance (as the minimal cost of a walk that traverses both timelines from beginning to end without backtracking, taking into account the location and ordering of the points along the timeline, with *Chebyshev* distance applied for a walk).
- Dynamic Time Warping (DTW) distance (similar to Fréchet distance, but computed using *Manhattan* distance).
- Constrained Dynamic Time Warping (cDTW) distance (by considering limited alignments to prevent excessive alignments between two time series [37]). Recently, the Fast Constrained Dynamic Time Warping (FcDTW) distance (operating within a fixed window area of the cDTW) was proposed [6].

The similarity of data sources is not equivalent to reliability of data which depends on various factors including data registration and communication. In order to establish the similarity of data sources in IoT based applications, the clustering of time series is used [31, 33]. In practice, distances like Euclidean, DTW and its variants have good behavior when considering the time complexity of the algorithms.

Reliable data are necessary when applying various models in order to establish strategies or assess risk. Outlier detection is an important task before any data analysis process and has various applications on digital time series data such as fraud

detection, fault detection, and data modification attack detection. Outlier detection by machine learning techniques succeeds in various projects (see [23]).

To give a view on the influence of data quality on final decision, in the following, three case studies will be presented belonging to different fields of application, but asking for good data quality: text recognition, electrocardiogram-based authentication, and data analysis.

3 Quality Metrics in Text Recognition

There are a number of factors affecting the text recognition accuracy: scanner characteristics, type of documents (laser printer, photocopied (including handwritten text)), and their linguistic complexity, paper quality, fonts used in the text, and dictionary of similarities. The accuracy of printed text recognition exceeds 99%, while handwriting text recognition (HTR) remains for improvements. Recognition of text from various documents, street images (advertising) or artistic imagery requires pre-processing steps to localize the text: applying image enhancement techniques to remove noise and increase the contrast in the image, removing the background containing any other scenes based on a specific threshold, separation of graphics from text, character segmentation to separate characters from each other, and morphological processing to enhance the characters appearance.

Mainly, the problem with character recognition is to divide the space of the images into sharp classes, each class corresponding to a unique ASCII code and representing an instance of the same character [27, 29]. Objects belonging to the same class are given by the “equality” of the degree of similarity with the class. Research and business applications use two methods for character recognition. The first method uses a large number of elements in the space of the minimally processed image code and encodes and analyzes the class separation hyperspace using a neural network. The second method uses elements of character properties expressed in the form of vectors, uses several image processing techniques, and carefully apply feature extraction methods.

The effectiveness of the back-propagation method used during training of the multilayer artificial neural network (ANN) was demonstrated using normalized binary images, according to the following ideas used during implementation:

- the image was resized (without distortion) until it has adequate width/height.
- the gray-values of the image were normalized (to remove shadows and to balance brightness and contrast by illumination compensation).
- the cursive style was transformed by a deslanting procedure.
- the dataset-size was augmented by applying further (random) transformations to the input images.

The HTR methodology applies partitioning (training, validation, and test) of the data, all partitions being preprocessed as above. Both Character Error Rate (CER), and Word Error Rate (WER) were used to validate the results.

These metrics were calculated using the Damerau-Levenshtein distance [10] between the desired result and the prediction, computed as $(i + s + d)/n$, where n is the total number of $c(w)$, i is the minimal number of $c(w)$ inserted, s is the number of $c(w)$ substituted, and d is the number of $c(w)$ deleted in order to transform the output text into the target, with $c(w)$ representing the characters (words) under recognition process.

Replacing the classical ANN approach to deep learning based on Convolution Neural Networks (CNN) has improved the performance of recognition both in speed and from a robustness point of view [38]. Convolutional Networks proved value when also were used for ECG based biometric identification, as shown in next section.

4 Quality Metrics in ECG Based Biometric Identification

Biometric identification through electrocardiogram (ECG) is a recent approach (as an alternative to fingerprint or iris-based analysis) as shown by [4, 28, 35].

According to [29], there are three types of ECG biometric techniques:

- (1) The first one is based on Fiducial Point detection: the feature wave duration, its slope, the curvature, its amplitude, and the direction are obtained and saved as registered data to be identified during the authentication phase.
- (2) The second technique will extract ECG features in the frequency domain using various signal transformations (wavelet, Fourier, and discrete cosine).
- (3) The third technique makes use of ANN and deep learning [28, 38].

The ECG based Biometric Recognition Algorithm (ECG-BRA) works on ECG input signals and provides as output the answer to the identification problem (accept/reject). It uses a database of templates (supporting the addition of a new person, and to search for a person).

The main modules of any ECG-BRA system are responsible for signal enhancement through noise elimination, feature extraction, enrollment, identification (authentication), and decision. A preprocessing step can be used to assess the quality of the ECG signal. For instance, in the case of ECGs collected using mobile phones, the four flags method of Liu et al. [24] or multi-expert annotation method [34] can be used.

The authors of [24] associate to every ECG an indicator called single signal quality index (SSQI) and the integrative signal quality index (ISQI) for multi-lead ECGs. If the ISQI of multi-lead ECGs is more than 10, the ECGs are identified as acceptable. Otherwise, the ECGs are identified as unacceptable.

The results of the ECG-BRA system described in [28] were based on several test series that are analyzed and compared using multilayer perceptron and deep learning. The Physionet.org's free ECG database [34] was used for training, validation and testing. A number of 310 unprocessed ECG signals for 90 people with a variable number of data records (20 s with a resolution of 12 bits and a sampling rate of 500 Hz) per person were selected.

The filtered records (considered as acceptable ECGs) from the database that were previously analyzed using plots for extraction are used in a total of 280 records from 81 people, allocated to the partitions for training, validation and testing. For the multilayer perceptron, the most stable and best results were achieved with 5 times more neurons in the hidden layer than with the number of training items plus 100.

The architecture of the convolution networks with the smallest vibrations consists of a convolution layer with the number of filters on the first layer corresponding to half the number of training items plus 2, followed by a Maxpool layer 2 and a dense layer with 50 activation units more than the number of classes. The number of iterations was selected as twice the number of classes plus 100 in order to achieve a correct classification of the training set and to prevent over-learning.

5 ML Based Timeline Data Analysis

In a classical analysis of data by regression, Klein and Rossin [20] found that data quality can affect the prediction accuracy of linear regression models in two ways. First, the training data used to estimate the model structure and its parameters can contain errors. Second, even if the training data does not contain errors, after using a linear regression model for prediction, a user can enter test data that contains errors in the model. One unexpected conclusion of [20] claims that perfectly accurate data may not always provide the best forecast. This is because regression models are not interpolation methods, but only data fitting approaches under some strong probabilistic assumptions.

Recently, Almeida et al. [2] explored the nature of the correlation in data to estimate the data quality to be used in decision-making processes. Also, they are promoting the gauge repeatability and reproducibility study to effectively identify possible data of poor quality leading to measurement errors.

In the following, we propose an empirical method to study the reliability of data collections. Let $d[1], d[2], \dots, d[n]$ be a number of sources of data (n at least 2) about the same event, registered as timeline sequences. To evaluate the reliability of sources, the following procedure, based on the Machine Learning (ML) approach, can be used. If $d[i, 1], d[i, 2], \dots, d[i, m[i]]$ is the data set from source i ($1 \leq i \leq n$), build a ML model to learn the timeline sequence $d[i]$. Validate the model and register the model hyper-parameters as an array $p[i]$. The number of parameters is the same for all models. Using a distance-based formula between $p[i]$ and $p[j]$, the similarity between $d[i]$ and $d[j]$ ($1 \leq i, j \leq n$) can be obtained. The sources are similar if the quality indicators are similar. Clustering algorithms can be used to classify the data sources using ideas presented by [1, 31–33].

In order to describe the idea, let be $n = 2$, $d[1] = y$ (the first time series), and $d[2] = z$ (a modified time series simulating the second data source). Using the R-package neuralnet [7, 11], the neural nets presented in Figs. 1 and 2, have learnt the associated data (x, y) , respectively (x, z) , from Table 3.

Fig. 1 The neural net associated with d[1]

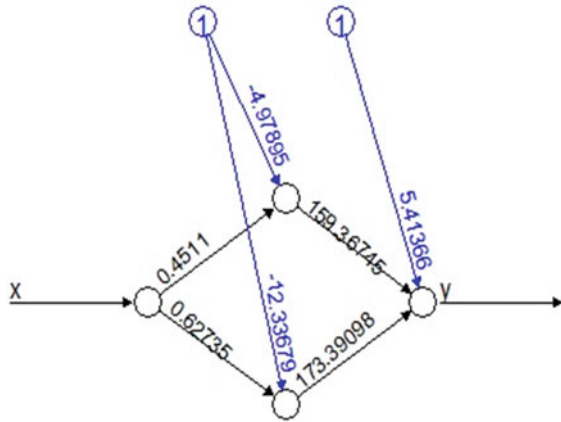
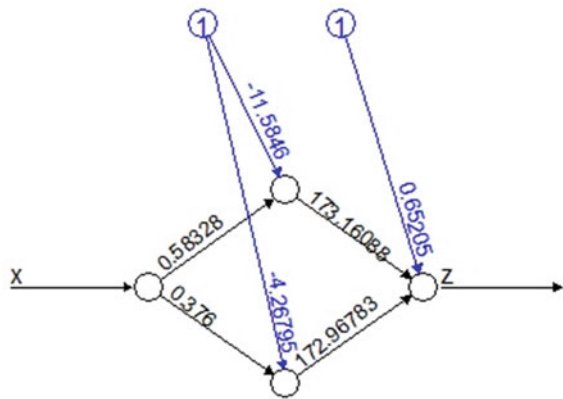


Fig. 2 The neural net associated with d[2]



Using the TS dist R package it is found that $\text{EuclideanDistance}(y, z) = 17.67$, and $\text{EuclideanDistance}(yp, zp) = 9.73$. If $p[1] = (-4.98, -12.34, 5.41, 0.45, 0.63, 159.37, 173.39)$, and $p[2] = (-11.58, -4.27, 0.65, 0.58, 0.38, 173.16, 172.97)$, then $\text{EuclideanDistance}(p[1], p[2]) = 23.07$. Moreover, $\text{EuclideanDistance}(y, yp) = 21.97$, $\text{EuclideanDistance}(z, zp) = 15.89$.

Based on DTW distance, it is found that $\text{DTW}(y, z) = 47$, $\text{DTW}(yp, zp) = 62.49$, $\text{DTW}(y, yp) = 137.81$ and $\text{DTW}(z, zp) = 100.93$.

Similarly, $\text{FréchetDistance}(y, z) = 17$, and $\text{FréchetDistance}(yp, zp) = 4.61$, while $\text{FréchetDistance}(y, yp) = 12.16$, and $\text{FréchetDistance}(z, zp) = 8.54$.

When consider Damerau-Levenshtein metric adapted for real numbers (edit distance) it is found that even y and z are almost similar, the procedure found series as different.

As seen above, the Fréchet distance based on min-max operator has small values compared against DTW based on min-sum operator. However, Euclidean distance has a good behavior and can be used only when the series has the same timeline as in

Table 3 Input data (y, z) versus predicted data (yp, zp)

x	y	yp	z	zp
1	2	7.12	2	4.11
2	7	8.07	4	5.65
3	11	9.55	8	7.84
4	13	11.82	13	10.93
5	17	15.25	17	15.23
6	24	20.34	24	21.11
7	29	27.67	29	28.94
8	40	37.81	39	39.05
9	44	51.07	46	51.53
10	69	67.19	69	66.18
11	85	85.18	85	82.42
12	94	103.53	94	99.39
13	133	120.84	116	116.23
14	141	136.44	141	132.46
15	148	150.75	148	148.23
16	157	165.24	157	164.45
17	182	182.08	182	182.61
18	209	203.31	209	204.21
19	229	229.38	229	229.67
20	257	257.81	257	257.26
21	282	284.04	282	283.47
22	306	304.50	306	305.04

our experiment, while Fréchet and DTW metrics can be used for series of different length.

The data scientist should consider the variability in the sequences of ML parameters due to non-linearity of artificial neural networks models when small changes in input data may modify internal parameters by a large offset.

6 Conclusion

The data quality has a strong impact on building models, validation and testing for performance. The most important dimensions of data quality are discussed along various metrics for similarity checking. Three case studies, from different fields, have been presented: text recognition influenced by quality of input material, electrocardiogram classification based on digital signal preprocessing before using deep learning, and timeline data analysis in order to measure the quality of data sources.

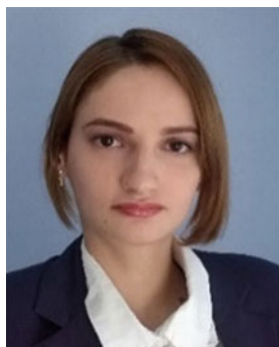
It was shown that aspects concerning the input quality and the performance of the models built according to machine learning strategies are specific to the application field.

References

1. Aghabozorgi S, Shirkhorshidi AS, Wah TY (2015) Time-series clustering—A decade review. *Inf Syst* 53:16–38 (Elsevier)
2. Almeida FA, Leite RR, Gomes GF, Gomes JHF, Paiva AP (2020) Multivariate data quality assessment based on rotated factor scores and confidence ellipsoids. *Decis Support Syst* 129. <https://doi.org/10.1016/j.dss.2019.113173> (Elsevier)
3. Arbitman G, Klein S, Peterlongo P, Shapira D (2021) Approximate hashing for bioinformatics. In: CIAA 2021—25th international conference on implementation and application of automata, Bremen, Germany, pp 1–12. hal-03219482
4. Belgacem N, Fournier R, Nait-Ali A, Bereksi-Reguig F (2015) A novel biometric authentication approach using ECG and EMG signals. *J Med Eng Technol* 39(4):226–238. <https://doi.org/10.3109/03091902.2015.1021429>, Taylor&Francis
5. Carvalho DV, Pereira EM, Cardoso JS (2019) Machine learning interpretability: a survey on methods and metrics. *Electronics* 8(8):832. <https://doi.org/10.3390/electronics8080832> (MDPI)
6. Choi W, Cho J, Lee S, Jung Y (2020) Fast constrained dynamic time warping for similarity measure of time series data. *IEEE Access* 8: 222841–222858. <https://doi.org/10.1109/ACCESS.2020.3043839> (IEEE)
7. Ciaburro G, Venkateswaran B (2017) Neural networks with R, Packt
8. Dai W, Yoshigoe K, Parsley W (2018) Improving data quality through deep learning and statistical models. In: Latifi S (ed) *Information technology—New generations. advances in intelligent systems and computing*, vol 558. Springer, https://doi.org/10.1007/978-3-319-54978-1_66
9. Dăneț N (2012) Some remarks on the Pompeiu-Hausdorff distance between order intervals. *ROMAI J* 8(2):51–60 (ROMAI)
10. Damerau - Levenshtein distance. https://en.wikipedia.org/wiki/Damerau%E2%80%93Levenshtein_distance. Accessed 12 June 2022
11. Fritsch S, Guenther F, Wright MN (2019) Package ‘neuralnet’, <https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>
12. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L (2019) Explaining explanations: an overview of interpretability of machine learning. <https://arxiv.org/pdf/1806.00069.pdf>
13. GridDB. <https://griddb.net/en/>. Accessed 12 June 2022
14. Gudivada VN, Apon A, Ding J (2017) Data quality considerations for big data and machine learning: going beyond data cleaning and transformations. *Int J Adv Softw* 10(1&2):1–20 (IARIA)
15. Heinrich M, Kaiser M, Klier M (2007) How to measure data quality? A metric based approach. In: *Twenty eighth international conference on information systems*, Montreal. <https://epub.uni-regensburg.de/23633/>. Accessed 12 June 2022
16. Hönigl J, Küng J (2015) Obtaining a data quality index with respect to case bases. *Vietnam J Comput Sci* 2:47–56. <https://doi.org/10.1007/s40595-014-0030-9>, Springer
17. Huffman CW, Kim J-L, Solé P (2021) *Concise encyclopedia of coding theory*. CRC Press
18. Jang W-J, Lee S-T, Kim J-B, Gim G-Y (2019) A study on data profiling: focusing on attribute value quality index. *Appl Sci* 9(23), Article 5054. <https://doi.org/10.3390/app9235054> (MDPI)
19. Jaro-Winkler distance. https://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance. Accessed 12 June 2022
20. Klein BD, Rossin DF (1999) Data quality in linear regression models: effect of errors in test data and errors in training data on predictive accuracy. *Inform Sci Inst* 2:33–43

21. Kljun M, Tersek M, Strumbelj E (2020) A review and comparison of time series similarity measures, ERK'2020, pp 367–370. [https://erk.fe.uni-lj.si/2020/papers/kljun\(a_review\).pdf](https://erk.fe.uni-lj.si/2020/papers/kljun(a_review).pdf)
22. Kullback–Leibler divergence. https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence. Accessed 12 June 2022
23. Lai K-H, Zha D, Wang G, Xu J, Zhao Y, Kumar D, Chen Y, Zumkhawaka P, Wan M, Martinez D, Hu X (2021) TODS: an automated time series outlier detection system. Proc AAAI Conf Artif Intell 35(18):16060–16062. <https://ojs.aaai.org/index.php/AAAI/article/view/18012>
24. Liu C, Li P, Zhao L, Liu F, Wang R (2011) Real-time signal quality assessment for ECGs collected using mobile phones. In: Computing in cardiology (IEEE Cinc2011, Hangzhou, China), vol 38, pp 357–360. <http://cinc.mit.edu/archives/2011/pdf/0357.pdf>. Accessed 12 June 2022
25. Mahalanobis distance. https://en.wikipedia.org/wiki/Mahalanobis_distance. Accessed 12 June 2022
26. Marcheggiani D, Sebastiani F (2017) On the effects of low-quality training data on information extraction from clinical reports. J Data Inf Qual 9(1). <https://doi.org/10.1145/3106235> (ACM)
27. Moloiu A-S (2014) Automatic character recognition (in Romanian). License thesis in informatics (under the supervision of Dana-Mihaela Vilcu), “Spiru Haret” University, Bucharest
28. Moloiu A-S (2017) ECG authentication system development (in Romanian). Master thesis in informatics (under the supervision of Dana-Mihaela Vilcu), “Spiru Haret” University, Bucharest
29. Moloiu A-S, Albeanu G, Popentiu-Vladicescu F (2020) Recent computational intelligence developments: technical and social aspects. Proc eLSE2020 1:429–436 (Editura Universitara)
30. Moloiu A-S, Albeanu G, Madsen H, Popentiu-Vladicescu F (2021) Data quality assessment for ml based decision-making systems. In: Proceedings of 26th ISSAT international conference on reliability and quality in design, RQD 2021, pp 39–43 (ISSAT)
31. Montero P, Vilat JA (2014) TSclust: an r package for time series clustering. J Stat Softw 62(1):1–43 (Foundation for Open Access Statistics)
32. Mori U, Mendiburu A, Lozano JA (2016) Distance measures for time series in R: the TSdlist package. The R J 8(2):451–459 (The R Foundation)
33. Paparrizos J, Gravano L (2017) Fast and accurate time-series clustering. ACM Trans Datab Syst 42(2):49. <https://doi.org/10.1145/3044711> (ACM)
34. Phionet: improving the quality of ECGs collected using mobile phones. <https://archive.physionet.org/challenge/2011/>. Accessed 12 June 2022
35. Pinto JR, Cardoso JS, Lourenco A (2019) Deep neural networks for biometric identification based on non-intrusive ECG acquisition. In: Arya KV, Bhadoria RS (eds) The biometric computing—Recognition and registration. Chapman and Hall/CRC, pp 217–234
36. Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. Commun ACM (ACM) 45(4):211–218. <https://doi.org/10.1145/505248.506010>
37. Sakoe H, Chiba S (1978) Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 26(1):43–49 (IEEE)
38. Sarker IH (2021) Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2:420 (Springer). <https://doi.org/10.1007/s42979-021-00815-1>
39. Scandura TA, Gower K (2009) Management today: best practices for the modern workplace. SAGE Publishing
40. Sessions V, Valtora M (2006) The effects of data quality on machine learning algorithms. In: Talburt JR, Pierce EM, Wu N, Campbell T (eds) Proceedings of the 11th international conference on information quality, MIT, Cambridge, MA, USA, pp 485–498
41. Shome N, Laskar RH, Das D (2019) Reference free speech quality estimation for diverse data condition. Int J Speech Technol 22:585–599. <https://doi.org/10.1007/s10772-018-9537-2> (Springer)
42. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Rob Fergus R (2014) Intriguing properties of neural networks. <https://arxiv.org/abs/1312.6199v4>

43. Tang M, Yu Y, Aref WG, Malluhi QM, Ouzzani M (2015) Efficient processing of hamming-distance-based similarity-search queries over mapreduce. In: Proceedings of the 18th international conference on extending database technology. EDBT, pp 361–372
44. Timmerman Y, Bronselaer A (2019) Measuring data quality in information systems research. *Decis Support Syst* 126, Article 113138. <https://doi.org/10.1016/j.dss.2019.113138> (Elsevier)
45. Wang F, Li M, Mei Y, Li W (2020) Time series data mining: a case study with big data analytics approach. *IEEE Access* 8:14322–14328. <https://doi.org/10.1109/ACCESS.2020.2966553>,IEEE
46. Watson Studio, IBM. <https://www.ibm.com/cloud/watson-studio>. Accessed 12 June 2022
47. Z-score. https://en.wikipedia.org/wiki/Standard_score. Accessed 12 June 2022
48. Zhang L, Jeong D, Lee S (2021) Data quality management in the internet of things. *Sensors* 21:5834. <https://doi.org/10.3390/s21175834> (MDPI)



Alexandra-Ștefania Moloiu TypingDNA, Bucharest, Romania

Alexandra-Ștefania Moloiu received her M.S. degree in Computer Science in 2017 from “Spiru Haret” University, Romania, and is currently pursuing the Ph.D. degree research program. She worked on numerous computer vision and deep learning projects mostly in the areas of self-driving cars and biometrics. She is a Senior Data Scientist at TypingDNA-Company (<https://www.typingdna.com/>). Her current interests include deep learning, biometrics, and face analysis.



Grigore Albeanu “Spiru Haret” University, Bucharest, Romania

Dr. Grigore Albeanu received his Ph.D. in Mathematics in 1996 from the University of Bucharest, Romania. During the interval 2004–2007, he was the Head of the UNESCO IT Chair at the University of Oradea and participated as principal investigator in a NATO project. Grigore Albeanu has authored or co-authored over 130 papers and 10 educational textbooks in applied mathematics and computer science. Since 2007, he is a Professor of computer science at “Spiru Haret” University in Bucharest. His current research interests include different aspects of scientific computing, including soft computing, modeling and simulation, software reliability, virtual reality techniques, and E-Learning.



Henrik Madsen Danish Technical University, Lyngby, Denmark

Dr. Henrik Madsen received a Ph.D. in Statistics at the Technical University of Denmark in 1986. He was appointed as Ass. Prof. in Statistics (1986), Assoc. Prof. (1989), and Professor in Mathematical Statistics with a special focus on Stochastic Dynamical Systems in 1999. In 2017, he was appointed as Professor II at NTNU in Trondheim. His main research interest is related to the analysis and modeling of stochastic dynamic systems. This includes signal processing, time series analysis, identification, estimation, grey-box modeling, prediction, optimization, and control. The applications are mostly related to Energy Systems, Informatics, Environmental Systems, Bioinformatics, Biostatistics, Process Modeling, and Finance. He has got several awards. Lately, in June 2016, he has been appointed as Knight of the Order of Dannebrog by Her Majesty the Queen of Denmark, and he was appointed as Doctor HC at Lund University in June 2017. He has authored or co-authored approximately 500 papers and 12 books. The most recent books are *Time Series Analysis* (2008), *General and Generalized Linear Models* (2011), *Integrating Renewables in Electricity Markets* (2013), and *Statistics for Finance* (2015).



Florin Popențiu-Vlădicescu University “Politehnica” of Bucharest & Academy of Romanian Scientists, Romania

Dr. Florin Popențiu-Vlădicescu was born on 17 September 1950, graduated in Electronics and Telecommunications from the University POLITEHNICA of Bucharest in 1974, and holds a Ph.D. in Reliability since 1981. Also, he is an Associate Professor at the University “Politehnica” of Bucharest, Faculty of Automatic Control and Computer Science. Professor Florin Popențiu-Vlădicescu is the founder of the first “UNESCO Chair of Information Engineering”, in UK, established at City University London, in 1998. He published over 150 papers in international journals and conference proceedings. Also, he is the author of one book and co-author of four books. Professor Florin Popențiu-Vlădicescu has worked for many years on problems associated with software reliability and has been Co-Director of two NATO Research Projects, in collaboration with Telecom ParisTech and Danish Technical University (DTU). He is an independent expert in the European Commission—H2020 programme, for Net Services—Software and Services, Cloud. Professor Popențiu-Vlădicescu is currently a Visiting Professor at renowned European and South Asian universities. He was the elected Fellow of the Academy of Romanian Scientists in 2008.

From Holistic Health to Holistic Reliability—Toward an Integration of Classical Reliability with Modern Big-Data Based Health Monitoring



Fengbin Sun

Abstract This paper aims to explore a new philosophical framework for integrating the classical reliability with modern big-data based health monitoring. The concepts of holistic health, holistic failure, and holistic reliability are introduced. Numerous algorithms of quantifying the holistic failure and holistic reliability are proposed.

Keywords Holistic health · Holistic failure · Holistic reliability · Degradation · Virtual failure · Complex number analogy · Yin-Yang analogy

1 Introduction

In classical reliability engineering, the status of each unit under test or in use is classified as either failure or survival. Life data analysis deals with two distinctive time categories: failure time and survival time. Here the failures are typically manifested as a “physical” loss of function or wrong output outside of the desired target range. In the meantime, however, the sub-healthy condition can be encountered at various scenarios, such as when performance degradation has crossed the pre-specified threshold but hasn’t manifested as a macro failure (cease to function physically) yet, etc. Such failures are typically classified as degradation-based virtual (or so-called pseudo) failures and can be quantified as fractional failures and fractional survivors [1]. In real world, both true failures and degradation-based virtual failures can co-exist within the same product. Scientifically integrating these two categories of failures provides holistic view of product reliability.

With the advancement of modern failure physics, sensor technology, and big data, the micro-level virtual failures become more and more “visible” and dominating. In addressing the opportunities and challenges when Reliability meets Big Data, Meeker and Hong [2] listed numerous examples of engineering systems that can generate large-scale product operating and environment data (i.e., high-frequency multivariate

F. Sun (✉)
Reliability Engineering, Tesla Inc, Palo Alto, CA, USA
e-mail: franksun9999@gmail.com

time series data), which can be used in reliability analysis. Such systems include, but are not limited to [2, 3]:

- Locomotive engines
- Aircraft engines
- Automobiles
- Power distribution transformers
- CT scanners and other large medical systems
- Wind turbines
- Solar energy power inverters and stationary energy storage systems
- Data storage device (HDD/SSD) and systems (data center)
- Gas turbines, large construction equipment, high-end printers/copiers, high-end computers, some home entertainment systems, smart phones, etc.

Over the past 36 years, the use of degradation data in reliability testing has received much attention. In addition to providing additional information about reliability and the opportunity to make reliability inferences with few or even no failures, degradation data provides information that is much richer for building and assessing the adequacy of physical/chemical models used for test acceleration. The typical use and application of modern reliability data includes [2–9]:

- Sensor data logging, retrieval, and time-series analysis
- Machine learning and failure prediction
- System prognostics, health management, and condition-based maintenance
- Simulation (physics based or Monte Carlo) and problem identification
- Early warning of emerging reliability issues
- Prediction of remaining life of individual systems
- Prediction of retirements/replacements in a fleet of systems
- Field data (ship/return) and warranty reserve forecast
- Prediction of maintenance costs
-

Introducing parametric-degradation-based virtual failure concept adds a micro-dimension to the traditional failure domain. Incorporating virtual failures into reliability assessment provides more insights about underlying product reliability and enhances the detection sensitivity of various reliability monitoring mechanisms, such as ongoing reliability test (ORT) during production and therefore can surface the poor vintage with potential high future failure rate due to “invisible” high degradation of critical parameters [4].

It is author’s belief that now it is time to develop a holistic reliability assessment framework that integrates the “macro-level” physical failure based classical reliability assessment technique with “micro-level” virtual failure based health monitoring technique. However, to author’s knowledge, up to this point, there is no publication on how to combine the physical failures with virtual failures to yield a holistic failure domain. To stimulate discussion and participation from both academia and fellow industry practitioners, author of this paper proposed numerous potential philosophies of describing the holistic failure and algorithms of quantifying

the holistic failure counts, including complex number (vector) approach, Yin-Yang theory, etc. An example is given to illustrate the potential value based on some preliminary work by the author. It is hoped that this work will be beneficial to a wide range of audience including reliability practitioners, theorists, and management.

The rest of the paper is arranged as follows. Section 2 reviews the manifestation and underlying mechanisms of physical and virtual failures. Section 3 gives a practical example of virtual failure determination from HDD industry with detailed step-by-step procedures. Section 4 looks into the health philosophy from human-machine analogy perspective and the need for holistic reliability assessment. Section 5 presents numerous philosophical algorithms for holistic integration of physical failure with virtual failure. Section 6 provides corresponding mathematical modeling for holistic failure quantification. Sect. 7 provides a numerical example for holistic reliability analysis considering parametric-degradation based virtual failures. Section 8 concludes the paper with future works.

2 Manifestation and Underlying Mechanisms of Physical and Virtual Failures

Product can fail and lose its function in various ways, such as:

- Fatigue
- Crack
- Corrosion
- Wear
- Electromigration
- Dielectric breakdown
- Electro-chemical migration (dendrite growth)
- Open circuit due to EOS or ESD.
-

Performance degradation induced virtual failures can also manifest in various ways depending on their failure governing nature. In general, there are two types of performance degradation; i.e., monotonic and fluctuating degradation [4].

For monotonic category, the product performance parameter changes monotonically; either monotonically increasing or monotonically decreasing as shown in Fig. 1. Their degradation is quantified by the max% change as follows where CP stands for critical parameter and ABS() stands for the absolute value of the parenthesis:

$$\text{Max} \left[\frac{\text{ABS}(\text{CP}_{\text{Current}} - \text{CP}_{\text{Initial}})}{\text{CP}_{\text{Initial}}} \right] \times 100\% \tag{1}$$

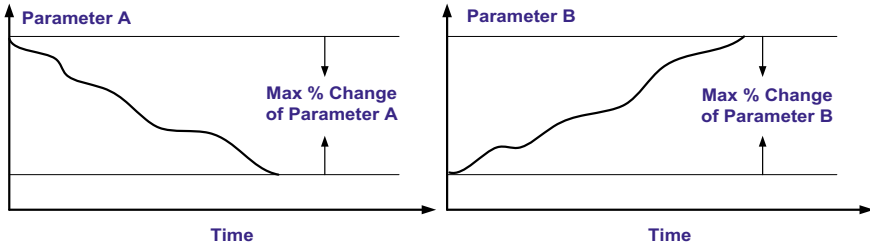
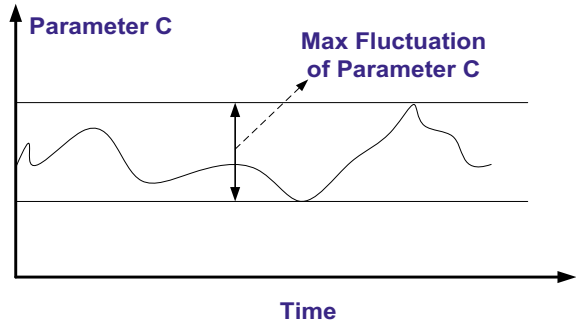


Fig. 1 Illustration of measuring the maximum degradation of monotonically degrading parameters [4]

Fig. 2 Illustration of measuring the maximum degradation of a fluctuating degradation parameter [4]



For non-monotonic but fluctuating category, the critical parameter goes up and down and reveals an unstable behavior as shown in Fig. 2. The performance degradation is quantified by the maximum fluctuation of the critical parameter as [4]:

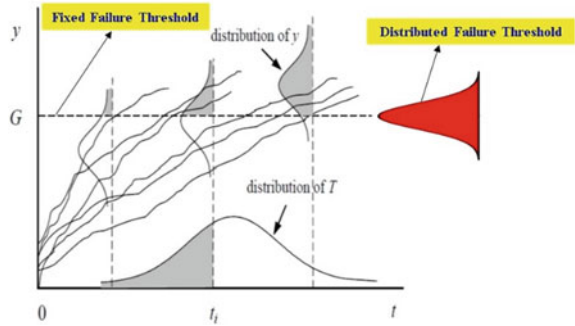
$$\text{Max CP Fluctuation} \sim (CP_{\text{Max}} - CP_{\text{Min}}) \tag{2}$$

In degradation analysis, it is often assumed that when the performance degradation reaches a pre-specified fixed threshold, the (virtual) failure happens. However, this is a conceptual assumption, not always true. In reality, the degradation to failure is not fixed, but a random variable itself as shown in Fig. 3. Reaching threshold only implies a certain failure probability, not necessarily a true failure.

For example, if the degradation-to-failure distribution has a pdf of $g(x)$, then the likelihood of failure when the performance degradation reaches a specified threshold, G , is given by

$$L_F = \int_{x \leq G} g(x) dx \tag{3}$$

Fig. 3 Reaching a pre-specified degradation threshold doesn't always imply a true failure



Empirically, Eq. (3) can be estimated by a reliability life test with performance degradation monitoring as follows:

$$L_F \approx \frac{\text{(Total \# of samples whose degradation exceeds specified trigger limit and actually failed)}}{\text{(Total \# of samples whose degradation exceeds specified trigger limit)}} \tag{4}$$

If we define a unit whose critical parameter exceeds the pre-specified threshold as a potential failure, then multiplying the above failure likelihood with the total number of potential failures will yield the total number of virtual failures; i.e.,

$$\text{Virtual Failure} = \text{Potential Failure} \times L_F \tag{5}$$

Since the failure likelihood is between 0 and 1, therefore, the virtual failure is often a fractional number.

3 An HDD Example of Virtual Failure Determination

3.1 Critical Parameters Identification, Their Degradation Pattern and Quantification

For 66 years since the invention of the first product in 1956 by IBM, the hard disk drive (HDD) industry has progressed tremendously not only in technology to meet the ever increasing demand of higher data storage capacity, better performance, and lower cost, but also higher product quality and reliability. For example, there may be more than one hundred critical parameters at both the magnetic head component level and hard disk drive system level that can be tracked and collected during HDD reliability tests [4]. They are expected to represent the product parametric performance

from various perspectives. The following is an exemplary list of some parameters as collected in reliability demonstration test (RDT) and ongoing reliability test (ORT):

- G-List: “growth” defect table for sectors that have gone bad after the drive was placed in use
- Error Rate (ER): raw bit error rate
- MR Bias: bias current applied on magneto resistive head
- Magneto-Resistive Asymmetry (MRA): channel amplitude asymmetry compensation
- MR Resistance: reader element resistance of magneto resistive head
- Non-Repeatable Run-Outs (NRRO): measurement of how much a platter wobbles, or moves off-center
- Spin Up Time: time required for the disk platters to get up to full operational speed from a stationary start
- Variable Gain Amplifier (VGA): channel amplitude gain compensation
- Write Current: current going through the writer element during write operation
-

Based on the extensive studies of available historical parametric data from ORT failures using degradation analysis and from field failure returns using multivariate analysis, four critical parameters (CP), three at head level and one at drive level, were identified as the most failure-indicative candidates [4]:

- Parameter A—Head Level
- Parameter B—Head Level
- Parameter C—Head Level
- Parameter D—Drive Level.

Figure 4 is a sample line plot of HDD CP degradation during ORT. It may be seen that the head-level parameters (Parameters A, B, and C) behave diversely from head to head, and drive-level parameter (Parameter D) behaves diversely from HDD to HDD. However, the significant change in the form of increase, decrease, and fluctuation are indicators for performance degradation while flatness is an indicator of performance robustness.

The failure of an HDD is governed by its weakest link. The head-level reliability performance of a HDD, as a whole, is governed by the worst-head (if one head fails, the whole HDD fails). Therefore, the CP values of the worst head should be used to represent the head-level parametric performance of an HDD. In general, Parameters A and B are monotonic [4]. Their degradation is quantified by the max % change of the worst head by Eq. (1), or

$$\text{Max} \left[\text{ABS}(\text{CP current day} - \text{CP 1st day}) / \text{CP 1st day} \right] \times 100\% \quad (6)$$

where ABS() stands for the absolute value of the parenthesis.

Parameter C is not monotonic but fluctuating in nature. Its degradation is quantified by the maximum fluctuation [4] of the worst head by Eq. (2), or

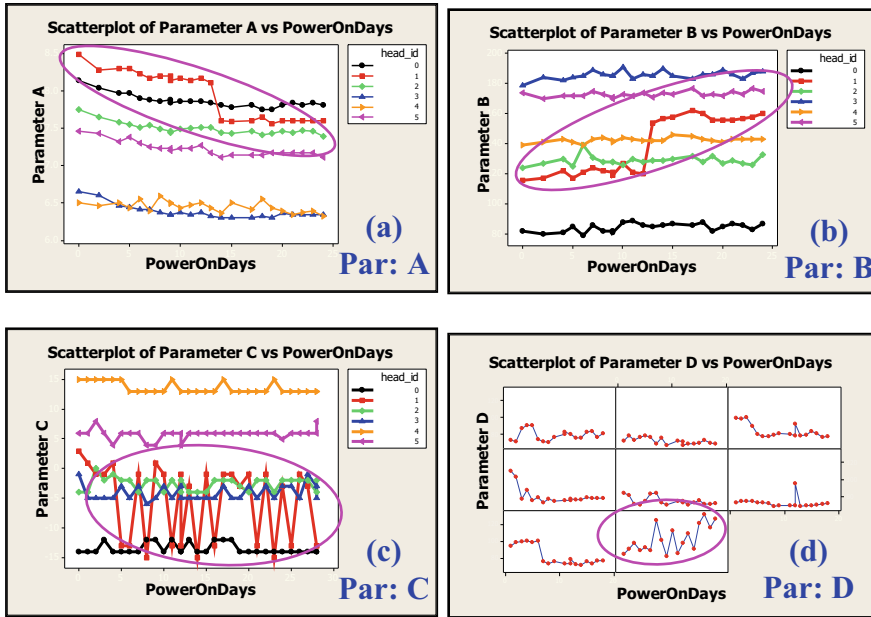


Fig. 4 A sample line plot of critical parameters A, B, C, and D for HDDs during ORT

$$\text{Max CP Fluctuation} = (\text{CP}_{\text{Max}} - \text{CP}_{\text{Min}})/32 \tag{7}$$

where

CP_{Max} the maximum DAC (digital-to-analog conversion) value of Parameter C for a head

CP_{Min} the minimum DAC value of Parameter C for a head
 32 (+16) – (–16) is the maximum allowable fluctuation range for Parameter C.

Note that the DAC (digital-to-analog converter) value is used directly here as Parameter C is linearly proportional to its DAC value.

Parameter D is also not monotonic but fluctuating. Its degradation is quantified by the maximum fluctuation [4] by Eq. (2), or

$$\text{Max CP Fluctuation} = (\text{CP}_{\text{Max}} - \text{CP}_{\text{Min}})/255 \tag{8}$$

where

CP_{Max} the maximum DAC value of Parameter D for an HDD
 CP_{Min} the minimum DAC value of Parameter D for an HDD
 255 ($2^8 - 1$) – 0 is the maximum allowable fluctuation range for Parameter D.

Note that similar to Parameter C, the DAC value is used directly here as Parameter D is linearly proportional to its DAC value.

3.2 Parametric Trigger Limits Determination

The trigger limits for the critical parameters are determined based on the actual CP data. The Individual-Moving Range (I-MR) charts are used to examine the process stability and upper and lower control limits (UCL and LCL) for the maximum % change or maximum fluctuation of each parameter [4]. The I-chart (individual chart) represents the process mean while the MR-chart (moving range chart) represents the process variance for each parameter. The UCL and LCL of the I-chart correspond to the $(\text{Mean} + 3\text{Sigma})$ and $(\text{Mean} - 3\text{Sigma})$ of the maximum % change or maximum fluctuation of each parameter and are used as the parametric trigger limits. It is required that both I-chart and MR-chart be stable for these limits to be valid.

The trigger limits determined above will be used to track and monitor the statistical behavior of each corresponding parameter for HDDs by vintage (work week) and identify the outliers as potential failures that will be included in the virtual failure analysis.

3.2.1 Population Dynamic Behavior Monitoring Via Box-Whisker Plotting

In order to track and monitor product overall parametric performance by vintage (week by week), Box-Whisker plots are used to describe the distribution of HDD population manufactured in that particular week [4]. They are also used to spot outliers. Figures 5 and 6 are sample Box plots by work week for Parameters A, B, C, and D from parametric ORT. The Y-axis represents the maximum % change or the maximum fluctuation of the plotted parameter for each HDD.

Applying the trigger limits on the Box-Whisker plot for each parameter yields “Potential Failure” candidates, the outliers outside the trigger limits. A potential failure candidate may have been triggered by a single or multiple parameters. These potential failures will be further studied individually in the next section by looking into their time trend so that those with the worst degradation will be selected for virtual failure analysis (“Virtual FA”). In the meantime, a weighting algorithm will

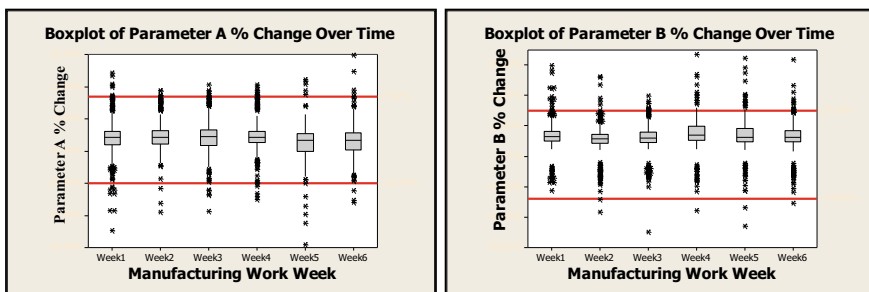


Fig. 5 A sample box plot by work week for parameters A and B

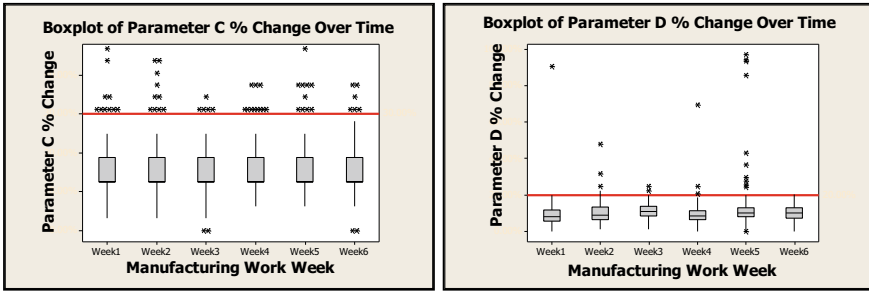


Fig. 6 A sample box plot by work week for parameters C and D

be applied to convert these potential failures to “Virtual Failures” for long-term reliability impact assessment.

3.2.2 Individual Behavior Examination for the Potential Failure Candidates Via Scatter (Line) Plotting

In order to get a deeper insight about the parametric degradation for each potential failure candidate identified by Box-Whisker plots against the trigger limits, the time trend for each parameter needs to be examined for each potential failure candidate individually. Scatter plots or line plots can be used to accomplish this task. Figures 7 and 8 are sample scatter plots versus test time of Parameters A, B, C, and D for each potential failure candidate. The Y-axis for the line plots represents the actual measurement of the parameter itself, not the maximum % change or the maximum fluctuation.

The graphical pattern of the line plot for each individual potential failure helps to identify potential failures for further virtual FA.

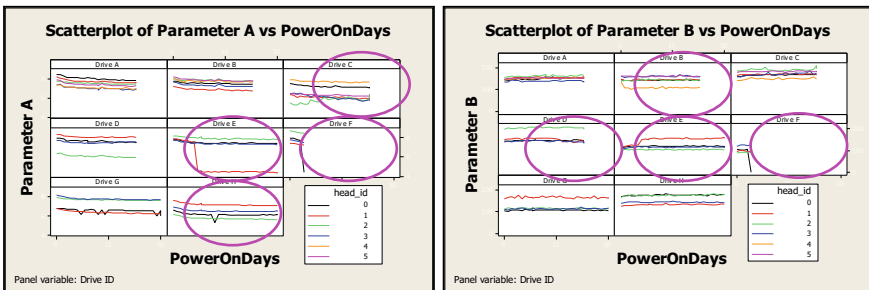


Fig. 7 A sample scatter plot versus test time of parameters A and B for the potential failures identified by box-whisker plots with trigger limits applied

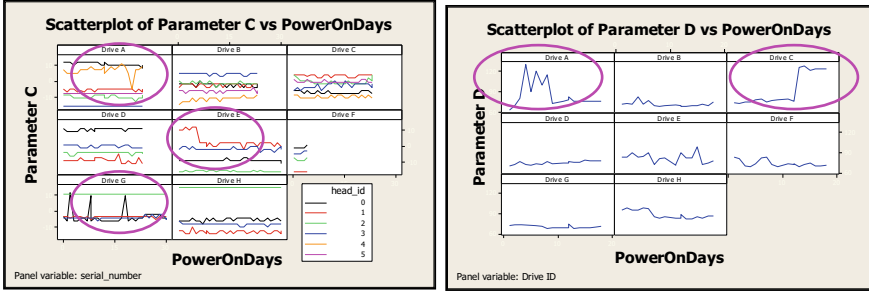


Fig. 8 A sample scatter plot versus test time of parameters C and D for the potential failures identified by box-whisker plots with trigger limits applied

3.2.3 Converting Potential Failures to Virtual Failures

Traditionally, product reliability during ORT, such as annual failure rate (AFR), is assessed based on the true failures. Since a deteriorating HDD may contribute to near-future true failures even though it may not manifest itself as a failure in ORT due to the test time limitation, the concept of “Virtual Failure” is introduced here from HDD parametric health point of view in order to quantify HDD reliability holistically. As stated previously, based on the extensive studies, Parameters A, B, C, and D have been identified as the most failure-indicative critical parameters. When the maximum % change or the maximum fluctuation of any of these parameters exceeds its trigger limit, this HDD will be identified as a potential failure candidate, an HDD with potential to fail in the next moment should the HDD is allowed to continue ORT. To account for the fact that not every potential failure candidate will fail in ORT but there is a correlation or likelihood between the incident of each CP exceeding the trigger limit and the incident of its actual failing in ORT, a weighted linear function is introduced to convert potential failures to their equivalent virtual failures, that is:

$$N_{VF} = (L_{F-A} \times N_{PF-A} + L_{F-B} \times N_{PF-B} + L_{F-C} \times N_{PF-C} + L_{F-D} \times N_{PF-D}) \tag{9}$$

where

- L_F failure likelihood, the failure-equivalence factor of converting the “Potential Failures” to “Virtual Failures” due to their CP % change exceeding the trigger limits. It is defined by Eq. (4) as the conditional probability that HDD will fail in ORT given that it’s CP maximum % change or maximum fluctuation during ORT exceeds the trigger limit; i.e.,
- L_F [Total # of ORT HDDs whose maximum % changes exceed the specified trigger limits and fail in ORT]/[Total # of ORT HDDs whose maximum % changes in ORT exceed the specified trigger limits]

- N_{PF-A} potential failure number induced by Parameter A, the total number of passing HDDs with their maximum Parameter A % change exceeding the specified trigger limits
- N_{PF-B} potential failure number induced by Parameter B, the total number of passing HDDs with their maximum Parameter B % change exceeding the specified trigger limits
- N_{PF-C} potential failure number induced by Parameter C, the total number of passing HDDs with their maximum Parameter C fluctuation exceeding the specified trigger limits
- N_{PF-D} potential failure number induced by Parameter D, the total number of passing HDDs with their maximum Parameter D fluctuation exceeding the specified trigger limits.

Note that the HDD serial numbers used in Eq. (9) should be all distinctive among the four brackets. If one HDD has more than one CP's whose maximum % changes exceed the specified trigger limits, the one with highest L_F value should be used in Eq. (9). A numerical example given in Sect. 6 shows a set of sample ORT test tracking data with both true failure count each week as well as calculated virtual failure count converted from corresponding potential failures due to parametrical degradation exceeding threshold.

4 Health Philosophy and the Need for Holistic Reliability Assessment: Human-Machine Analogy

Human health and machine health bear many similarities. People see their doctors periodically for health check up and get their physical health indication parameters, such as:

- Body temperature
- Blood pressure
- Pulse rate, rhythm, and quality
- Heartbeat, or heart sound
- Respiratory rate, effort, and quality
- Cholesterol level
- Blood sugar level
-

Comparing with these physical health indicators, which are typically quantitative, the mental health indicators are more subjective and therefore more qualitative, such as:

- Self confidence
- Optimism
- Respect for self and others

- Empathy
- Resiliency to stress
- Flexibility
- Spirituality
- Sense of belonging
-

In human health science, it is believed that the mental health affects physical health and is not separable from, and even more important than physical health. In the meantime, both physical health and mental health are affected by environmental and operational conditions of human body, such as weather conditions (temperature, humidity, sunshine, rain, etc.), family environment, working environment, workload and stress, etc.

In analogy to human health, the “physical” loss of function for a hardware or wrong output outside of the desired target range corresponds to the physical disability or misbehavior of human being, while the parametric degradation seems more similar to human aging and mental degradation. The physical health is often more visible and quantifiable than mental health. Of course, it is well known that product health is directly affected by both environmental and operational conditions, such as temperature, humidity, altitude, vibration, workload (stress), etc.

The modern health science promotes holistic view of health and its management; i.e., treat physical and mental health as a whole. This reminds us we should take a similar approach to machine health; i.e., treat the physical failure and virtual failure as a whole. For example, when one compares two groups of product samples under the same test conditions, we not only compare their actual “visible” physical failures, but also compare their “invisible” performance (or characteristic) degradation.

It is understood that some physical failures may happen to be physical manifestation of degradation failure. However, it should be pointed out that not all physical failures are due to degradation, especially failures due to infant mortality, manufacturing defects, random overstress such as ESD, EMI, etc. In other words, the physical failures and degradation process maybe dependent, and maybe independent as well. Many believe that the so-called hard (or catastrophic) failures and so-called soft failures are always competing with each other [5–24].

5 Holistic Integration of Physical Failure with Virtual Failure

As discussed in the preceding section, physical failures and virtual failures are two parallel and competing aspects of product failure. Even though one is “visible” and the other is “not visible”, both of them should be considered in the holistic representation of product failure. Such relationship can be most vividly mimicked by the following two theories:

- (1) Complex number theory
- (2) Ying Yang theory.

5.1 Complex Number (or Vector) Analogy

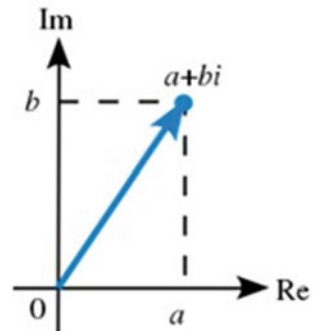
In Complex Number theory, a complex number is expressed in the form of $a + bi$, where a and b are real numbers, and i is a solution of the equation $x^2 = -1$. For the complex number $a + bi$, a is called the real part, and b is called the imaginary part. Despite the historical nomenclature “imaginary”, complex numbers are regarded in the mathematical sciences as just as “real” as the real numbers, and are fundamental in many aspects of the scientific description of the natural world [25].

Graphically, a complex number can be visually represented as a pair of numbers (a, b) forming a vector on a diagram called an Argand diagram, representing the complex plane as shown in Fig. 9 where “Re” is the real axis, “Im” is the imaginary axis. Geometrically, complex numbers extend the concept of the one-dimensional number line to the two-dimensional complex plane by using the horizontal axis for the real part and the vertical axis for the imaginary part. The complex number $a + bi$ can be identified with the point (a, b) in the complex plane. A complex number whose real part is zero is said to be purely imaginary; the points for these numbers lie on the vertical axis of the complex plane. A complex number whose imaginary part is zero can be viewed as a real number; its point lies on the horizontal axis of the complex plane.

In analogy, the holistic failure can be represented by “real failure” (physical failure) part and “imaginary failure” (virtual failure due to performance degradation) part:

$$\text{HolisticFailure} = (\text{PhysicalFailure}) + (\text{VirtualFailure})i \tag{10}$$

Fig. 9 Graphical representation of holistic failure analogy to a complex number



Therefore, the classical reliability theory deals with a special case when there is no performance degradation induced “virtual” failures, but “hard” failures only. In contrary, health monitoring based on pure parametric degradation, excluding hard failures, is another special case where only imaginary (or virtual) failure is considered.

5.2 *Ying-Yang Analogy*

In Chinese philosophy, Yin and Yang (“dark-bright”, “negative–positive”) describes how seemingly opposite or contrary forces may actually be complementary, interconnected, and interdependent in the natural world, and how they may give rise to each other as they interrelate to one another. Many tangible dualities (such as light and dark, fire and water, expanding and contracting) are thought of as physical manifestations of the duality symbolized by Yin and Yang. This duality lies at the origins of many branches of classical Chinese science and philosophy, as well as being a primary guideline of traditional Chinese medicine [26]. Figure 10 is a typical graphical illustration of Yin-Yang theory.

Yin and Yang can be thought of as complementary (rather than opposing) forces that interact to form a dynamic system in which the whole is greater than the assembled parts. According to this philosophy, everything has both yin and yang aspects (for instance, shadow cannot exist without light). Either of the two major aspects may manifest more strongly in a particular object, depending on the criterion of the observation. The Yin Yang shows a balance between two opposites with a portion of the opposite element in each section.

In analogy, the physical failure corresponds to the “Yang” element of the failure that is visible, while the virtual failure due to performance degradation corresponds to the “Yin” element of the failure that is “not visible”:

Fig. 10 Graphical representation of holistic failure analogy to Yin-Yang theory



$$\text{HolisticFailure} = (\text{PhysicalFailure})_{\text{Yang}} \cup (\text{VirtualFailure})_{\text{Yin}} \quad (11)$$

According to Yin-Yang theory, physical failure and virtual failure are two complementary elements of overall holistic failure. They can be interrelated and give rise to each other. For example, significant performance degradation may eventually manifest as a physical failure, while physical damage (such as that due to shock, vibration, UV exposure, etc.) may also accelerate or decelerate internal performance degradation. For instance, too much sun exposure leads to sunburn on human body, accelerate aging of the skin, and increase the risk of skin cancers. However, sunlight exposure can boost the body’s vitamin D supply, promotes bone health and decelerates bone degradation, etc.

6 Holistic Failure Quantification and Applications

From Sect. 5, it is understood that physical failure and virtual failure can be interpreted as two dimensions of holistic failure. However, in order to quantify the overall magnitude of the overall failure (or unhealthy conditions), we need to come up with something holistic.

This can be accomplished through the Complex Number analogy; i.e., the holistic failure can be represented by the following two attributes:

- (1) The absolute value (or modulus or magnitude), Z

$$Z = \sqrt{(\text{Physical Failure})^2 + (\text{Virtual Failure})^2} \quad (12)$$

- (2) The phase angle (or argument), θ

$$\theta = \arctan[(\text{Virtual Failure})/(\text{Physical Failure})] \quad (13)$$

The absolute value indicates the overall (combined) effects of physical failures and virtual failures while the phase angle indicates the relative contribution ratio between physical and virtual failures.

It is interesting to notice that the complex number field is intrinsic to the mathematical formulations of quantum mechanics, where complex Hilbert spaces provide the context for one such formulation that is convenient and perhaps most standard. The original foundation formulas of quantum mechanics—the Schrödinger equation and Heisenberg’s matrix mechanics—make use of complex numbers as well.

For multiple degradation parameter scenario, where there are more than one failure governing degradation parameters, Eq. (12) could potentially be extended to a more generalized format, such as the following:

$$Z = \left[(\text{Physical Failure})^m + \sum_i (\text{Virtual Failure}_i)^m \right]^{1/m} \quad (14)$$

where $m \geq 1$ represents the total number of distinctive failure dimensions due to both physical and virtual failure categories.

As an extreme case with $m = 1$, the holistic failure can be calculated as follows:

$$Z = (\text{Physical Failure}) + (\text{Virtual Failure}) \quad (15)$$

This implies that when both physical and virtual failures are treated as the same dimension, then they can be simply added together.

7 Holistic Reliability Analysis Considering Parametric-Degradation Based Virtual Failures—An Example

Traditionally, reliability metric during ongoing reliability test (ORT) of a consumer electronic device, such as annual failure rate (AFR), is assessed based on the true failures. Since a deteriorating device may contribute to near-future true failures even though it may not manifest itself as a failure in ORT due to the test time limitation, the concept of “Virtual Failure” is introduced from product parametric health point of view in order to quantify its reliability holistically [4].

Table 1 is an example of ORT data summary within 20 production weeks including:

- Weekly sample size
- Test duration
- Confirmed physical failures
- Parametric degradation induced virtual failures.

It is known that the underlying time-to-failure (TTF) distribution is Weibull with a known shape parameter of $\beta = 0.74$. It is desired to predict, by weekly vintage, field annual failure percent (AFR) at 60% confidence level under annual power on hours (POH) assumption of 3,120 h. Based on internal DOE study, the overall acceleration factor due to applied temperature and workload in ORT with respect field is about $AF = 13$.

Under Weibull life distribution assumption with known shape parameter, the field predicted AFR can be calculated as follows

$$\widehat{AFR} = 1 - e^{-\left(\frac{POH}{\hat{\eta}}\right)^\beta} \quad (16)$$

where

$$(\hat{\eta})^\beta = \frac{2\sum_{i=1}^n N_i (T_i \times AF)^\beta}{\chi_{CL;2}^2(\sum_{i=1}^n r_i) + 2} \quad (17)$$

Table 1 ORT data summary within 20 production weeks

Work week	Sample size	Run time, (h)	True failures	Virtual failure
WW01	100	672	0	0
WW02	100	672	0	0
WW03	100	672	0	0
WW04	100	672	1	3.32
WW05	100	672	2	1.07
WW06	100	672	0	1.22
WW07	100	672	1	1.31
WW08	100	672	0	0.21
WW09	100	672	2	0.22
WW10	100	672	1	1.10
WW11	100	672	0	0.57
WW12	100	672	2	0.99
WW13	100	672	2	6.04
WW14	100	672	1	4.09
WW15	100	672	2	0.57
WW16	100	672	1	0.71
WW17	100	672	1	0.99
WW18	100	672	0	1.04
WW19	100	672	0	0.85
WW20	100	672	1	0.80

- n total number of work weeks = 20,
- N_i sample size of week $i = 100$,
- T_i test duration of week $i = 672$ h,
- r_i failure number of week i ,
- β Weibull shape parameter = 0.74,
- AF ORT acceleration factor with respect to field = 13,
- CL confidence level = 60%.

Table 2 summarizes the holistic failure counts calculated using Eqs. (12) and (15). Table 3 summarizes the predicted field AFR using Eqs. (16) and (17) based on true failure, holistic failure #1, and holistic failure #2.

Figure 11 is graphical comparison of predicted field AFR using Eqs. (16) and (17) based on true failure, holistic failure #1, and holistic failure #2. It may be seen that AFR based on holistic failure provides added sensitivity than that based on true failure only. Holistic failure #1 as a direct sum of true and virtue failure treats true and virtual failures as one dimension, which is an extreme and most aggressive case. Holistic failure #2 as a vector modulus or magnitude of true and virtue failures is more realistic in reflecting their joint effects.

Table 2 ORT holistic failure by production week

Work week	Holistic failure #1: (True + Virtual) failures	Holistic failure #2: $\sqrt{\text{True}^2 + \text{Virtue}^2}$
WW01	0.00	0.00
WW02	0.00	0.00
WW03	0.00	0.00
WW04	4.32	3.47
WW05	3.07	2.27
WW06	1.22	1.22
WW07	2.31	1.65
WW08	0.21	0.21
WW09	2.22	2.01
W10	2.10	1.49
WW11	0.57	0.57
WW12	2.99	2.23
WW13	8.04	6.36
WW14	5.09	4.21
WW15	2.57	2.08
WW16	1.71	1.23
WW17	1.99	1.41
WW18	1.04	1.04
WW19	0.85	0.85
WW20	1.80	1.28

Similar to AFR plot in Fig. 11, one can also calculate and plot product mission reliability for each production vintage as follows:

$$\widehat{R}(t) = e^{-\left(\frac{t}{\hat{\eta}}\right)^\beta} \quad (18)$$

where t is any specified mission duration in terms of POH in the field and $\hat{\eta}$ is given by Eq. (17).

8 Concluding Remarks

In classical reliability engineering and life data analysis, the state of an item is binary: survive or fail, which is typically judged at macro or physical level. However, with the advancement of modern failure physics, sensing technology, and big data, the

Table 3 ORT predicted field AFR using Eqs. (16) and (17) based on true failure, holistic failure #1, and holistic failure #2

Work week	AFR based on true failure (%)	AFR based on holistic failure #1 (%)	AFR based on holistic failure #2 (%)
WW01	2	2	2
WW02	2	2	2
WW03	2	2	2
WW04	4	10	8
WW05	6	8	6
WW06	2	4	4
WW07	4	6	5
WW08	2	2	2
WW09	6	6	6
WW10	4	6	4
WW11	2	3	3
WW12	6	7	6
WW13	6	17	14
WW14	4	12	10
WW15	6	7	6
WW16	4	5	4
WW17	4	5	4
WW18	2	4	4
WW19	2	3	3
WW20	4	5	4

micro-level virtual failures become more and more “visible” and dominating. Introducing parametric-degradation-based virtual failure concept adds a micro-dimension to the traditional failure domain. Incorporating virtual failures into reliability assessment provides more insights about underlying product reliability and enhances the detection sensitivity of various reliability monitoring mechanisms, such as ongoing reliability test (ORT) during production and therefore can surface the poor vintage with potential high future failure rate due to “invisible” high degradation of critical parameters.

Author of this paper proposed numerous potential philosophies of describing the holistic failure and algorithms of quantifying the holistic failure counts, including complex number (or vector) approach, Yin-Yang theory, etc. An example is given to illustrate an application of applying holistic failures in parametric ORT reliability assessment. It is hoped that more discussions, explorations, and collaborations will be conducted to establish a solid and robust framework for holistic reliability assessment of modern era.

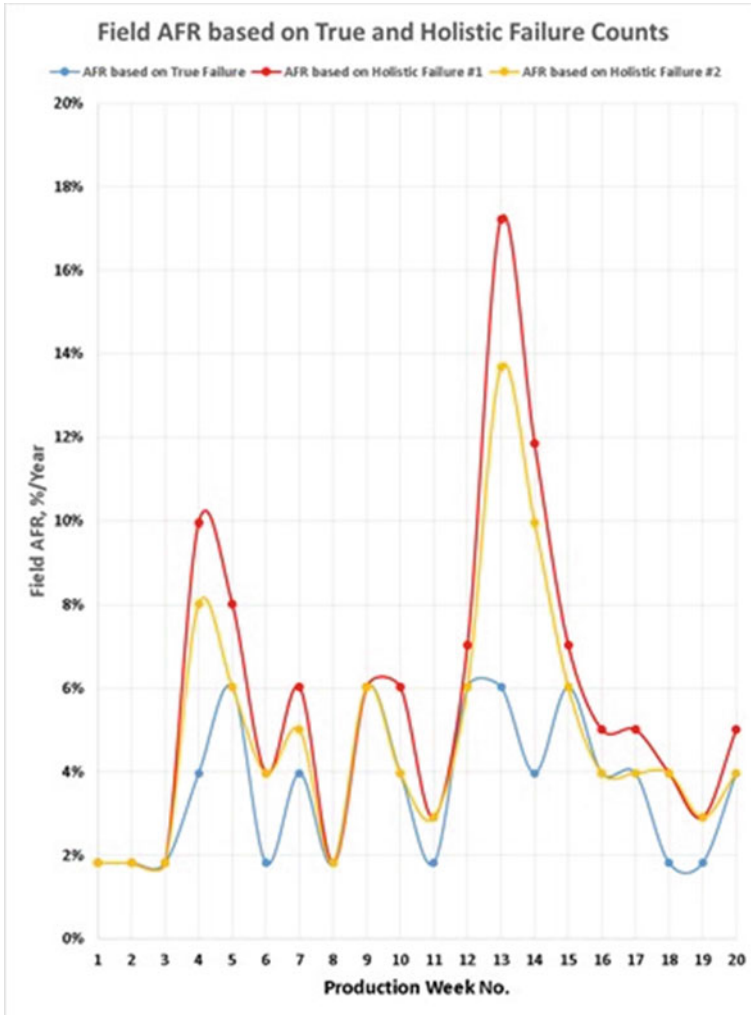


Fig. 11 ORT predicted field AFR based on true failure, holistic failure #1, and holistic failure #2

References

1. Sun F, Zhang J (2017) Dealing with fractional failures and fractional survivors in reliability engineering. In: Proceedings of the annual reliability and maintainability symposium, Orlando, Florida
2. Meeker Q, Hong Y (2014) Reliability meets big data: opportunities and challenges. Qual Eng 102–116
3. Hong Y, Zhang M, Meeker WQ (2018) Big data and reliability applications: the complexity dimension. Qual Eng 102–116
4. Sun F-B, Ou E, Zhang S (2009) Quantifying and controlling hard disk drive reliability through parametric monitoring during product ongoing reliability test. In: Proceedings of 15th ISSAT

- international conference on reliability and quality in design, San Francisco, California
5. Zhao S, Makis V, Chen S, Li Y (2016) Reliability evaluation for an electronic system subject to competing risks of dependent soft and hard failures. In: Proceedings of prognostics and system health management conference (PHM-Chengdu)
 6. Ma Z, Krings AW (2008) Competing risks analysis of reliability, survivability, and prognostics and health management (PHM). *Aerosp Conf (IEEE)*
 7. Bocchetti D, Giorgio M, Guida M, Pulcini G (2009) A competing risk model for the reliability of cylinder liners in marine diesel engines. *Rel Eng Syst Saf* 94(8):1299–1307
 8. Wenbiao Z, Elsayed EA (2004) An accelerated life testing model involving performance degradation. In: Proceedings of the annual reliability and maintainability symposium (RAMS)
 9. Guo C, Wang X, Guo B (2011) Competing risk model for long-stop-short-run systems. In: Proceedings of prognostics and system health management conference (PHM-Shenzhen). IEEE
 10. Wang H, Gao J (2014) A reliability evaluation study based on competing failures for aircraft engines. *Eksploatacja i Niezawodnosc-Maintenance and Reli*, vol 16
 11. Tang D, Yu J (2015) Optimal replacement policy for a periodically inspected system subject to the competing soft and sudden failures. *Eksploatacja i Niezawodnosc-Maintenance and Reli* 17(2):228–235
 12. Huang W, Askin RG (2003) Reliability analysis of electronic devices with multiple competing failure modes involving performance aging degradation. *Qual Reliab Eng Int* 19(3):241–254
 13. Peng H, Feng Q, Coit D (2010) Reliability and maintenance modeling for systems subject to multiple dependent competing failure processes. *IIE Trans* 43:12–22
 14. Jiang L, Feng Q, Coit D (2012) Reliability and maintenance modeling for dependent competing failure processes with shifting failure thresholds. *IEEE T Reliab* 61:932–948
 15. Rafiee K, Feng Q, Coit D (2014) Reliability modeling for dependent competing failure processes with changing degradation rate. *IIE Trans* 46:483–496
 16. Jiang L, Feng Q, Coit D (2015) Modeling zoned shock effects on stochastic degradation in dependent failure processes. *IIE Trans* 47:460–470
 17. Song S, Coit D, Feng Q (2014) Reliability for systems of degrading components with distinct component shock sets. *Reliab Eng Syst Saf* 132:115–124
 18. Song S, Coit D, Feng Q (2014) Reliability analysis for multi-component systems subject to multiple dependent competing failure processes. *IEEE T Reliab* 63:331–345
 19. Rafiee K, Feng Q, Coit D (2017) Reliability assessment of competing risks with generalized mixed shock models. *Reliab Eng Syst Safe* 159:1–11
 20. Fan M, Zeng Z, Zio E (2017) Modeling dependent competing failure processes with degradation-shock dependence. *Reliab Eng Syst Saf* 165:422–430
 21. Liu H, Yeh R, Cai B (2017) Reliability modeling for dependent competing failure processes of damage self-healing systems. *Comput Ind Eng* 105:55–62
 22. An Z, Sun D (2017) Reliability modeling for systems subject to multiple dependent competing failure processes with shock loads above a certain level. *Reliab Eng Syst Saf* 157:129–138
 23. Liu J, Song B, Zhang Y (2018) Competing failure model for mechanical system with multiple functional failures. *Adv Mech Eng* 10(5):1–16
 24. Liu C, Chen N, Ji P (2017) Operational reliability assessment using multivariate covariates. In: Proceedings of annual reliability and maintainability symposium
 25. Wikipedia, the free encyclopedia (2018a) Complex number. https://en.wikipedia.org/wiki/Complex_number
 26. Wikipedia, the free encyclopedia (2018b) Yin and yang. https://en.wikipedia.org/wiki/Yin_and_yang

Fengbin Sun Tesla Inc., California, USA

Dr. Feng-Bin (Frank) Sun is currently a Technical Lead and Principal Manager of Reliability Engineering at Tesla Inc. with over 30 years of industry and academia experience. He has published two books and more than 50 papers in various areas of reliability, maintainability, and

quality engineering. Dr. Sun served in the editorial board as an Associate Editor for the IEEE Transactions on Reliability from 1999 to 2003 and from 2016 to the present, Program Chair of the ISSAT International Conference on Reliability and Quality in Design since 2014, and a committee member and session moderator of numerous international conferences on reliability. He received his Ph.D. in Reliability from the University of Arizona and is a senior member of ASQ and the President of the Society of Reliability Engineers Silicon (SRE) Valley Chapter. Dr. Sun received the RAMS 2013 A.O. Plait Best Tutorial Award and RAMS 2020 P.K. McElroy Best Paper Award.

On the Aspects of Vitamin D and COVID-19 Infections and Modeling Time-Delay Body's Immune System with Time-Dependent Effects of Vitamin D and Probiotic



Hoang Pham

Abstract This chapter provides a summary of recent views on the aspects of vitamin D levels and the relationship between the prevalence rates of vitamin D deficiency and COVID-19 death toll of several countries in Europe and Asia. The chapter also discusses a new modified time-delay immune system model with time-dependent of the body's immune healthy cells, vitamin D, and probiotic. The model can be used to assess the timely progression of healthy immune cells with the effects of the levels of vitamin D and probiotics supplement. It also can help to predict when the infected cells and virus particles free state can ever be reached as time progresses with and without considering the vitamin D and probiotic supplements.

Keywords Vitamin D deficiency · Immune system · Virus particles · COVID-19 · Delay differential equation

1 Introduction

The COVID-19 pandemic has made a significant impact on the global economy and the daily life of each household around the world [1–3]. The COVID-19 symptoms such as sore throat, headache or shortness of breath, of infected persons are the same for both unvaccinated and vaccinated people but generally milder and lesser severe to vaccinated people [4]. COVID-19 has killed at least 1 million people and infected more than 90 million in the United States since the pandemic began [5]. Authorities in at least 220 countries have reported over 560 million COVID-19 cases and more than 6.3 million global deaths since the COVID-19 pandemic beginning in December 2019 [6]. Many researchers around the world [7, 8] have developed models to estimate the COVID-19 deaths and its effects due to various COVID-19 pandemic restrictions [9–11].

Some recent studied has observed that infected people with COVID-19 increase the risk of severe infection if they had lower vitamin D levels. Several other studies,

H. Pham (✉)

Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ 08854, USA

e-mail: hopham@soe.rutgers.edu

Table 1 Levels of vitamin D

Classification	Vitamin D deficiency “25(OH)D”
Severe deficiency	≤ 12 ng/mL (or ≤ 30 nmol/L)
Deficiency	≤ 20 ng/mL (or ≤ 50 nmol/L)
Insufficient	20–29 ng/mL (or 50–75 nmol/L)
Normal	30–50 ng/mL (or 75–125 nmol/L)

Notation ng = nanogram; mL = millilitre; nmol = nanomoles; L = litre; 1 ng/mL = 2.5 nmol/L

however, found no association between vitamin D levels and COVID-19 infected cases [12]. Similarly, other clinical studies have observed that vitamin D may reduce the COVID-19 infection rates and the mortality but the evidence for vitamin D prevention of the COVID virus will still need further studies.

Vitamin D is one of many vitamins our bodies need to stay healthy and supporting body’s immune system. The immune system needs vitamin D to fight off invading bacteria and viruses. Vitamin D deficiency (i.e. “25(OH)D”) is a serum 25-hydroxyvitamin D concentrations ≤ 20 ng/mL (50 nmol/L). The classifications of the vitamin D levels are shown in Table 1.

Some recent studies [13, 14] have shown vitamin D deficiency has been linked to colorectal cancer, prostate cancer, heart disease, and breast cancer. Several recent studies have also shown that people with sufficient levels of vitamin D have a lower risk of infection [15–17]. Some recent reviews [18, 19] and empirical designed studies [20–22] observed that vitamin D supplement can help to reduce symptoms for patient at high risk of respiratory infection. Several recent findings based on observational studies have illustrated a strong correlation between high COVID-19 mortality rate and vitamin D deficiency in most European countries.

Many researchers [23, 24] have actively studied and widely discussed with various guidelines about the sufficient vitamin D levels in some countries in the past decades but there is yet consent on what will be the best level of vitamin D [25, 26].

This chapter first discusses some recent views about the relationship between the vitamin D and disease infections related to COVID-19 pandemic and then presents a correlation analysis result between the prevalence rates of the deficiency levels of vitamin D and COVID-19 death tolls of 23 countries in Europe and 24 countries in Asia and all 47 countries in both continents Europe and Asia. The chapter also presents a modified mathematical model of time-delay interactions between body’s immune healthy cells, infected cells, and virus particles with consideration of vitamin D deficiency and probiotic supplements. It finally provides a brief conclusion and findings.

2 Prevalence Rates Analysis of Vitamin D Deficiency in European and Asian Countries

The prevalence rates of vitamin D deficiency (i.e. $25(\text{OH})\text{D} < 50 \text{ nmol/L}$) have reported of 24%, 37%, and 40% in US, Canada, and 40% in Europe [1, 27–29], respectively. The vitamin D deficiency has shown strongly related with various health outcomes and increased risk for COVID-19-related severity and high mortality rate based on related mortality distributions [30]. Recent study by Dror et al. [31] has demonstrated that people with lower vitamin D levels are more likely to have severe COVID-19 infections than people with higher vitamin D levels.

This section discusses the correlation analysis of the vitamin D deficiency and COVID-19 death rates per million people of many countries in Europe and Asia and a combination of all 47 countries in both continents. The number of COVID-19 reported cases (see column 7, Table 2) and deaths (column 8, Table 2) per one million as of January 15th, 2022 by countries in Asia and Europe were obtained from the Worldometer website [5]. The corresponding prevalence of vitamin D deficiency of 23 countries in European and 24 countries in Asia were obtained from the PubMed database [19] as shown in Table 2.

We observe that among European countries, the lowest and highest prevalence of vitamin D deficiency were reported in Finland and Ukraine with vitamin D deficiency values ranging from 6.9 to 81.8%, respectively, as shown in Table 2 [1]. The lowest and highest number of COVID-19 infected cases per million of the country population in Europe were from Finland (66,824 deaths per million) and Slovenia (253,591 cases per million) accordingly where the lowest and the highest number of COVID-19 deaths per million were from Norway (252 deaths per million) and Bulgaria (4,660 per million), respectively.

Figure 1 shows the vitamin D deficiency levels with respect to European countries (x-Axis), COVID-19 deaths (y-axis), and COVID-19 cases (z) for European countries. Similarly, the lowest and highest prevalence of vitamin D deficiency in Asia were reported in Vietnam and Oman with vitamin D deficiency values ranging from 2.0 to 87.5%, respectively. The lowest and the highest number of COVID-19 infected cases per million of the country population among Asian countries were from Vietnam (20,347 cases per million) and Bahrain (168,999 cases per million) respectively where the lowest and the highest number of COVID-19 deaths per million of the total population were from Vietnam (360 deaths per million) and Iran (1542 per million), respectively. Figure 2 shows the vitamin D deficiency levels with respect to Asian countries (x-Axis), COVID-19 deaths (y-axis), and COVID-19 infected cases (z) for Asian countries. Figure 3 is similar to Figs. 1 and 2 when both European and Asian continents were combined.

Table 3 presents the results of the correlation analysis of the prevalence of vitamin D deficiency, COVID-19 death, and infected cases rates per million in European and Asian continents [1]. We observe that

Table 2 The prevalence of vitamin D deficiency and COVID-19 death rate of 47 countries* [1]

No	Country	Continent	Sample	Age (yrs)	VitDdef	C19 cases per million	C19 deaths per million
<i>Europe</i>							
1	Belgium	3	697	20–69	51.1	206,627	2452
2	Bosnia	3	2483	>18	58.7	96,380	4229
3	Bulgaria	3	2016	20–80	75.8	118,495	4660
4	Croatia	3	3791	45.5	46.1	199,613	3215
5	Denmark	3	2565	18–69	51.5	183,639	598
6	Finland	3	798	30–64	6.9	66,824	310
7	France	3	297	18–65	75.1	207,186	1935
8	Germany	3	6995	18–79	61.5	94,020	1380
9	Greece	3	1084	>18	32.4	159,508	2116
10	Ireland	3	17,590	>18	51.5	217,571	1202
11	Italy	3	74,235	>18	33.3	141,515	2335
12	Norway	3	4465	40–69	24.7	90,464	252
13	Poland	3	5775	16–90	65.8	113,765	2707
14	Portugal	3	3092	>18	66.6	182,513	1898
15	Romania	3	812	>21	56.5	99,568	3110
16	Russia	3	1544	18–75	45.7	73,781	2196
17	Slovakia	3	578	>15	15	160,413	3158
18	Slovenia	3	125	18–64	58.2	253,591	2739
19	Spain	3	1262	20–83	33.9	172,992	1940
20	Sweden	3	268	18–80	22.4	153,038	1517
21	Switzerland	3	367	25–70	38.2	189,587	1435
22	Ukrain	3	1575	20–95	81.8	86,500	2266
23	United Kingdom	3	222	48–94	64.9	221,341	2220
<i>Asia</i>							
1	Bahrain	1	314	>30	79.9	168,999	780
2	Bangladesh	1	160	<70	63.7	9644	168
3	Brunei	1	446	>18	52	35,480	221
4	China	1	14,302	18–65	50.3	73	3
5	India	1	424	>18	55.2	26,499	347
6	Iran	1	7504	18–65	65.3	72,613	1542
7	Iraq	1	300	25–70	75.3	50,730	586
8	Japan	1	107	20–69	28.2	145,539	146
9	Jordan	1	3007	<83	67.9	105,948	1252
10	Kazakhstan	1	1347	>18	70	54,200	683

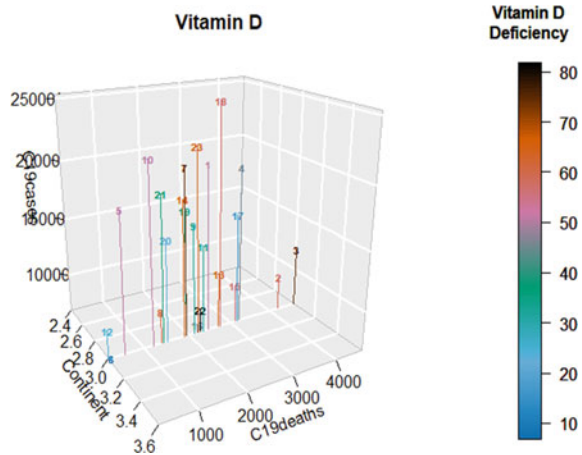
(continued)

Table 2 (continued)

No	Country	Continent	Sample	Age (yrs)	VitDdef	C19 cases per million	C19 deaths per million
11	Kuwait	1	960	>20	83	104,480	566
12	Lebanon	1	142,131	>19	35.5	119,291	1380
13	Malaysia	1	858	>18	67.4	84,911	962
14	Mongolia	1	320	20–58	35.5	120,499	620
15	Nepal	1	300	>18	51.3	28,363	388
16	Oman	1	206	18–55	87.5	58,492	776
17	Pakistan	1	1255	<84	36	5803	128
18	Qatar	1	102,342	18–65	71.4	105,143	222
19	Saudi Arabia	1	20–40	74.6		17,112	250
20	Singapore	1	114	21–100	42.1	48,984	142
21	Thailand	1	120	25–60	19.1	33,059	313
22	Turkey	1	11,734	>18	70.6	121,344	987
23	UAE	1	12,346	>18	72	79,631	217
24	Vietnam	1	637	18–87	2	20,347	360

* “1” in the category indicates Asian countries; “3” indicates European countries

Fig. 1 Vitamin D deficiency levels with respect to Continent (x-axis); COVID-19 deaths (y), and COVID-19 cases (z) among 23 countries in Europe



- The prevalence of vitamin D deficiency has a strong positive correlation with the COVID-19 deaths in European continent (i.e. correlation value (cv) = 0.4151) as shown in Table 3.
- The prevalence of vitamin D deficiency also has a positive correlation with the COVID-19 deaths in Asian continent with $cv = 0.2655$ as shown in Table 3.

Fig. 2 Vitamin D deficiency levels with respect to Continent (x-axis); COVID-19 deaths (y), and COVID-19 cases (z) among 24 countries in Asia

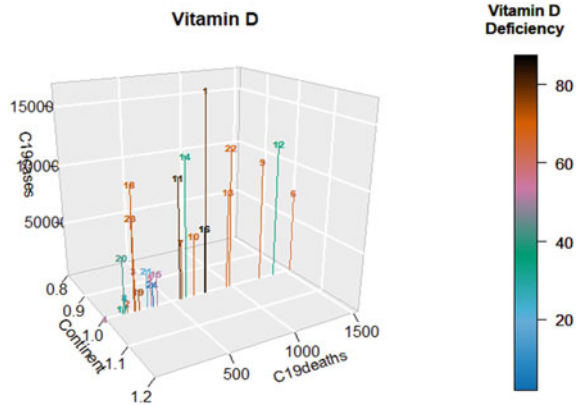


Fig. 3 Vitamin D deficiency levels with respect to Continent (x-axis); COVID-19 deaths (y), and COVID-19 cases (z) among countries in Asia and Europe

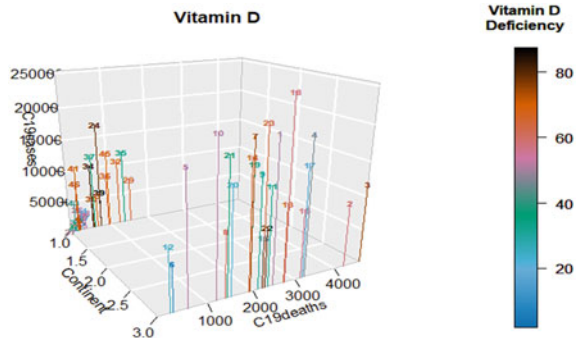


Table 3 Correlation values with the prevalence of vitamin D deficiency

Continent	Deaths (per million)	Infected cases (per million)
Europe continent	0.4151	0.1347
Asia continent	0.2655	0.3901
Europe & Asia continents	0.0949	0.0589

- The prevalence of vitamin D deficiency has a strong positive correlation with the COVID-19 infection cases in Asian continent (cv = 0.3901).

3 Mathematical Model Development

As we discussed in previous section, lower vitamin D levels such as vitamin D deficiency have been associated with increased susceptibility to infection, disease, and autoimmunity. The immune system needs vitamin D to fight off invading bacteria and

viruses. Recent research shows that high levels of vitamin D status can help to keep body's immune system healthy and likely protect against respiratory illnesses. Hospitalized patients with COVID-19 infections who had sufficient vitamin D levels has a decreased risk of adverse outcomes and death [31]. The immune system provides all the protective mechanisms needed to fight off infections or other dangerous agents [1].

Developing mathematical models to assess the growth of disease infections and immune healthy cells have been of interest in the area of cancer epidemiology, tumor growth and infectious disease in the past few decades [32–38]. Many models [35–38] have been studied in the past decades using the ordinary differential equations to characterize the tumor-immune dynamic growth of virus-infected cells of immune system [39]. Lestari et al. [40] studied an epidemic cancer model with chemotherapy in the form of a system of non-linear differential equations with three sub-populations. They presented the point of equilibrium and numerically determined the reproduction number of cancer cells. Pham [3] presented a model to estimate the US deaths toll related to COVID-19. Pham [7] recently discussed a time-dependent model considering various factors related to COVID-19 including social distancing and reopening states in different communities. Pham [41] recently developed a virus-immune time-delay model of immune system, and later extended Pham's earlier work [41] by taking the factor of chemotherapy drug treatment into the model consideration [42].

This section presents a modified mathematical time-delay model, from [1] considering time-dependent interactions between body's immune healthy cells, infected cells, and virus particles with the effect of vitamin D levels and probiotic. We describe model assumptions based on recent studies in [40–42] and discuss a model formulation consisting of three time-dependent variables in terms of partial differential equations that represent the time-dependent growth rate of immune cells, infected cells, and virus particles. We present numerical results to illustrate the model and discuss various cases to predict whether the infected cells or virus particles can be reached free state level or not as the time progresses.

Notation

The following notation are used in the model formulation:

a	infected cells death rate per unit time
s	coefficient of time-dependent growth rate of healthy cells
c	death rate of healthy cells unit time
g	coefficient of time-dependent healthy cells growth rate with probiotic supplement
x	the virus rate per unit time that enters the body
m	the rate that virus infects healthy cells
m_1	coefficient of time-dependent vitamin D rate that can enhance healthy cells
u	the rate that the healthy cells fail to fight infected cells
b	coefficient of time-dependent rate that the healthy cells fights and destroy infected cells

- n the rate that the healthy cells manages to fight off the virus due to healthy immune system
- k the rate that infected cells produce virus
- d virus death rate per unit time
- q the half saturation constant that related to healthy cells
- w the half saturation constant that related to virus
- $H(t)$ the time-dependent function of number of healthy cells at time t
- $I(t)$ the time-dependent function of number of infected cells at time t
- $V(t)$ the time-dependent function number of virus in the body at time t .

The assumptions of a modified model of time-delay interactions between body’s immune healthy cells, infected cells, and virus particles with the effects of vitamin D deficiency and probiotic supplement are as follows [1]:

1. The growth rate of healthy cell has a constant coefficient s .
2. The healthy cell has a natural death rate, c .
3. There is a reduced number of healthy cells due to the virus infects healthy cell at the rate m .
4. Vitamin D time-dependent growth function to enhance healthy cells that can fight off the virus at the constant coefficient m_1 .
5. The healthy cells fail to fight infected cells with the rate u .
6. Time-dependent healthy cells growth coefficient g by using probiotic supplement with a delay τ_1 time.
7. The healthy cells are able to fight infected cells at the rate b and also manages to fight off the virus at the rate n with a time delay τ_2 .
8. The infected of virus particles and the healthy cells became the infected cells at the rate m .
9. The infected cell has a death rate per unit time a
10. The healthy cells manage to fight off the infected cells at a time-dependent function with coefficient b .
11. The infected cells produce the virus cells at the rate k
12. The virus particles have a death rate per unit time d
13. The healthy cells manage to fight off the virus particles at the time-dependent function with coefficient n .
14. The virus enters the body at the rate x per unit time with a time delay τ_3 .

The mathematical model includes three time-dependent variables: immune healthy cells $H(t)$, infected cell $I(t)$, and virus particles $V(t)$. Virus particles invaded healthy cells with the rate m (i.e., $mH(t) V(t)$) and transform them into infected cells. Healthy cells, infected cells, and virus particles die at the rate c , a , and d , respectively.

The rate functions of the number of immune healthy cells $H(t)$, infected cells $I(t)$, and virus particles $V(t)$ over time t of the modified model from [1] can be written based on the assumptions (1–7), (8–10), and (11–14) respectively, as follows:

$$\frac{\partial H(t)}{\partial t} = st^{0.01} - cH - mHV + m_1t^{0.02}HV - uHI$$

$$\begin{aligned}
 &+ \frac{gt^{0.001}H(t - \tau_1)I(t - \tau_1)V(t - \tau_1)}{q + I(t - \tau_1)V(t - \tau_1)} \\
 &+ bnI(t - \tau_2)V(t - \tau_2)H(t - \tau_2)
 \end{aligned} \tag{1}$$

$$\frac{\partial I(t)}{\partial t} = mHV - aI - bt^{0.01}IH \tag{2}$$

$$\frac{\partial V(t)}{\partial t} = kI - dV - nt^{\frac{1}{5}}VH + \frac{xV(t - \tau_3)H(t - \tau_3)}{W + H(t - \tau_3)}. \tag{3}$$

For the model without considering the time-dependent function, the above model can be rewritten as:

$$\begin{aligned}
 \frac{\partial H(t)}{\partial t} = &s - cH - mHV + m_1HV - uHI + \frac{gH(t - \tau_1)I(t - \tau_1)V(t - \tau_1)}{q + I(t - \tau_1)V(t - \tau_1)} \\
 &+ bnI(t - \tau_2)V(t - \tau_2)H(t - \tau_2)
 \end{aligned} \tag{1a}$$

$$\frac{\partial I(t)}{\partial t} = mHV - aI - bIH \tag{2a}$$

$$\frac{\partial V(t)}{\partial t} = kI - dV - nVH + \frac{xV(t - \tau_3)H(t - \tau_3)}{W + H(t - \tau_3)}. \tag{3a}$$

Figure 4 illustrates the modified mathematical presentation of the above equations. We develop an algorithm using *R* software to numerically obtain the results for $H(t)$, $I(t)$, and the number of virus particles $V(t)$ as we will discuss with various cases in numerical examples in Sect. 3.1.

$$\begin{aligned}
 f_1 &= \frac{gt^{0.001}H(t - \tau_1)I(t - \tau_1)V(t - \tau_1)}{q + I(t - \tau_1)V(t - \tau_1)} \\
 f_2 &= bnI(t - \tau_2)V(t - \tau_2)H(t - \tau_2) \\
 f_3 &= \frac{xV(t - \tau_3)H(t - \tau_3)}{W + H(t - \tau_3)}.
 \end{aligned}$$

3.1 Numerical Examples

This section presents some numerical results based on the above delay system of differential equations using the numerical values based on some existing studies [40–42] as shown in Table 4 for the illustration of the model. Note that one can easily apply the proposed model to obtain the numerical results for any other sets of parameter values.

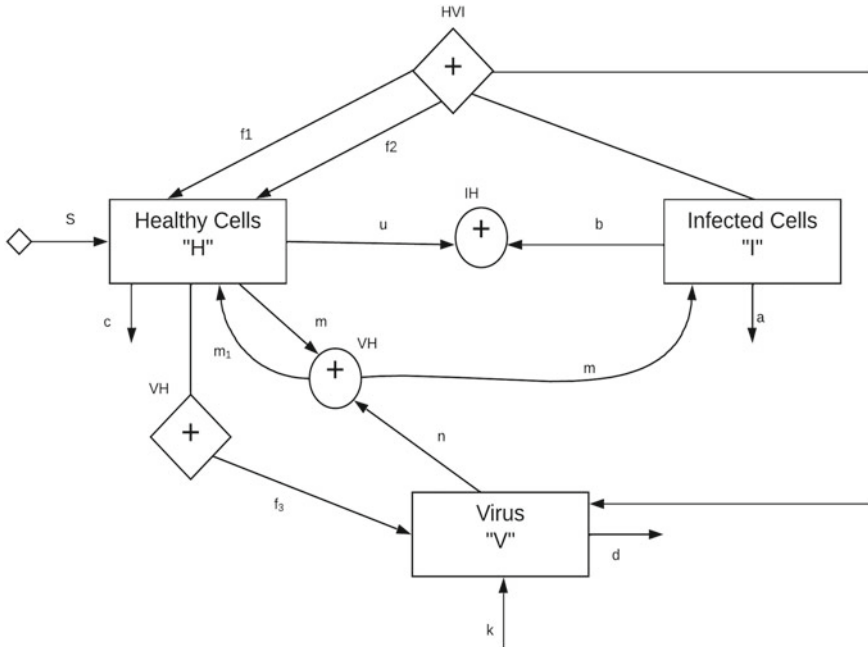


Fig. 4 Mathematical process presentation of time-delay interactions between body’s immune healthy cells, infected cells, and virus particles model [1]

Table 4 Model parameter values

$s = 0.43/\text{day}$	$c = 0.00212/\text{day}$	$m = 0.0004/\text{day}$
$u = 0.0002/\text{day}$	$g = 0.01/\text{day}$	$q = 0.20/\text{day}$
$b = 0.0001/\text{day}$	$n = 0.00005/\text{day}$	$a = 0.004/\text{day}$
$k = 0.0070/\text{day}$	$d = 0.004/\text{day}$	$x = 0.006/\text{day}, w = 0.90/\text{day}$

In this analysis, we assume that the initial number of virus particles, infected cells, and healthy cells are 30,000, 35,000 and 75,000, respectively, for the algorithm and developed program can search to find the best solutions. It is worth to note that the results do not necessary depend on the assumptions of these initial numbers but analyst can provide any numbers as initial values for the purpose of numerically solving the system of differential equations. We now present various cases with some parameter values of the healthy cell growth rate as well as the rate that infected cells and virus particles where the first, second and third time-delay are 1, 2, and 2 days, respectively.

Case 1: Time-dependent function with $m_I = 0$ and $g = 0$. Here we do not take into considerations of having the vitamin D and probiotic supplement, that is $m_I = 0$ and $g = 0$.

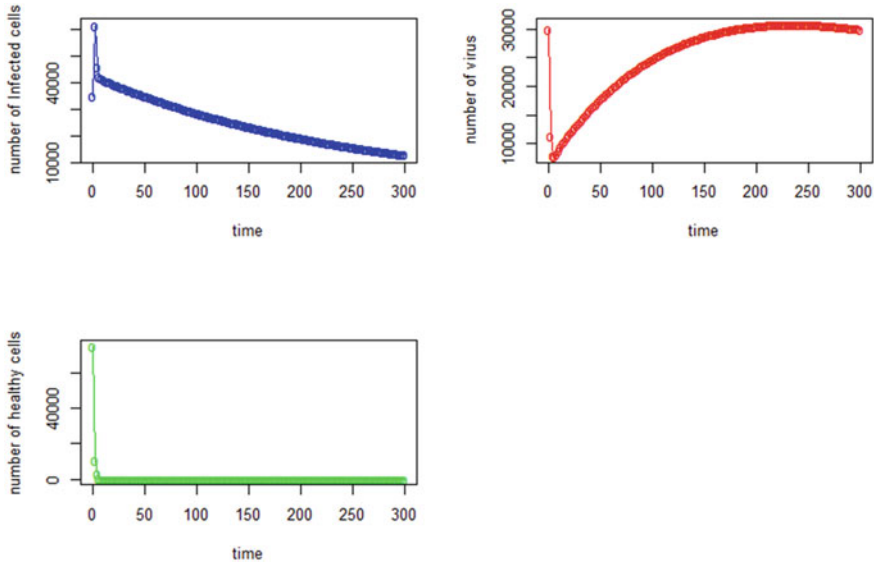


Fig. 5 Time-dependent functions where $m_1 = 0$ and $g = 0$

From Fig. 5, we observe as expected the initial number of virus particles, infected cells, and healthy cells are 30,000, 35,000 and 75,000, respectively. The infected cells begins to increase sharply until the 2nd day then starts to decrease until it slowly stable after the 290th day at the level of around 13,000 (see Fig. 5 in “blue”). The virus particles begins to decrease sharply until the 6th day but starts to increase until it slowly stable at around 30,000 which is about the 280th day (see Fig. 5 in “red”). The healthy cells (see Fig. 5 in “green”) decreases sharply at the beginning and it reaches the level of empty healthy cell at the level at about the 40th day.

For without considering the time-dependent function as given in Eqs. 1a, 2a and 3a, there are slightly but not significant different on the numerical results as shown in Fig. 6.

Case 2: Time-dependent functions (see Eqs. 1–3).

Here we assume that with a sufficient vitamin D level, it can enhance the number of healthy cells that can help to fight off the infected cells and virus particles since the healthy cells can be boosted due to the vitamin D sufficiency level. In the model, $m_1 = 0.00005$. Likewise, we assume that by taking a sufficient probiotic supplement, it can enhance the number of healthy cells that can help to fight off the infected cells and virus particles where in the model $g = 1.8$ for example. In other words, in this example we take into considerations of taking the vitamin D and probiotic supplement where $m_1 = 0.00005$ and $g = 1.8$. As seen in Fig. 7, the healthy cells (see Fig. 7 in “green”) decreases sharply at the beginning until around the 7th day and starts to increase quickly that reaches the maximum number of healthy cells level at around the 27th day and then begins to decrease until they slowly stabilize

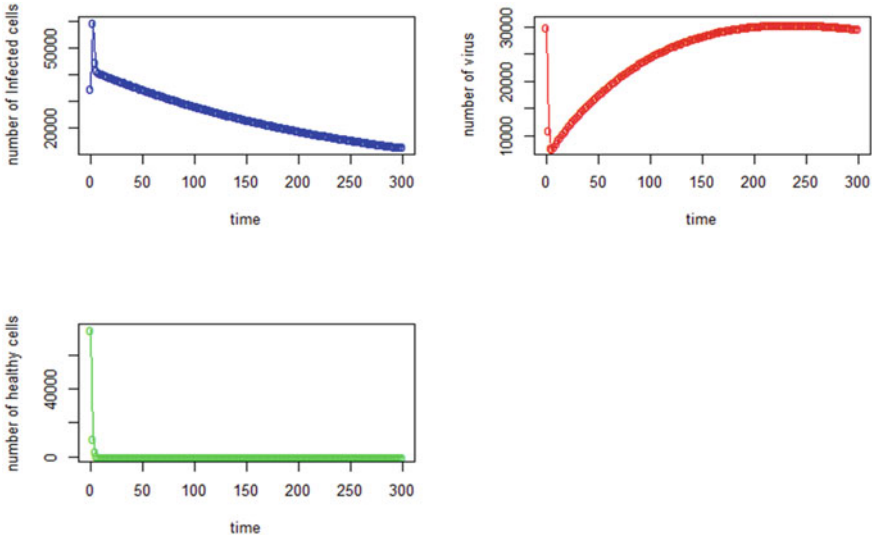


Fig. 6 Without time-dependent function and $m_1 = 0, g = 0$

at around the 325th day at the level of 260,000. Unlike the previous cases, although the infected cells slightly increase at the beginning, both the infected cells and virus particles decrease significantly to until they reach the infected cells (see Fig. 7 in “blue”) free state and virus particles (Fig. 7 in “red”) after around the 20th day as shown in Fig. 7.

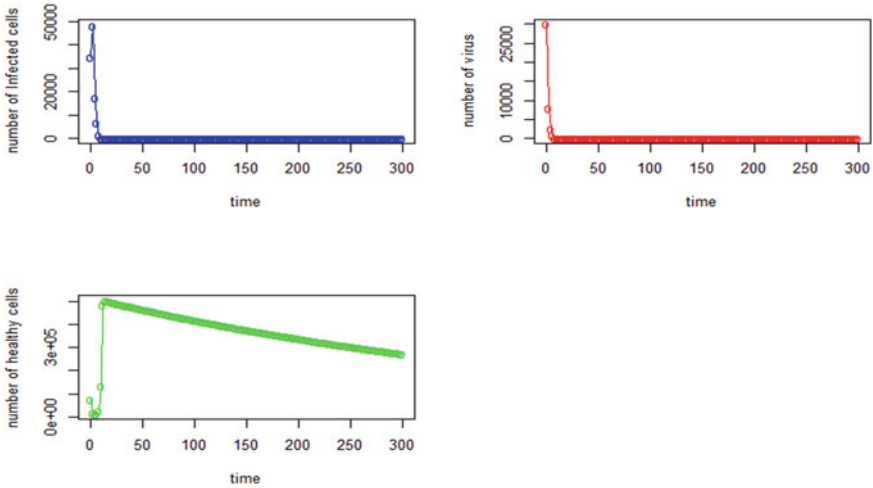


Fig. 7 Time-dependent functions

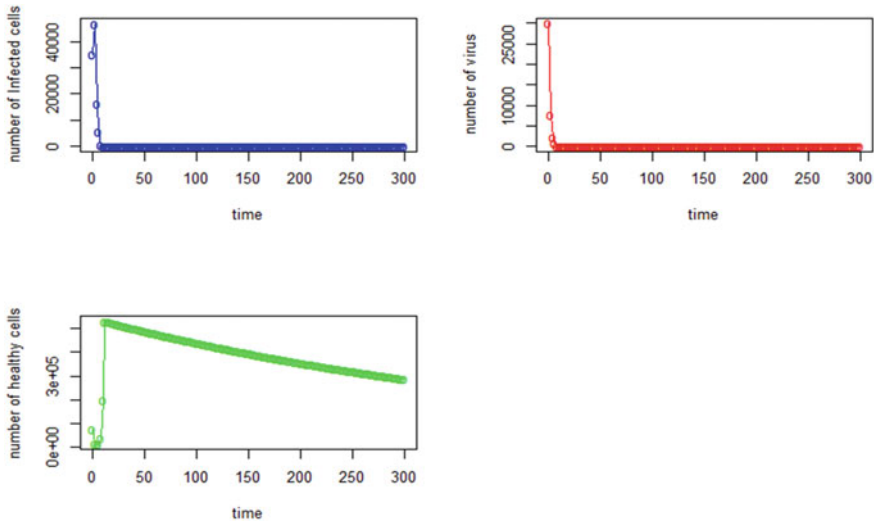


Fig. 8 Without time-dependent function

For not considering the time-dependent function case as given in Eqs. 1a, 2a and 3a, there are slightly but not significant different results from the above. As can be seen from Fig. 8, the healthy cells (see Fig. 8 in “green”) first decreases sharply at the beginning until around the 4th day and starts to increase quickly that reaches the maximum number of healthy cells level at around the 25th day and then begins to decrease until they slowly stabilize at around the 300th day at the level of 288,000. Similarly to the above case,, both the infected cells and virus particles decrease significantly at the beginning to until they reach the infected cells (see Fig. 8 in “blue”) free state and virus particles (Fig. 8 in “red”) free state after around the 20th day as shown in Fig. 8. Figures 9, 10 and 11 shows various values for the level of vitamin D and probiotic supplement, where $m_1 = 0.0005$ and $g = 1$, $m_1 = 0.005$ and $g = 0.7$, and $m_1 = 0.000005$ and $g = 0.01$, respectively.

4 Conclusion

This chapter studied the relationship between the vitamin D deficiency and COVID-19 infected cases and deaths toll among the countries in Europe and Asia. The results agreed with some recent studies that there seems to be a strong association between vitamin D deficiency, the COVID-19 deaths, and COVID-19 infected cases, and that the lower vitamin D levels were associated with higher COVID-19 deaths. The modeling results show an important finding that the appropriate of time-delay duration that will take those cells (i.e., body’s immune healthy cells, infected cells

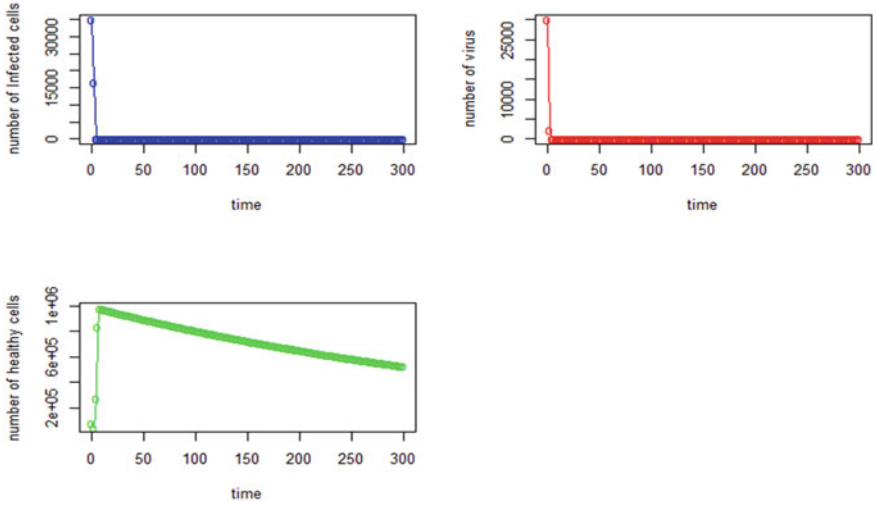


Fig. 9 Time-dependent function with and $m_1 = 0.0005, g = 1$

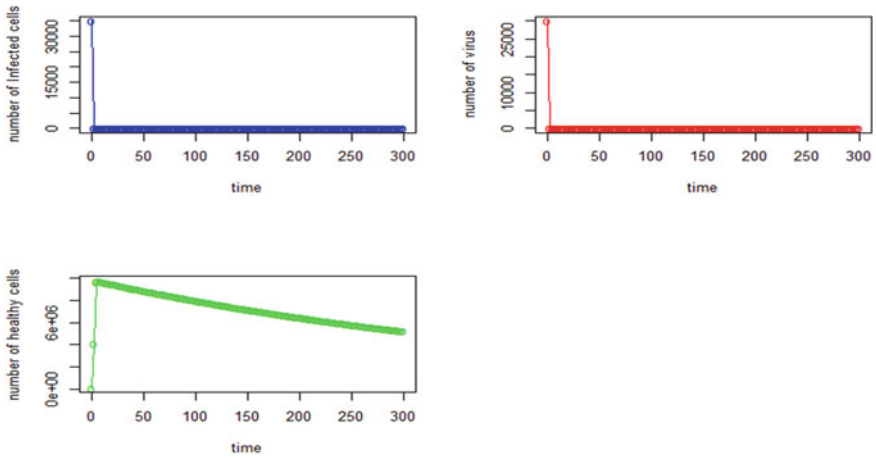


Fig. 10 Time-dependent function with and $m_1 = 0.005, g = 0.7$

and virus particles) to effectively spread in the body is crucial in the model and that can make a significant difference of the modeling results.

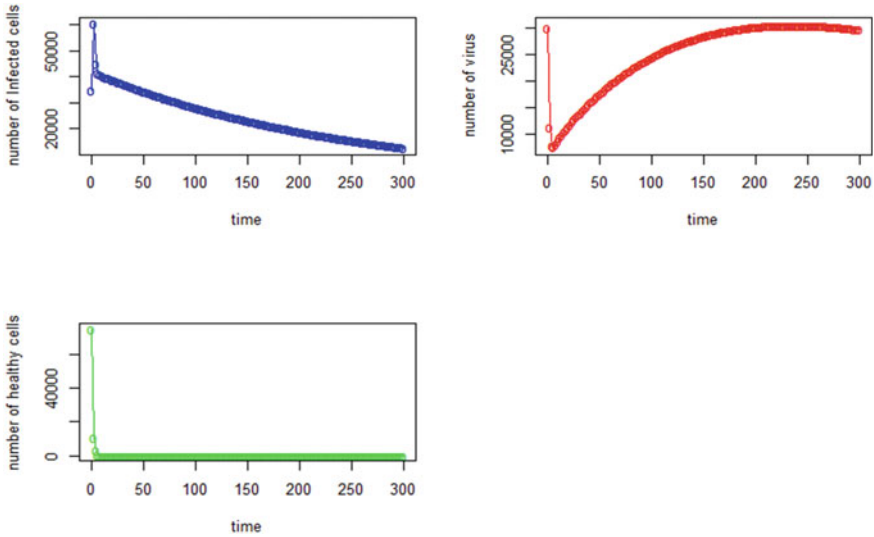


Fig. 11 Time-dependent function with $m_1 = 0.000005$, $g = 0.01$

References

1. Pham H (2022) Analyzing the relationship between the vitamin D deficiency and COVID-19 mortality rate and modeling the time-delay interactions between body’s immune healthy cells, infected cells, and virus particles with the effect of vitamin D levels. *Math Biol Eng* 19(9). <https://doi.org/10.3934/mbe.2022417>
2. <https://www.osha.gov/SLTC/covid-19/medicalinformation.html>. Accessed 17 Mar 2020
3. Pham H (2020) On estimating the number of deaths related to Covid-19. *Mathematics* 8:655. <https://doi.org/10.3390/math8050655>
4. Centers for Disease Control and Prevention. <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html>. Accessed 8 Apr 2020
5. Worldometers (2022). https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?#countries. Accessed 1 Apr 2022
6. World Health Organization. <https://covid19.who.int/>. Accessed 1 Apr 2022
7. Pham H (2020) Estimating the COVID-19 death toll by considering the time-dependent effects of various pandemic restrictions. *Mathematics*
8. Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 20:533–534. [https://doi.org/10.1016/s1473-3099\(20\)30120-1](https://doi.org/10.1016/s1473-3099(20)30120-1)
9. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Jit M, Klepac P, Flasche S, Clifford S et al (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* 20(5):e261–e270. [https://doi.org/10.1016/s2468-2667\(20\)30073-6](https://doi.org/10.1016/s2468-2667(20)30073-6)
10. Sebastiani G, Massa M, Riboli E (2020) Covid-19 epidemic in Italy: evolution, projections and impact of government measures. *Eur J Epidemiol* 35:341–345. <https://doi.org/10.1007/s10654-020-00631-6>
11. Wang X, Zhang Y, Fang F (2021) Role of vitamin D in COVID-19 infections and deaths. *J Evid Based Med* 14:5–6
12. Urashima M, Segawa T, Okazaki M, Kurihara M, Wada Y, Ida H (2010) Randomized trial of vitamin D supplementation to prevent seasonal influenza A in schoolchildren. *Am J Clin Nutr* 91:1255–1260

13. Pham H, Pham DH (2020) A novel generalized logistic dependent model to predict the presence of breast cancer based on biomarkers. *Concurr Comput: Pract Exp* 32:1
14. Boersma D, Demontiero O, Mohtasham AZ, Hassan S, Suarez H, Geisinger D, Suriyaarachchi P, Sharma A, Duque G (2012) Vitamin D status in relation to postural stability in the elderly. *J Nutr Health Aging* 16:270–275
15. de Haan K, Groeneveld ABJ, de Geus HRH, Egal M, Struijs A (2014) Vitamin D deficiency as a risk factor for infection, sepsis and mortality in the critically ill: systematic review and meta-analysis. *Crit Care* 66:6604
16. Ingham TR, Jones B, Camargo CA, Kirman J, Dowell AC, Crane J (2014) Association of vitamin D deficiency with severity of acute respiratory infection: a case-control study in New Zealand children. *Eur Respir J Eur Respir Soc* 44:124
17. Martineau AR, Jolliffe DA, Hooper RL, Greenberg L, Aloia JF, Bergman P (2017) Vitamin D supplementation to prevent acute respiratory tract infections: Systematic review and meta-analysis of individual participant data. *BMJ* 65:6583
18. Richardson DP, Lovegrove JA (2021) Nutritional status of micronutrients as a possible and modifiable risk factor for COVID-19: a UK perspective. *Br J Nutr* 125:678–684
19. Grant WB, Lahore H, McDonnell SL, Baggerly CA, French CB, Aliano JL, Bhatta HP (2020) Evidence that vitamin D supplementation could reduce risk of influenza and COVID-19 infections and deaths. *Nutrients* 12:988
20. Maghbooli Z, Sahraian MA, Ebrahimi M, Pazoki M, Kafan S, Tabriz HM, Hadadi A, Montazeri M, Nasiri M, Shirvani A et al (2020) Vitamin D sufficiency, a serum 25-hydroxyvitamin D at least 30 ng/mL reduced risk for adverse clinical outcomes in patients with COVID-19 infection. *PLoS ONE* 15:e0239799
21. Martineau AR, Forouhi NG (2020) Vitamin D for COVID-19: a case to answer? *Lancet Diabetes Endocrinol* 8:735–736
22. Ali N (2020) Role of vitamin D in preventing of COVID-19 infection, progression and severity. *J Infect Public Health* 13:1373–1380
23. Holick MF, Binkley NC, Bischoff-Ferrari HA, Gordon CM, Hanley DA, Heaney RP et al (2011) Evaluation, treatment, and prevention of vitamin D deficiency: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 96:1911–1930. <https://doi.org/10.1210/jc.2011-0385>
24. Pfeifer M, Begerow B, Minne HW, Abrams C, Nachtigall D, Hansen C (2000) Effects of a short-term vitamin D and calcium supplementation on body sway and secondary hyperparathyroidism in elderly women. *J Bone Miner Res* 15:1113–1118. <https://doi.org/10.1359/jbmr.2000.15.6.1113>
25. Pietras SM, Obayan BK, Cai MH, Holick MF (2009) Vitamin D2 treatment for vitamin D deficiency and insufficiency for up to 6 Years. *JAMA Int Med* 169:1806–18. <https://doi.org/10.1001/archinternmed.2009.361>
26. Rusińska A, Płudowski P, Walczak M, Borszewska-Kornacka MK, Bossowski A, Chlebna-Sokół D et al (2018) Vitamin D supplementation guidelines for general population and groups at risk of vitamin D deficiency in Poland—Recommendations of the Polish Society of Pediatric Endocrinology and Diabetes and the expert panel with participation of national specialist consultants and representatives of scientific societies. *Front Endocrinol* 9. <https://doi.org/10.3389/fendo.2018.00246>
27. Jain A, Chaurasia R, Sengar NS, Singh M, Mahor S, Narain S (2020) Analysis of vitamin D level among asymptomatic and critically ill COVID-19 patients and its correlation with inflammatory markers. *Sci Rep* 10(1):20191. <https://doi.org/10.1038/s41598-020-77093-z>
28. Schleicher RL, Sternberg MR, Looker AC, Yetley EA, Lacher DA, Sempos CT et al (2016) National estimates of serum total 25-Hydroxyvitamin D and metabolite concentrations measured by liquid chromatography–Tandem mass spectrometry in the US population during 2007–2010. *J Nutr* 146:1051–61. <https://doi.org/10.3945/jn.115.227728>
29. Sarafin K, Durazo-Arvizu R, Tian L, Phinney KW, Tai S, Camara JE et al (2015) Standardizing 25-hydroxyvitamin D values from the Canadian health measures survey. *Am J Clin Nutr* 102:1044–1050. <https://doi.org/10.3945/ajcn.114.103689>

30. Pham H (2011) Modeling U.S. mortality and risk-cost optimization on life expectancy. *IEEE Trans Reliab* 60(1):125–133
31. Dror A, Morozov N, Daoud A, Namir Y, Yakir O, Shachar YR, Lifshitz M, Segal E, Fisher L, Mizrahi M, Eisenbach N, Rayan D, Gruber M, Bashkin A, Kaykov E, Barhoum M, Edelstein M, Sela E (2022) Pre-infection 25-hydroxyvitamin D3 levels and association with severity of COVID-19 illness. *PLOS*. <https://doi.org/10.1371/journal.pone.0263069>
32. Ng J, Stovecky YR, Brenner DJ, Formenti SC, Shuryak I (2021) Development of a model to estimate the association between delay in cancer treatment and local tumor control and risk of metastases. *JAMA Netw Open* 4:1–10
33. Talkington A, Durrett R (2015) Estimating tumor growth rates in vivo. *Bull Math Biol* 77:1934–1954
34. Vaghi C, Rodallec A, Fanciullino R, Ciccolini J, Mochel JP, Matri M (2020) Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PLoS Comput Biol* 16:e1007178. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007178>. Accessed 25 Apr 2021
35. Yin DJ, Moes AR, van Hasselt JGC, Swen JJ, Guchelaar HJ (2019) A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors. *CPT Pharmacomet Syst Pharmacol* 8:720–737
36. Jackson TL, Byrne HMA (2000) A mathematical model to study the effects of drug resistance and vasculature on the response of solid tumors to chemotherapy. *Math Biosci* 104:17–38
37. Meacham CE, Morrison SJ (2013) Tumour heterogeneity and cancer cell plasticity. *Nature* 501:328–337
38. Sun X, Bao J, Shaob Y (2016) Mathematical modeling of therapy-induced cancer drug resistance: connecting cancer mechanisms to population survival rates. *Sci Rep* 6:22498
39. Taniguchi K, Okami J, Kodama K, Higashiyama M, Kato K (2008) Intratumor heterogeneity of epidermal growth factor receptor mutations in lung cancer and its correlation to the response to gefitinib. *Cancer Sci* 99:929–9354
40. Lestari ER, Arifah H (2019) Dynamics of a mathematical model of cancer cells with chemotherapy. *J Phys Conf Ser* 1320:1–8
41. Pham H (2021) A dynamic model of multiple time-delay interactions between the virus-infected cells and body's immune system with autoimmune diseases. *Axioms* 10:216. <https://doi.org/10.3390/axioms10030216>
42. Pham H (2022) Mathematical modeling the time-delay interactions between tumor viruses and the immune system with the effects of chemotherapy and autoimmune diseases. *Mathematics* 10(5), 756. <https://doi.org/10.3390/math10050756>

Hoang Pham Rutgers University, New Jersey, USA

Dr. Hoang Pham is a Distinguished Professor and former Chairman (2007–2013) of the Department of Industrial and Systems Engineering at Rutgers University. His research areas include reliability modeling and prediction, software reliability, and statistical inference. He is editor-in-chief of the *International Journal of Reliability*, and editor of *Springer Series in Reliability Engineering*. Dr. Pham is the author or coauthor of 7 books and has published over 200 journal articles, 100 conference papers, and edited 17 books including *Springer Handbook in Engineering Statistics* and *Handbook in Reliability Engineering*. He is a Fellow of the IEEE, AAIA, and IISE.

A Staff Scheduling Problem of Customers with Reservations in Consideration with Expected Wait Time of a Customer without Reservation



Junji Koyanagi

Abstract In cell phone stores in Japan, they serve customers with reservations and ones without reservations. The manager of the shop assigns a staff to each customer with reservations (CR) at the start of or before the day. For the customers without reservations (CN), they are served by the staff who does not serve other customers. However, no staff is available to serve CN, CN must wait until some staff can serve CN. The assignment of CR to the staff should be scheduled not to wait CR. The wait time of CN should be as short as possible, though they are served not to delay the service of CR. The minimization of the expected wait time of the first CN is considered in this research. Taboo search method is adopted for the optimization of the evaluation function to assign each reservation to each staff and to minimize the waiting time of the first CN.

1 Introduction

There are several service systems such as nursing care and customer service in the shop. It is important to assign the staff to the customers or reservations suitably. Usually, there are several constraints in the assignment. If the reservations are made without considering the actual constraints, it may be a problem to find the feasible schedule [1, 2].

In this research, we deal with a staff assignment problem in a cell phone shop. In cell phone shop in Japan, the reservations are mainly made by telephone and web site. Therefore, feasibility of the service is usually ensured, when the reservation is accepted. The customers with reservation are served when they arrive the shop. On the other hand, there are customers without reservations because of the reasons

J. Koyanagi (✉)
Graduate School of Tottori University, Koyama-Minami 4-101, Tottori-City, Tottori 680-8552,
Japan
e-mail: junji@tottori-u.ac.jp

such as the sudden trouble of the phone. These customer will have to wait until their service starts and the wait time for these customers should be as short as possible. Thus, we consider the wait time of the customer without reservation. Though it is desirable to consider all possible customers without reservation (CN), it consumes long time to calculate. Therefore, we consider to minimize the wait time of the first CN in this research.

2 Reservation and Assignment

We assume the following conditions of the shop.

1. The number of reservations is n . The reservation is numbered from 0 to $n - 1$.
2. The number of staff is m . The staff is numbered from 0 to $m - 1$.
3. The business hours of the shop is divided to time zones, and each time zone has equal length (15 min in this research). The time zone is numbered from 0 to $z - 1$. The start time of the time zone t is called time t and the end time of the time zone t is called time $t + 1$. The reservation occupies several consecutive time zones and the start time of the reservation coincides to the start of time zone and the end time of the reservation also coincides to the end of time zone.

For the reservation r , we assume the following conditions.

1. The reservation is made before the day, usually by web site. For the reservation r , the customer inputs the start time zone b_r of the reservation and the reservation type such as purchase of new phone and request of repair. According to the reservation type, the standard end time zone e_r is set.
2. If the reservation r is assigned to staff s , the start time zone is changed to $b_r - \alpha_{rs}$ and the end time zone is changed to $e_r + \beta_{rs}$ (Fig. 1).
 - (a) For $\alpha_{rs} \geq 0$, α_{rs} is considered as the preparation time if the reservation r is assigned to staff s . If staff s is a novice, the staff may need preparation to deal with reservation r . The second role of α_{rs} is to indicate the reservation r can be assigned to staff s . If $\alpha_{rs} = -1$, it indicates that the reservation r cannot be assigned to staff s , for example, because the skill level of staff s is low for reservation r .
 - (b) β_{rs} is to adjust the end time of reservation r . If staff s is a novice, the end time tends to be late. In that case, $\beta_{rs} > 0$. If staff s is a skilled staff, the end time tends to be earlier. In that case, $\beta_{rs} < 0$.

For the assignment, we assume the following conditions.

- A1. Each reservation is served by one staff. If reservation r needs to be served by more than two staff, the reservation r can be copied for the number of staff.
- A2. The staff can serve only one reservation in a time zone. The staff cannot serve parallelly more than two reservations.

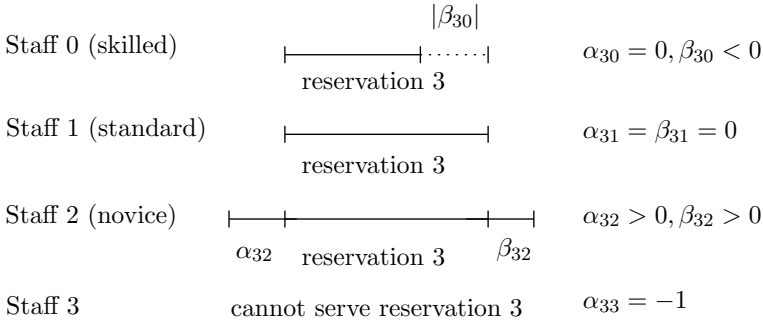


Fig. 1 Skill adjustment of the reservation

A3. The staff takes the lunch time of fixed length, while the lunch time zone. For example, the staff takes 45 min lunch time in lunch time zone from 11:00 to 14:30.

If the reservations are made by web site, the above condition can be satisfied because the system will reject the reservation beyond the capacity. Since the reject is based on standard time of reservation r , it may be difficult to find the assignment to satisfy the conditions when α_{rs} and β_{rs} are considered. Moreover, the assignment should consider the wait time of CN, then we have to consider better assignment of CR to minimize the wait time of CN.

3 Assignment Evaluation

The evaluation of the schedule is the wait time of the first CN. For the assignment, there are blank time zones, which indicate the time zones without reservations. If CN comes, the customer is assigned to the blank time zones with enough length. It is assumed that CN arrives at the start of time zone and fixed number of time zones (3 time zones, 45 min here) is required for the service of CN. Since the customer is served in the time zones without reservation, the customer may wait until the suitable time zone for the service. For example, we explain the wait time in Table 1 (three staffs and three reservations). In the table, the number shows the reservation number that the staff serves.

1. If CN arrives at 15:00, the customer can be served by either staff 0 or 1, then the wait time is 0.
2. If CN arrives at 15:15, the customer can be served by staff 0, then the wait time is 0. Staff 1 cannot serve the customer because staff 1 must start reservation 2 at time 15:45.
3. If CN arrives at 15:30, the customer must wait until 16:30 for staff 2, then the wait time is 60 min.

Table 1 The wait time of the first customer without reservation

Time zone (start time)	Staff 0	Staff 1	Staff 2	Wait time (min)
15:00			1	0
15:15			1	0
15:30			1	60
15:45		2	1	45
16:00	0	2	1	30
16:15	0	2	1	15
16:30	0	2		0
16:45	0	2		0
17:00	0			0
17:15	0			0
17:30	0			–
17:45	0			–

4. In a similar way, the wait time is 45, 30, 15, 0, if CN arrives at 15:45, 16:00, 16:15, 16:30, respectively.
5. Let the shop close at 18:00, then CN after 17:30 is out of service.

Expected Wait Time

Expected wait time for the first CN is calculated from the wait time, if the arrival probability for each time zone is available. However, it is difficult to obtain the probability, then we assume the probability is same for each time zone here. The minimization of the expected wait time is the objective in this research. For the assignment of Table 1, the expected wait time is $(60 + 45 + 30 + 15)/10 = 15$ because the arrival probabilities are equal for the ten time zones.

4 Tabu Search

This problem is a kind of combinatorial optimization [3]. To express the problem as combinatorial problem, the assignment of the reservations is expressed by $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})$ for n reservations. The element x_r shows the staff number who serves the reservation r . For example $x_2 = 3$ means the reservation 2 is assigned to staff 3. By this expression, condition A1 is automatically satisfied. It is difficult to find the assignment which satisfies A2 and A3. We use penalty function method and tabu search to minimize the expected wait time and to satisfy A2 and A3. For the assignment \mathbf{x} , we can make the timetable as shown Fig. 2. From this timetable, we can calculate the following three functions. In Fig. 2, an example of 3 staffs and 6

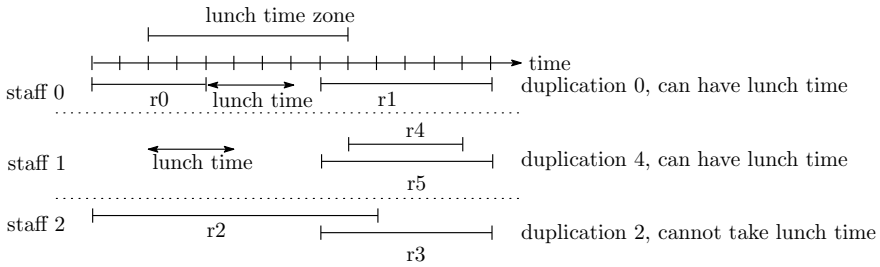


Fig. 2 Calculation of $D(\mathbf{x})$ and $L(\mathbf{x})$

reservations, reservation 2 is denoted by r2 and other reservations are denoted in a similar way. The assignment vector $\mathbf{x} = (0, 0, 2, 2, 1, 1)$.

$D(\mathbf{x})$: This function counts the sum of duplication of the reservations for each staff. $D(\mathbf{x})$ can be calculated by checking the start time and the end time for all reservation combinations assigned for each staff. If reservation r and r' is assigned to staff s , let $b_r - \alpha_{r's} \leq b_{r'} - \alpha_{r's}$ (reservation r starts earlier than (or at the same time as) reservation r'). Note that reservation r occupies the time zones from $b_r - \alpha_{r's}$ to $e_r + \beta_{r's}$, then it ends at time $e_r + \beta_{r's} + 1$.

1. If $e_r + \beta_{r's} + 1 \leq b_{r'} - \alpha_{r's}$, then no duplication between reservation r and r' . In Fig. 2, it is the case of reservation 0 and 1 for staff 0, ($r = 0, r' = 1, s = 0$)
2. If $e_r + \beta_{r's} + 1 > b_{r'} - \alpha_{r's}$, duplication is calculated for the following two cases.
 - a. If $e_r + \beta_{r's} \geq e_{r'} + \beta_{r's}$, the time zones for reservation r' exists in the time zones for reservation r . Then the duplication is estimated by $e_{r'} + \beta_{r's} + 1 - (b_{r'} - \alpha_{r's})$, which is the case of reservation 4 and 5 for staff 1 in Fig 2.
 - b. If $e_r + \beta_{r's} < e_{r'} + \beta_{r's}$, the duplication occurs in the time zones from $b_{r'} - \alpha_{r's}$ to $e_r + \beta_{r's}$. Then the duplication is estimated by $e_r + \beta_{r's} + 1 - (b_{r'} - \alpha_{r's})$, which is the case of reservation 2 and 3 for staff 2 in Fig 2.

Thus, for the assignment vector $\mathbf{x} = (0, 0, 2, 2, 1, 1)$, the value of $D(\mathbf{x})$ is $0 + 4 + 2 = 6$.

$L(\mathbf{x})$: This function counts the number of staff who cannot take lunch time. In Fig. 2, $L(\mathbf{x}) = 1$, because staff 2 cannot take lunch time (lunch time length is three time zones). $L(\mathbf{x})$ can be calculated by the following loop checking consecutive time zones of lunch time length in the lunch time zone.

Step 0. For the staff s , let the start time zone of the lunch time zone be l_s . If time zone l_s is occupied by some reservation, increase l_s until the time zone l_s becomes blank. Then we assume time zone l_s is a blank time zone.

- Step 1. If $l_s + h_s - 1$ is larger than the end of lunch time zone, then staff s cannot take lunch time, and exit the loop, where h_s is the lunch time length for staff s .
- Step 2. If time zones $l_s + 1, \dots, l_s + h_s - 1$ are not occupied by any reservations, then staff s can take lunch time and exit the loop. If time zone l'_s ($l_s + 1 \leq l'_s \leq l_s + h_s - 1$) is occupied by some reservation r , then increase l_s until the blank time zone after the end of reservation r . Then return to Step 1.

$L(\mathbf{x})$ is the number of the staffs who cannot take lunch time, derived by the above process.

$W(\mathbf{x})$: This function calculates the expected wait time of the first CN. The way of calculation is shown in Table 1. For each staff, the earliest time zone where the staff can serve CN with arrival time zone t is searched in a similar way to calculate $L(\mathbf{x})$. Then the staff who can serve CN as early as possible is selected and the wait time of CN is the difference between the arrival time zone of CN and the service start time zone. After calculating the wait time for every possible arrival time zone, the expected wait time is calculated by assuming some arrival probabilities for all time zones.

For the large coefficient C_1, C_2 (100,000 for example), tabu search is used for objective function $V(\mathbf{x}) = C_1 D(\mathbf{x}) + C_2 L(\mathbf{x}) + W(\mathbf{x})$. By minimizing the objective function, it is expected that $D(\mathbf{x}^*) = L(\mathbf{x}^*) = 0$ for the (nearly) optimal solution \mathbf{x}^* .

Tabu Search

Tabu search [4] is performed as follows with initial assignment \mathbf{x} and taboo list set $T = \{\mathbf{x}\}$

1. Make the neighborhood set X . In this research, X is made by one element is changed from \mathbf{x} .
2. Calculate $V(\mathbf{x}')$ for $\mathbf{x}' \in X$ if \mathbf{x}' does not exist in taboo list set T . Then choose \mathbf{x}^* among \mathbf{x}' which minimize $V(\mathbf{x}^*)$.
3. Add \mathbf{x}^* to T . Then repeat 1. by replacing \mathbf{x} by \mathbf{x}^* .

After repeating the loop by fixed time, we adopt the optimal solution which minimizes the $V(\mathbf{x})$ in the above process.

5 Numerical Example

We could get the reservation data at a cell phone shop from Seiryō Electric Corporation. By this data, we made Table 2 on 2020/Dec/1 and expected wait time (assume the uniform distribution of the arrival) is 1.25 min for the first CN with 45 min request. The number in the table shows the reservation number, and L shows lunch time for 45 min. Lunch time can be moved in 12:00–14:30 if the first CN arrives.

Table 2 Time Table on 2020/Dec/1

Time zone	Staff 0	Staff 1	Staff 2	Staff 3	Staff 4	Staff 5	Staff 6
10:00	29	24	20		11	0	5
10:15	29	24	20		11	0	5
10:30	29	24	20		12	0	5
10:45	29	24	20		12	0	5
11:00	29	24	6		12	0	22
11:15	29	24	6		12	0	22
11:30					12	1	22
11:45					12	1	22
12:00	2	16	L	L	L	L	22
12:15	2	16	L	L	L	L	22
12:30	L	16	L	L	L	L	23
12:45	L	16					23
13:00	L	L	30			17	23
13:15		L	30			17	23
13:30		L	30		7	17	L
13:45		25	30		7	17	L
14:00		25	30		7	3	L
14:15		25	30		7	3	
14:30		25			7	3	
14:45		25			7	3	
15:00		25			13	3	8
15:15		18			13	3	8
15:30		18	4	31	13	26	8
15:45			4	31	13	26	8
16:00		34	4	31	9	26	14
16:15		34	4	31	9	26	14
16:30		34	4	31	9	27	14
16:45		34	4	31	9	27	14
17:00				10	32	27	14
17:15				10	32	27	15
17:30		33		10	32		15
17:45		33		10	32		15
18:00			28	19			15
18:15			28	19			15
18:30			28	19			15
18:45			28	19			15
19:00							
19:15							
19:30							
19:30							
19:45							

6 Conclusion

We deal with the assignment problem in cell phone shop. By considering the wait time of the first customer, we can make the assignment which minimizes the expected wait time of the customer. This method seems to be useful for problem size 30 reservations.

Acknowledgements This research is supported by SEIRYO Electric Corporation. The author would like to thank the company for providing the shop reservation data and financial support.

References

1. Sigeno M, Ikegami A (2018) Periodic schedule in staff schedule: making and application. *Oper Res Soc Jpn* 3:134–138 (In Japanese)
2. Ikegami A, Uno T et al (2012) Optimization problem mainly considering running cost— Available in small company. *Oper Res Soc Jpn* 695–704 (In Japanese)
3. Korte B, Vygen J (2012) *Combinatorial optimization: theory and algorithms*. Springer
4. Gendreau M, Potvin J-Y (eds) (2018) *Handbook of metaheuristics*. Springer

Junji Koyanagi Tottori University, Tottori, Japan

Dr. Junji Koyanagi is an Associate Professor in the graduate school of sustainability science, Tottori University. His initial research deals with the maintenance problem of queueing systems. Then, his research area is the optimization problems of various systems, usually including stochastic factors.

Decision Support System for Ranking of Software Reliability Growth Models



Devanshu Kumar Singh, Hitesh, Vijay Kumar, and Hoang Pham

Abstract The requirement for Software Reliability Growth Models (SRGMs) has increased exponentially in response to the growing demand for strong and reliable software systems. During the testing phase of the Software Development Life Cycle (SDLC), SRGMs are particularly effective for estimating fault content, minimizing testing expenses, and maximizing software reliability. There has been a lot of research into selecting the best SRGMs for a certain failure dataset and then ranking all the SRGMs against the dataset. In this chapter, we have studied the mentioned problem and the solution to automate it with the developed compact Decision Support System (DSS), which includes all the functionalities and computational analysis of error logs and ensure error-free software to achieve the desired objective. The DSS is developed in Python utilizing several well-known packages such as Numpy, Scipy, Tkinter, and Pandas. To rank SRGMs employed in the DSS, we used Entropy & Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) ranking methodology. The implemented schema provides highly accurate performance indexes for the SRGMs required for efficient ranking, emphasizing the significance of the proposed prototype of DSS in the open literature, being a novel and ingenious development in the domain of software reliability.

Keywords SRGMs · Software testing · Decision support system (DSS) · Python · TOPSIS · Entropy

D. K. Singh · Hitesh

Department of Computer Science & Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

V. Kumar (✉)

Department of Mathematics, Amity Institute of Applied Sciences, Amity University, Noida, Uttar Pradesh, India

e-mail: vijay_parashar@yahoo.com

H. Pham

Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

e-mail: hopham@soe.rutgers.edu

1 Introduction

With the gradual and perpetual growth of the computer technology, the modern society has witnessed the significance of the reliable computer software in variety of real-world applications. The management of computer system has a vital role, which needs long-term consistency and reliability in the software framework. Computer industry has vastly impacted the functioning and operability of almost all the domains which in turn raises the significance of the development constraints such as time and budget constraints. With rising competition in the professional domain, the work quality needs a constant and consistent pace. Of course, the deep-rooted computer systems that are the very foundation of the framework of the job completion in any organization, the need for the dexterous and reliable computer software. Nevertheless, when a new service or product is introduced to the customers, it often suffers from the problems due to the ever-changing market scenario and field environment, which necessitates the regular updates to the system. Considering the sustainable and broad spectrum of software systems, the most important aspect of the entire procedure is the software reliability that is directly proportionate to the growth and enhancement of any organization or human life.

Software reliability is defined as a program's ability to perform its functions under specific conditions for a specific time [1]. Software reliability, according to ANSI, is defined as the likelihood of fault-free software functioning in a predefined environment for a specified time [2]. Software quality, usability, functionality, serviceability, performance, maintainability, capability, and documentation all contribute to software reliability. Due to the high convolution of software, obtaining software reliability is difficult. Due to the rapid growth of system space and the feasibility of upgrading the software, system developers tend to increase the level of complexity within layers of software. It is difficult to achieve a satisfactory level of reliability for a highly complex system containing software [3]. The number of users is directly influenced by software reliability. In practice, a reliable piece of software should have backup (redundant) code that runs the exception handling checks. As a result, the space complexity increases, necessitating more storage, which has an impact on the execution speed. However, for the following reasons, reliability still takes precedence over efficiency since compared to reliability, efficiency can easily be solved by utilizing better hardware resources. A user (or organization) would never use a less reliable or unreliable software that would hinder its growth and productivity. Furthermore, for deadline-based applications, an unreliable software might lead to the loss of invaluable information due to system crash. Origin of unreliability is unclear in a software system with distributed faults through its structure.

Failure, according to the Institute of Electrical and Electronics Engineers (IEEE), is the beginning of a fault, or an anomaly in the software's execution path. This means that, while software reliability is an important factor in evaluating software performance, it is also important to distinguish between reliability and failure. It is said that a reliable product has had its infrequently occurring flaws removed. This emphasizes the need for tools and techniques to organize reliability and fault logs

so that bug-free software can be produced. The SRGMs oversee cost-estimation, release time, and software reliability. The SRGM stands for a *Software Reliability Growth Model* that describes how software faults behave over time. SRGMs aid in the time-consuming and costly process of extensive testing and debugging. Software engineers can use SRGMs to predict expected software reliability, which allows software managers to profile a framework for allocating money, time, and resources to a project. These models also make estimating software completion simple. One of the most important customer-oriented attributes of software quality is software reliability. Effective approaches to developing reliable software, as well as quantitatively estimating software reliability, are critical. Many time-dependent Non-Homogenous Poisson Process (NHPP) based SRGMs have been developed over the last 6 decades to determine the reliability of software systems.

Various models have been investigated based on various assumptions. Some models, for example, assume perfect debugging, while others consider imperfect debugging, as well as fault removal efficiency, error generation, and the learning process. Other factors in the real-world development process, such as testing coverage, testing effort, time-delay fault correction, and fault reduction factor, are also factored in to improve the estimation accuracy of SRGMs [4–6]. The underlying assumption in most models is that the operating environment is the same as the testing environment, and that the software used in the operating environment has the same failure-occurrence behavior as the software used in the software testing environment. As a result, models that are appropriate for software testing data are also appropriate for field data [7]. However, this assumption may not always be true, especially for software that will be used in a variety of field environments after it is released. The in-house testing environment is frequently a more controlled environment with less variation than the field environments; however, the field operation environments for software vary significantly from one location to the next or from one application to the next [4–7]. It has also been observed that software reliability in the field differs from that in the testing environment, so a proportional constant of the environmental factor has been defined to characterize the discrepancy between the testing and field environments in terms of the severity of usage conditions, and several studies on field-oriented software reliability assessment have been carried out.

The SRGMs exist in abundance, such as Goel and Okumoto model [8], PNZ model [9], Teng and Pham model [10], Yamada Imperfect model [11], etc., that are associated with and developed with a specific structure and type of dataset and features in mind. It should go without saying that a specific SRGM designed for one type of dataset may not be appropriate for another type of dataset. This leads one to believe that there may be an SRGM that is better suited to dealing with a given dataset than others, as that SRGM may have been more concerned with the type of dataset at the time of development, making it more appropriate to choose.

Because of the difficulty in determining the ranking of SRGMs on a consistent basis, there is an urgent need for a gauging tool or mechanism that can determine the efficiency of that SRGM in relation to the dataset used. The technique, in theory, is to use optimization algorithm concepts to identify the ranking of SRGMs based on performance indices such as MSE, R^2 , and others, by first estimating the model

parameters from the initial parameters input. The ranking of SRGMs is then determined using Multi Criteria Decision Making (MCDM) methodologies. To address the dilemma of ranking of SRGMs, we researched and developed a DSS that allows software managers to analyze and rate commonly used SRGMs in software management during the testing phase of the software development life cycle (SDLC). The proposed DSS is built with the Python programming language 3.9.6, with GUI support provided by Tkinter, Pandas 1.3.1 for data processing, and Numpy 1.19.5 and SciPy 1.7.1 are used for all mathematical procedures. Aside from the DSS tool, the proposed study includes a mechanism for ranking models based on Entropy and TOPSIS.

This Chapter is further divided into following sections describing its contributions, structuring, etc.:

1. Section 2 reveals the literature about SRGMs and selection methods in the field of MCDM approaches along with the chosen approach of ranking.
2. Section 3 deals with the description of the performance indexes with a brief account and formulae of each index.
3. Section 4 discusses the ranking methodology used to develop the DSS.
4. Section 5 discusses the DSS environment, going over the working and implementation of the DSS, and touching on some of the DSS's complexities.
5. Section 6 gives an outline of the results and analysis of the SRGMs over the employed dataset after implementing the DSS.
6. Section 7 describes the conclusion of the chapter.

2 Literature Review

Over the last six decades, a plethora of SRGMs have been developed to assess the reliability growth parameters during the SDLC testing phase. Goel and Okumoto [8] created a game-changing model that can be used during the testing phase. They proposed a model that depicted fault detection as a NHPP and assumed a link between software faults and detection rate. By incorporating the unpredictability of operating environments and establishing a normalized criteria distance method, Pham [12] proposed a V-tub shaped fault detection rate model to rank SRGMs. Teng and Pham [10] proposed a new approach to software reliability prediction based on the idea that a unit-free environmental element can detect all the arbitrary effects of the field environments, which are represented as a randomly distributed variable. For the operating and testing stages of software development, the authors developed a random field environment software reliability model. On the scenario of variability and anomalies in operating environments, Chang et al. [13] developed an innovative testing-coverage software reliability model. The Pham Nordmann Zhang (PNZ) model [9] is a concave and inflection S-shaped model that predicts component-based system or software reliability using fault tolerance structures techniques. The PNZ model [9], which is based on the NHPP, is thought to be one of the best.

Yamada et al. [11] presented two software reliability models that used simple debugging techniques. The proposed models consider the assumption of frequent

occurrence of errors, as faults that have been dormant in the software system are discovered and corrected during the testing phase. Pham and Zhang [14] presented a new model based on the NHPP concept, as well as a comparison study of the presented model with other NHPP models that looked at the variability of operating environments. Zhang et al. [15] developed a fault removal model to assess software reliability measures by considering factors like learning phenomenon, imperfect debugging, and failure rate, as well as integrated fault introduction rate, fault removal proficiency, and failure rate. Zhu and Pham [16] developed a model with stochastic fault process. Ohba et al. [17] developed an inflection S-shaped model that established a more feasible estimation base than the conventional exponential type SRGM. Furthermore, parameter evaluations for datasets related to calendar time are more accurate. Roy et al. [18] used an NHPP-based model with a static error detection rate and an exponential detection rate function to model the fault generation phenomenon and imperfect debugging. Yamada et al. [19] investigated the software fault detection process using a delayed S-shaped model. For the given data of the detected software faults, the observed growth curve was S-shaped. Kumar et al. [20] developed a model with testing effort in dynamic environment. Kumar et al. [21] also presented reliability modelling for fault detection and fault correction processes.

One of the most difficult challenges in the sphere of MCDM problems is to assign weights to the various criteria or parameters for ranking the options. This difficulty can be tackled with the implementation of Shannon's Entropy to evaluate Entropy Weight Method (EWM). The definition of entropy is derived from the thermodynamics principle. The quantity of valid information produced by the system is measured by entropy, which indicates the degree of randomness of the system. Shannon's metric of entropy [22, 23] is a basic concept in information theory. Information can be regarded as variables that can be stored or transferred, according to the Information Theory. As a result, a variable may be thought of as a storage unit that can hold different values from a preset pool of values at different points in time with respect to activities. In simple terms, information is obtained by reading the values of variables, which can be understood by reading the text of an email, for example. As a result, the data can be thought of as the process that controls the variable [22].

The Entropy of a Variable is a measure of how much information a variable store. However, the amount of information provided by a variable is not solely determined by the size of the pool of values the variable can acquire, just as the amount of information provided by an email is not solely determined by the number of words in it or the number of unique words in the email's vernacularism. As a result, the amount of information available is proportional to the amount of "disclosure" provided by its reading. For example, if an email is a repeat of a previous email, it is not considered informational. However, if an email reveals the outcome of a contentious election, it is priceless information. As a result, the quantity of astonishment disseminated owing to the contents of the variable [23] can be linked to the information stored by a variable.

In recent decades, the viable selection of SRGMs has piqued the interest of many researchers. It accelerated the development and implementation of many MCDM techniques [24]. The major MCDM methods that are commonly used are Technique

for Order of Preference by Similarity to Ideal Solution (TOPSIS) [25], Combinative Distance Based Approach (CODAS) [26], VIKOR [27], DBA [28], and Elimination et Choice Translating Reality (ELECTRE) [29]. To solve the problem of optimal selection of SRGMs, Saxena et al. [30] developed the methodology of CRITIC-TOPSIS approach. Saxena et al. [31] also used the ELECRE approach to select the optimal SRGMs.

3 Performance Indexes

The performance index identification is frequently disregarded when choosing the best SRGM. An exclusive set of performance indices governs the selection of SRGMs. There are a total of 12 similar indices [32] in the free literature. A major stage in the selection of SRGMs is correctly identifying and categorizing these performance indices. The detailed literature review aids in the identification of the most relevant performance indicators for selecting, assessing, and ranking SRGMs. For the DSS, we looked at five different performance indicators. The following is a brief description of each of the five indexes:

1. Mean Squared Error (MSE): It evaluates the discrepancies among the predicted data with the real data. The formula of MSE is as follows

$$\text{MSE} = \frac{1}{n - N} \sum_{i=1}^n (y_i - \hat{m}(t_i))^2 \quad (1)$$

2. R Squared (R^2): It decides whether the fit is successful in describing the variation of the data as follows–

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{m}(t_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

3. Akaike Information Criterion (AIC): It is a tool for calculating out-of-sample prediction error and, as a result, the relative quality of statistical models for a given set of data. AIC calculates the quality of each model in relation to the other models in the following way for different models:

$$\text{AIC} = n \cdot \ln\left(\frac{\text{SSE}}{n}\right) + 2N \quad (3)$$

where,

$$\text{SSE} = \sum_{i=1}^n (y_i - \hat{m}(t_i))^2 \quad (4)$$

4. Bayesian Information Criterion (BIC): It is similar to AIC, except the penalty term is determined by the sample size.

$$\text{BIC} = n \cdot \ln\left(\frac{\text{SSE}}{n}\right) + N \log n \quad (5)$$

5. Pham's Criterion (PC): When the sample size is too small, the penalty is increased slightly each time parameters are added to the model.

$$\text{PC} = \left(\frac{n - N}{2}\right) \log \frac{\text{SSE}}{n} + N \left(\frac{n - 1}{n - N}\right) \quad (6)$$

4 Ranking Methodology

For the proposed work, we used TOPSIS [25] technique which is a Multi-Criteria Decision Making (MCDM) approach established by Ching-Lai Hwang and Yoon [25]. It stands for Technique for Order of Preference by Similarity to Ideal Solution. It is based on the principle that the preferred option must have the shortest geometric distance from the positive ideal solution (PIS) while also having the largest geometric distance from the negative ideal solution (NIS). When it comes to multi-criteria decision making, TOPSIS is a powerful tool to use, with the basic strategy being to determine the positive ideal solution and negative ideal solution first. Clearly, the worst ideal hypothesis value is chosen as the negative ideal solution in each situation, and the best ideal hypothesis value is chosen as the positive ideal solution, according to each criterion that is taken into account during the decision-making process. The weighted Euclidean distance between each alternative and the positive and negative ideal solutions is calculated once this process is complete. We can calculate the value of similarity for each alternative using this newly calculated data, which consists of distances between each alternative and the positive and negative ideal solutions. The SRGMs are sorted based on these similarity values, and the best solution is then found. An illustration for the sequential steps of the employed Entropy-TOPSIS methodology is provided in Fig. 1.

5 Environment of the DSS

The GUI of the DSS is built using Tkinter library of Python programming language. The main screen of the DSS is shown in Fig. 2. The DSS is divided in three major sections, namely.

- Computation Section
- Models Section

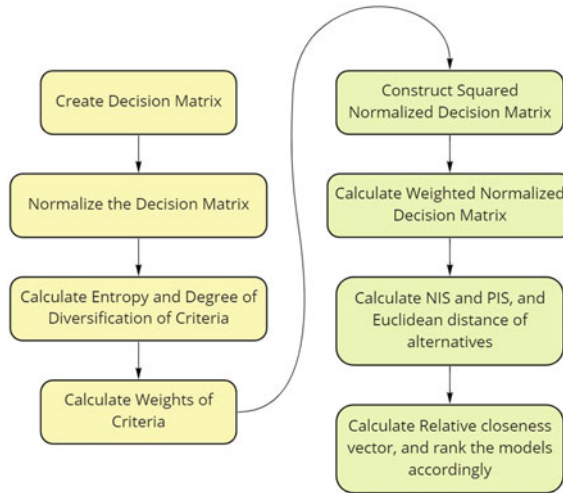


Fig. 1 Illustration for the steps of Entropy-TOPSIS methodology

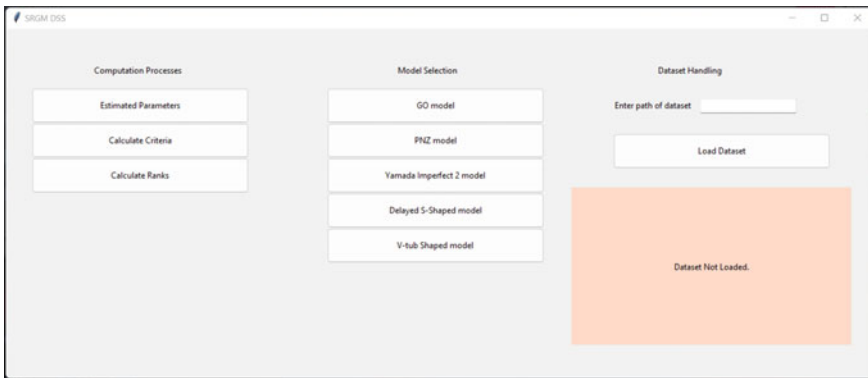


Fig. 2 Main screen of the DSS before dataset is loaded

- Dataset Section.

The first step is to load the dataset in the DSS, this takes place in the dataset section. The entry field requires the path of the dataset in the form which is acceptable by the *read_csv* method of Pandas package of Python, which is used for handling the dataset. The path must be start with the name of the root storage disk for example C, D, G etc. followed by a colon and two backslashes *':'//'*. After that each subdirectory must be separated by a backslash until the directory where the dataset csv file resides is reached, and the path ends with the name of the dataset csv file. It must be sure that the independent variable column or the time column must be named 'Time' and the dependent variable column must be named 'CDF' in the dataset csv file.

On clicking the load dataset button, the DSS looks up the path entered and generates an error message if the path entered is wrong, and on the contrary if the path entered is correct a successful load message is prompted.

The message that is prompted on successful loading of the dataset is shown in Fig. 3. If the path entered is wrong or the dataset csv file does not exist, then in that case the DSS prompts an error message as shown in Fig. 4.

After loading the dataset successfully, the DSS is now able to display the dataset. Earlier until the dataset is not loaded, the dataset window displays the message “Dataset Not Loaded”. But once the DSS loads the dataset, the dataset window displays the Time and Cumulative Faults columns of the dataset as Time and CDF columns respectively. This is shown in Fig. 5.

Now that the dataset has been loaded, the selection of models is to be performed. The DSS provisions 5 models to select from for comparison and ranking over the dataset. These models are.

1. GO Model [8]
2. PNZ Model [9]
3. Yamada Imperfect 2 Model [11]
4. Delayed S-Shaped Model [19]
5. V-tub Shaped Model [12].

The selection process requires clicking on the models to be selected for ranking, and input of their initial model parameters to estimate the optimal model parameters using Least Squares Estimation (LSE) method. The Numpy and SciPy package of Python are used for the simplicity of implementation and accuracy of results. Figure 6 shows windows for input of initial model parameters. On clicking the Submit button

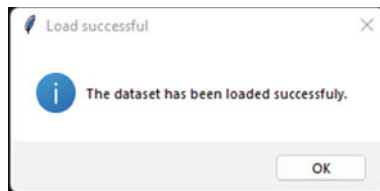


Fig. 3 Message for successful load of dataset



Fig. 4 Message for load failure of dataset

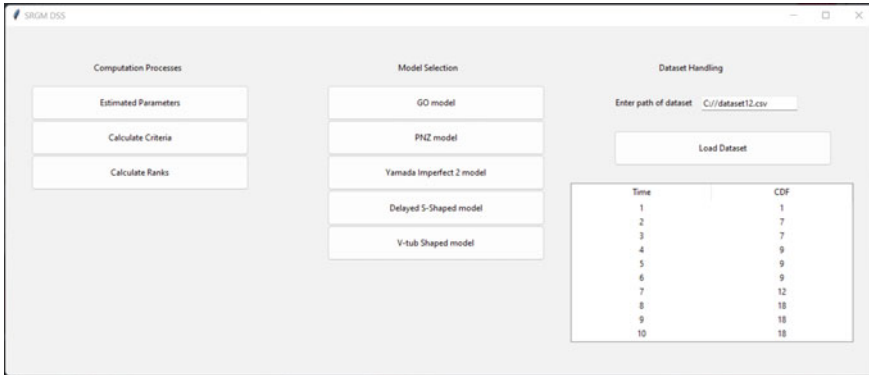


Fig. 5 Main screen of the DSS after dataset is loaded

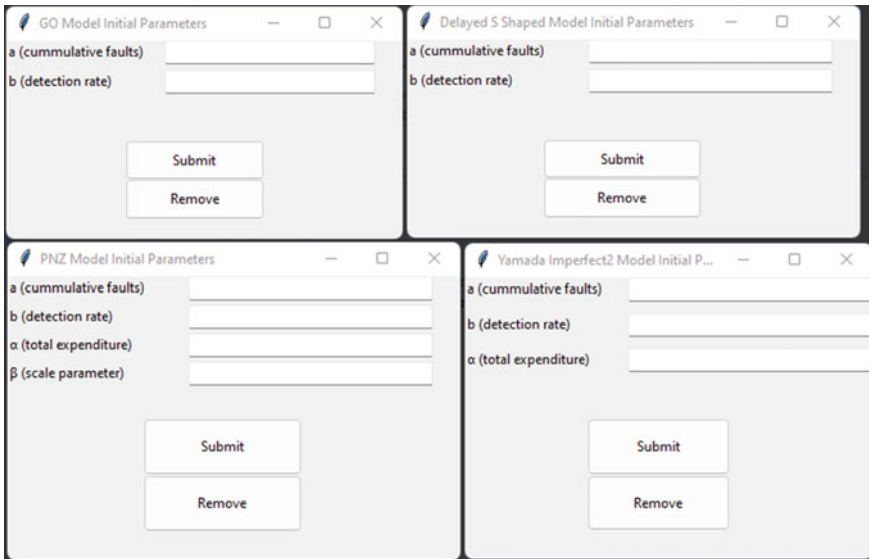


Fig. 6 Windows for input of initial model parameters

of the respective model, the DSS executes the estimation process for that model, and keeps a record of the estimated optimal parameters for each model that is selected.

The DSS also provides an option to remove a model, on clicking the Remove button of the model that is to be removed, and the model is removed from the ranking list. The record of the model's parameters is also dropped. When the model is successfully removed, the DSS prompts the message shown in Fig. 7.

Estimated model parameters can be viewed by clicking on the Estimated Parameters button as illustrated in Fig. 8. Apart from the optimal model parameters, the

Fig. 7 Message for removal of model from the ranking procedure

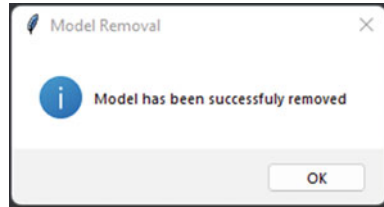
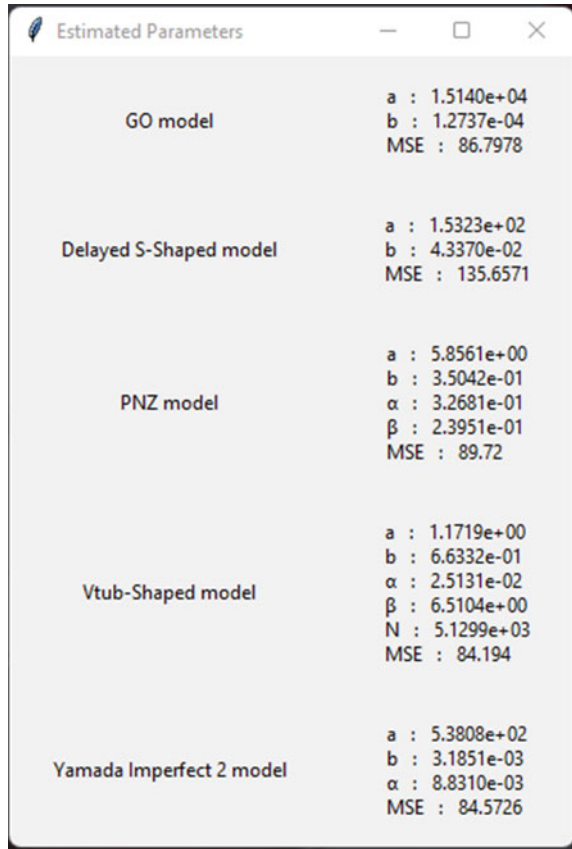
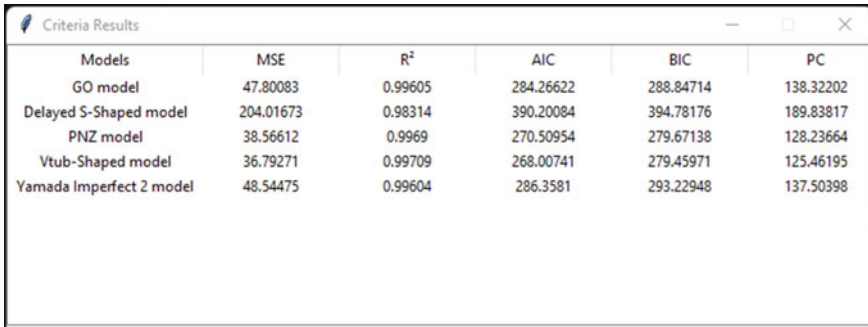


Fig. 8 Window for estimated model parameters



window also displays the lowest value of MSE attained at the optimal model parameters. This process of parameter estimation can again be carried out using different initial model parameters, and the DSS dynamically updates the record of the new optimal model parameters and the value of MSE that is attained.



Models	MSE	R ²	AIC	BIC	PC
GO model	47.80083	0.99605	284.26622	288.84714	138.32202
Delayed S-Shaped model	204.01673	0.98314	390.20084	394.78176	189.83817
PNZ model	38.56612	0.9969	270.50954	279.67138	128.23664
Vtub-Shaped model	36.79271	0.99709	268.00741	279.45971	125.46195
Yamada Imperfect 2 model	48.54475	0.99604	286.3581	293.22948	137.50398

Fig. 9 Window for criteria of models

After the model parameters are estimated, the ranking criteria can be viewed by clicking on the Calculate Criteria button. The DSS opens another window with the tabulated data as shown in Fig. 9. This is a clear depiction of the decision matrix that is created in order to rank the models using the ranking methodology.

The system keeps track of estimated model parameters and the models that are selected for ranking. When the Calculate Criteria button is clicked, a subroutine that takes the estimated model parameters as a parameter runs and calculates the performance criteria. Finally, to calculate rank of models, the Calculate Ranks button in the computation section is to be clicked.

On clicking the Calculate Ranks button, the DSS opens another window as shown in Fig. 10 which asks for choice of ranking methodology. Currently the choice only includes Entropy & TOPSIS.

On choosing the option and clicking on submit button, a new window with the ranks of the selected models is spawned. These are the ranks that are assigned to the models by the ranking algorithm according to the ranking methodology used in the DSS. This is shown in Fig. 11.

Fig. 10 Window for choice of ranking methodology

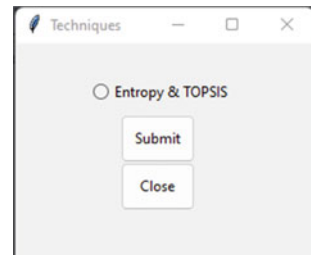


Fig. 11 Window for ranks of models

Models	Rank
GO model	3
Delayed S-Shaped model	5
PNZ model	2
Vtub-Shaped model	1
Yamada Imperfect 2 model	4

6 Result and Analysis

The DSS is tested over a dataset from Military Digital Control System [7]. We employ one real-world dependability data set (Data Set 1, DS-1) given by [33] to validate the usefulness of the proposed model in actual application. This data collection was gathered from a military command and control system through system testing and on-board testing. Table 1 shows the failure data in detail, with the day as the time unit. In total, 367 flaws were discovered in just 73 days. To fit the models and estimate the parameters for the descriptive power comparison, all data points are needed. The model parameters are estimated using the LSE approach in this case.

According to the ranking methodology, the decision matrix needs to be constructed. The alternatives and criteria that are required to create the decision matrix are displayed in the criteria window in Fig. 8. The decision matrix is also shown in the Table 2.

Once the decision matrix is created, next comes calculating the weight of each criterion. Using EWM which uses Shannon Entropy governed by Eq. (7),

$$\text{ShannonEntropy} = - \sum_i P(x_i) \log(P(x_i)) \tag{7}$$

the weights are calculated using the data of the decision matrix. The final weight vector is shown in Table 3.

The weight of the criteria determines how much priority each criterion is to be given while calculating the rank of the models using the ranking methodology.

The decision matrix needs to be normalized before the weight vector is used to get the weighted normalized matrix. After the normalization of the decision matrix, the weight vector needs to be multiplied to each row of the normalized decision matrix in order to get the weighted normalized decision matrix.

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \tag{8}$$

Table 1 Data set 1, DS-1

Day no	Faults	Cumulative faults	Day no	Faults	Cumulative faults	Day no	Faults	Cumulative faults	Day no	Faults	Cumulative faults
1	2	2	20	6	124	39	5	204	58	3	303
2	0	2	21	2	126	40	1	205	59	4	307
3	2	4	22	3	129	41	4	209	60	8	315
4	3	7	23	2	131	42	6	217	61	3	318
5	3	10	24	3	134	43	3	220	62	4	322
6	6	16	25	8	142	44	2	222	63	5	327
7	8	24	26	6	148	45	6	228	64	6	333
8	8	32	27	7	155	46	13	241	65	0	333
9	12	44	28	8	163	47	9	250	66	4	337
10	10	54	29	2	165	48	6	256	67	5	342
11	6	60	30	3	168	49	7	263	68	4	346
12	5	65	31	4	172	50	3	266	69	5	351
13	4	69	32	3	175	51	3	269	70	5	356
14	6	75	33	3	178	52	4	273	71	5	361
15	10	85	34	4	182	53	5	278	72	3	364
16	6	91	35	4	186	54	6	284	73	3	367
17	7	98	36	5	191	55	6	290			
18	10	108	37	4	195	56	5	295			
19	10	118	38	4	199	57	5	300			

Table 2 Decision matrix

	MSE	R ²	AIC	BIC	PC
GO model	47.80083	0.99605	284.26622	288.84714	138.32202
Delayed S-shaped model	204.01673	0.98314	390.20084	394.78176	189.83817
PNZ model	38.56612	0.9969	270.50954	279.67138	128.23664
Vtub-shaped model	36.79271	0.99709	268.00741	279.45971	125.46195
Yamada imperfect 2 model	48.54475	0.99604	286.3581	293.22948	137.50398

Table 3 Weight vector of criteria

MSE	R ²	AIC	BIC	PC
8.9729e-1	4.5287e-5	3.3807e-2	2.9998e-2	3.8853e-2

Table 4 Values for S_i^+ , S_i^- and C_i

	Go model	Delayed S-shaped model	PNZ model	Vtub-shaped model	Yamada imperfect 2 model
S_i^+	0.044609	0.677215	0.007189	0	0.047620
S_i^-	0.632620	0	0.670030	0.677215	0.629606
C_i	0.934130	0	0.989383	1	0.926829
Rank	3	5	2	1	4

According to TOPSIS, after the weighted normalized decision matrix is calculated, the Euclidean distance of each alternative from the Positive Ideal Solution (S_i^+) and Negative Ideal Solution (S_i^-), which determine the vectors S_i^+ and S_i^- respectively. Further S_i^+ and S_i^- are used to calculate the relative closeness vector C_i . This is governed by Eq. 8. Rank of the models are calculated using the value of C_i of each model, the higher the value, higher the rank. This is shown in Table 4.

7 Conclusion

Many researchers, software engineers, and practitioners have been bothered by the difficulty of picking optimal SRGMs, which has led to this field's evolution to the point where the time-consuming procedure of finding the rank for all SRGMs has been totally automated. The presented work explains how an efficient hybrid Entropy-TOPSIS approach is utilised to compute rankings for an interval-domain dataset, by evaluating 5 NHPP SRGMs for 5 performance indicators. We have used an interval-domain dataset to develop the DSS prototype. The TOPSIS approach has been used because over other MCDM methods, the Entropy-TOPSIS methodology has

numerous advantages, such as calculating the weights of performance measures, simplicity, and so on.

The unique feature of this research study is that it uses a prototype written in the Python programming language to automate the time-consuming and tedious job of parameter estimation, weight evaluations, and model rank computing. Furthermore, it is worth noting that no single model is universally applicable to all datasets. The selection of SRGMs is based on the context's elements as well as how they process data. As shown in Fig. 10, the DSS provides a ranking framework to choose the most optimal SRGM for the employed dataset, on the basis of the weights of the considered performance indexes, which have a significant impact on the ranking of the SRGMs. The testing is carried out using a dataset that has already been released. In terms of SRGM development and evaluation, the proposed DSS contains endless possibilities for future research with a provisional and exploratory essence.

References

1. Lyu MR (1995) Handbook of software reliability engineering. McGraw-Hill publishing. ISBN 0-07-039400-8. <http://portal.research.bell-labs.com/orgs/ssr/book/reliability/introduction.html>
2. ANSI/IEEE (1991) Standard glossary of software engineering terminology, STD-729-1991. ANSI/IEEE
3. Rook P (ed) Software reliability handbook. Centre for software reliability, City University, London, U.K
4. Okamura H, Dohi T, Osaki S (2001) A reliability assessment method for software products in operational phase-proposal of an accelerated life testing model. *Electron Commun Jpn* 84(8):25–33
5. Morita H, Tokuno K, Yamada S (2005) 'Markovian operational software reliability measurement based on accelerated life testing model. In: Proceedings of the 11th ISSAT international conference on reliability and quality in design, pp 204–208
6. Tokuno K, Yamada S (2007) User-oriented and perceived software availability measurement and assessment with environmental factors. *J Oper Res Soc Jpn* 50(4):444–462
7. Li Q, Pham H (2019) A generalized software reliability growth model with consideration of the uncertainty of operating environments. *IEEE Access* 7:84253–84267. <https://doi.org/10.1109/ACCESS.2019.2924084>
8. Goel AL, Okumoto K (1979) Time-dependent error-detection rate model for software reliability and other performance measures. *IEEE Trans Reliab* 28(3):206–211
9. Pham H, Nordmann L, Zhang Z (1999) A general imperfect-software debugging model with S-shaped fault detection rate. *IEEE Trans Rel* 48(2):169–175
10. Teng X, Pham H (2006) A new methodology for predicting software reliability in the random field environments. *IEEE Trans Reliab* 55(3):458–468
11. Yamada S, Tokuno K, Osaki S (1992) Imperfect debugging models with fault introduction rate for software reliability assessment. *Int J Syst Sci* 23(12):2241–2252
12. Pham H (2014) A new software reliability model with Vtub-shaped fault-detection rate and the uncertainty of operating environments. *Optimization* 63(10):1481–1490
13. Chang IH, Pham H, Lee SW, Song KY (2014) A testing-coverage software reliability model with the uncertainty of operating environments. *Int J Syst Sci Oper Logist* 4(1):220–227
14. Pham H, Zhang X (1997) An NHPP software reliability model and its comparison. *Int J Reliab Qual Saf Eng* 4(03):269–282

15. Zhang X, Teng X, Pham H (2003) Considering fault removal efficiency in software reliability assessment. *IEEE Trans Syst Man Cybern Part a Syst Hum* 33(1):114–120
16. Zhu M, Pham H (2020) A generalized multiple environmental factors software reliability model with stochastic fault detection process. *Annals Oper Res* 1–22
17. Ohba M (1984) Inflection S-shaped software reliability growth model. In: *Stochastic models in reliability theory*. Springer, Berlin, Heidelberg, pp 144–162
18. Roy P, Mahapatra GS, Dey KN (2014) An NHPP software reliability growth model with imperfect debugging and error generation. *Int J Reliab Qual Saf Eng* 21(02):1450008
19. Yamada S, Ohba M, Osaki S (1983) S-shaped reliability growth modeling for software error detection. *IEEE Trans Reliab* 32(5):475–484
20. Kumar V, Kapur PK, Taneja N, Sahni R (2017) On allocation of resources during testing phase incorporating flexible software reliability growth model with testing effort under dynamic environment. *Int J Oper Res* 30(4):523–539
21. Kumar V, Mathur P, Sahni R, Anand M (2016) Two-dimensional multi-release software reliability modelling for fault detection and fault correction processes. *Int J Reliab Qual Saf Eng* 23(03):1640002
22. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
23. Entropy (information theory), Wikipedia, the free encyclopaedia. [http://en.wikipedia.org/wiki/Entropy_\(information_theory\)](http://en.wikipedia.org/wiki/Entropy_(information_theory))
24. Kumar V, Ram M (eds) (2021) *Predictive analytics: modeling and optimization*. CRC Press
25. Hwang CL, Yoon K (1981) *Multiple attribute decision making: methods and applications*. Springer, New York
26. Keshavarz G, Zavadskas EK, Turuskis Z, Antucheviciene J (2016) A new combinative distance-based assessment (CODAS) method for multi-criteria decision-making. *Econ Comput Econ Cybern Stud Res* 50(3):25–44
27. Chang C (2010) A modified VIKOR method for multiple criteria analysis. *Environ Monit Assess* 168(1):339–344
28. Sharma K, Garg R, Nagpal CK, Garg RK (2010) Selection of optimal software reliability growth models using a distance-based approach. *IEEE Trans Rel* 59(2):266–276
29. Yang Z, Xu Q, Qiu X, Wang H (2008) An applied study on the method for supplier selection with PCA and ELECTRE. In: *IEEE international conference on service operations and logistics, and informatics*, pp 2151–2156. <https://doi.org/10.1109/SOLI.2008.4682890>
30. Saxena P, Kumar V, Ram M, A novel CRITIC-TOPSIS approach for optimal selection of software reliability growth model (SRGM). *Qual Reliab Eng Int*. <https://doi.org/10.1002/qre.3087>
31. Saxena P, Kumar V, Ram M (2021) Ranking of software reliability growth models: a entropy-ELECTRE hybrid approach. *Reliab: Theory Appl (SI 2 (64))*:95–113
32. Garg R, Raheja S, Garg RK (2022) Decision support system for optimal selection of software reliability growth models using a hybrid approach. *IEEE Trans Reliab* 71(1):149–161. <https://doi.org/10.1109/TR.2021.3104232>
33. Bao X, Lu M, Cong M (2000) *Software reliability measurement*. Software dependability research Center. Beihang University, Beijing, China, Tech. Rep. p 15

Devanshu Kumar Singh Amity University, Uttar Pradesh, Noida, India

Devanshu K. Singh is pursuing his Bachelor in Technology specializing in Artificial Intelligence from the Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India. He has co-authored articles at international conferences of distinguished repute and has been conducting research work in the field of software reliability, Artificial Neural Networks (ANNs), and optimization. His research interest includes machine learning, neural networks, and optimization.

Hitesh Amity University, Uttar Pradesh, Noida, India

Hitesh is pursuing his Bachelor in Technology specializing in Artificial Intelligence from Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India. He has been conducting research work for more than 2 years in the field of software reliability, optimization, and machine learning. He has a particular interest in the field of Artificial Neural Networks (ANNs).

Vijay Kumar Amity University, Uttar Pradesh, Noida, India

Dr. Vijay Kumar received his M.Sc. in Applied Mathematics and M.Phil. in Mathematics from the Indian Institute of Technology (IIT), Roorkee, India, in 1998 and 2000, respectively. He has completed his Ph.D. from the Department of Operational Research, University of Delhi. Currently, he is an Associate Professor in the Department of Mathematics, Amity Institute of Applied Sciences, Amity University Uttar Pradesh, Noida, India. He has co-authored one book and has published more than 55 research papers in the areas of software reliability, mathematical modeling, and optimization in international journals and conferences of high repute. His current research interests include software reliability growth modelling, optimal control theory, and marketing models in the context of innovation diffusion theory. He has reviewed many papers for *Soft Computing* (Springer), *IJRQSE*, *IJQRM*, *IJSAEM*, and other reputed journals. He has edited special issues of *IJAMS*, *CMES* (Taylor & Francis), and *RIO* journal. He is an editorial board member of *IJMEMS*. He is a life member of the Society for Reliability Engineering, Quality, and Operations Management (SREQOM).

Hoang Pham Rutgers University, New Jersey, USA

Dr. Hoang Pham is a Distinguished Professor and former Chairman (2007–2013) of the Department of Industrial and Systems Engineering at Rutgers University. His research areas include reliability modeling and prediction, software reliability, and statistical inference. He is editor-in-chief of the *International Journal of Reliability*, and editor of *Springer Series in Reliability Engineering*. Dr. Pham is the author or coauthor of 7 books and has published over 200 journal articles, 100 conference papers, and edited 17 books including *Springer Handbook in Engineering Statistics* and *Handbook in Reliability Engineering*. He is a Fellow of the IEEE, AAIA, and IISE.

Human Pose Estimation Using Artificial Intelligence



Himanshu Sharma, Anshul Tickoo, Avinash K. Shrivastava, and Umer Khan

Abstract Artificial intelligence is currently grabbing the attention of the developers, reason being that its vast applications. One of such applications is Human Pose Estimation. Estimating postures of human is amongst the trending topics in the research field nowadays. In this, the system is provided with the trained model and by the help of which the system can look for joints in body of the person standing in front of it. It has vast applications like it can be utilized for checking fitness of an athlete, to check if a person is doing exercise properly or not. In this paper we discuss modelling a gym tracker using artificial intelligence. Here we are counting repetitions of 4 exercises: squats, pushups, curls, pullups.

Keywords AI (Artificial intelligence) · Mediapipe · Pose estimation · Human body models

1 Introduction

Computer vision's most essential application is posture estimation. We may use it to track various portions of the human body, from which we can generate mechanical models on the computer. Human posture estimate, as the name suggests, is a technique for assessing the human body's stance. We can assess human posture using computer vision. Figure 1 shows the numerous human body endpoints.

It is a computer vision task that uses graphics to represent a person's orientation. This method is widely used to forecast the location of a person's body parts or joints. It's gaining a lot of attention as it has a wide range of applications that can benefit from it. In photos or videos, HPE (human pose estimate) recognizes and classifies the positions of human body parts and joints. It is used to represent human

H. Sharma · A. Tickoo · U. Khan
Amity University, Noida, Uttar Pradesh, India
e-mail: aticoo@amity.edu

A. K. Shrivastava (✉)
International Management Institute, Kolkata, West Bengal, India
e-mail: kavinash1987@gmail.com



Fig. 1 Illustration of human body joints [1]

body positions in 2D, and 3D space, a model-based method is utilized. It captures coordinates by identifying human body joints such as the wrist, shoulder, knees, eyes, ears, ankles, and arms, which is a vital point in images & videos that characterize a person's position. When an image or video is sent into the pose estimator model, it returns the coordinates of the identified joints & confidence score, indicating how accurate the estimations were. Pose estimation is a function that helps in predicting the location of a person or an item using computer vision. This is done by using the stance and orientation of a person or an item together. Human pose estimate uses key locations on the body to properly distinguish people's postures from a picture. These computations can be conducted in two or three dimensions.

Because of recent developments in machine-learning algorithms, machines can now comprehend human body language by performing position identification and stance tracking. The accuracy of these detections and the technology required to carry them out. Furthermore, the coronavirus pandemic is having a significant impact on the development of technology, where high-performance and real-time posture identification and tracking will lead to advancements in computer vision. It may then be utilized for social distancing by combining human position estimate and distance projection techniques. It assists people in maintaining physical separation from one another in a crowded work setting. HPE will have a tremendous influence on security, health, safety, and entertainment. One application where this technology has demonstrated its effectiveness is autonomous driving.

The basic step for determining human posture entails two steps:

- (i) Detecting critical points/joints in the human body
- (ii) Assembling those joints into a valid human position.

Pose estimation may be divided into two categories. Single Person and Multi-Person Pose Estimation because the number of individuals in an image are uncertain, a multi-person pose estimate is more complex than a single-person posture estimation. A straightforward way is to start with a person detector, then estimate the

components, and last calculate the poses for each person. Another strategy is to identify all portions in the image (i.e., components of everyone), then associate/group parts that belong to different people. The bottom-up strategy is the term for this procedure. There are two techniques that a computer vision system can use to gain an understanding of the data it is given. They are the Bottom-Up and Top-Down methods, respectively. The bottom-up technique includes utilizing collected knowledge to undertake an additional investigation into a random observation; finally, all the gathered knowledge leads to a solution or a broad comprehension of the whole observation object. Automatic number plate identification is an example of where a Bottom-Up technique is utilized in a computer vision application. This sort of software is a logical addition to a traffic speeding camera. Background knowledge is used in the top-down method to build an understanding from an observation. The background information serves as a reference point for selecting the best parameters for a model (like the approach of Deep Learning techniques). This method may be characterized as the process of breaking down a picture into sub-components and obtaining knowledge of the fragmented information to give a comprehension of the full image.

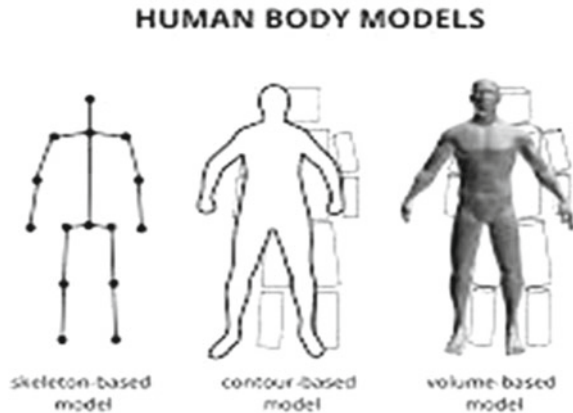
Tracking and measuring the human activity and movement is one of the most prominent dimensions applicable to pose estimation. Many architectures, including Open Pose, Pose Net, and Dense Pose, are used to recognize actions, gestures, and gaits. The following are some applications of human activity tracking:

- Personal gym trainers or AI-powered sports instructors.
- Detecting sitting motions.
- Disabled people can communicate via sign language.
- Traffic police officers make a detection signal.
- Signal detection by the cricket umpire.
- Detecting dance methods.
- Security and surveillance moves are being tracked.
- For retail businesses, crowd counting, and tracking.

1.1 Human Body Models

- Skeleton-based model [2]
It is composed of group of joints that constitute the skeletal structure of human body. This model can be utilized in both 2D & 3D human pose estimation techniques.
- Contour-based model [3]
It comprises of the contour and rough width of body torso and limbs. In this model, body parts are represented with rectangle and boundary.
- Volume-based model
It consists of a 3-D depiction of the human body and postures represented by volume-based models with geometric meshes and forms, which are often recorded via 3D scans (Fig. 2).

Fig. 2 Human body models
[4]



1.2 Risks and Errors in Fitness Apps Based on Human Pose Estimation

Anything that exists on the earth has certain number of risks associated with it. Same is the case with human pose estimation-based fitness apps. They are described below:

- **Men and Women Body Specifics**
While training human pose estimation models for any fitness app. While doing this, one needs to consider the structural difference of male and female bodies. If this is not done, then it will give inaccurate result either for male body or female body.
- **Physiology Specifics**
The body structure can be different even for two persons having same gender. So, it might be possible that it doesn't work for some persons.
- **Detection of the Exercise Start**
When estimating the training, we must recognize the start along with end of the training, so the fitness app must analyze the duration of the training directly. The error may occur when the arm angles briefly exceed the arbitrary threshold value.

1.3 Human Pose Estimation Metrics

There are basically two estimation metrics that we can use. They are as follows:

PCP (Percentage of Correct Parts): This statistic assesses the rate of limb detection, with a limb deemed identified if the discrepancy between the two anticipated joint positions and the genuine limb joint locations is less than half the limb length. It does, however, have limitations, such as penalizing shorter, more difficult-to-identify limbs.

PDJ stands for Percentage of Detected Joints. It is another joint detection-based metric & used to solve the drawbacks of the previous approach. If the gap between the expected and actual joint is less than a specific proportion of the torso diameter, the joint is considered detected. This fraction can be changed to generate detection.

1.4 2D Versus 3D Pose Estimation

- **2D Pose Estimation**
In this method, we simply estimate the positions of the body joints in 2D space in relation to the input data (i.e., image or video frame). Each key point's position is given by X and Y coordinate.
- **3D Pose Estimation**
This sort of pose estimation involves converting a 2D picture into a 3D object by adding a Z-dimension to the forecast. We can forecast the precise spatial placement of a depicted person or item using 3D pose estimation as it includes the complexity involved in creating datasets and algorithms that estimate multiple parameters, such as an images or videos background, light conditions, and so on.

1.5 Applications of Pose Estimation

- **Exam proctoring**
It is employed in all the exam proctoring tools that use webcam for proctoring the exam. By help of this, we can check, whether the candidate is looking at screen only or somewhere else.
- **Driver Drowsiness Detection**
Driving while feeling drowsy is very risky. We can check if the driver is feeling drowsy and take appropriate action.
- **Games**
Currently XBOX is using this technique. In this, what happens is that pose of person in front of the device is reflected as it is in the game.
- **Robotics**
Traditional robots depend on 2D vision systems, which have several drawbacks. A 3D pose estimation approach can be used instead of programming the robots manually to learn motions. This method results in robotics systems that are responsive, versatile, and realistic. It allows robots to understand motions and movements by monitoring the tutor's pose, appearance, or appearance (Figs. 3, 4 and 5).

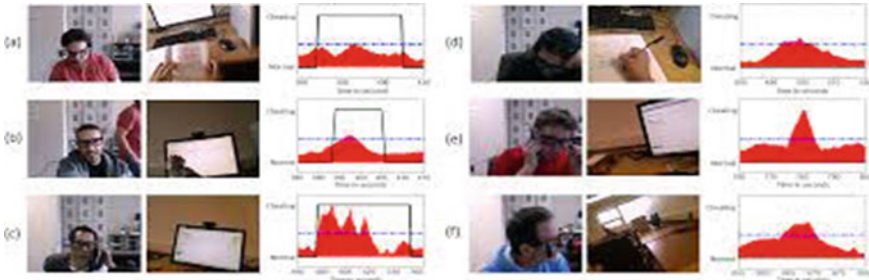


Fig. 3 Exam proctoring [5]

Fig. 4 Driver drowsiness detection [6]

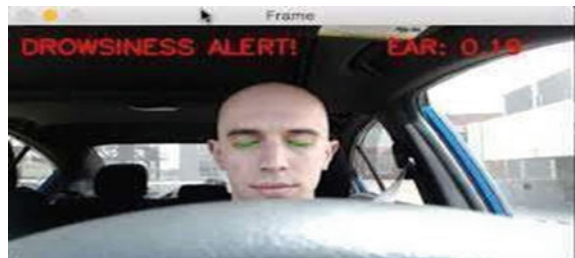


Fig. 5 Game using pose estimation [5]



1.6 Working of Pose Estimation

There are two approaches.

- Top-down
In this, first we segregate the humans in the image using object detection and then crop that and detect joints in the body.
- Bottom-up
In this, first we detect joints in the body, then we connect them all and detect various human bodies in the image (Figs. 6 and 7).

These are the two primary techniques that a computer vision system can use to gain an understanding of the data it is given. The bottom-up technique includes utilizing collected knowledge to undertake an added investigation into a random observation; finally, all the gathered knowledge leads to a solution or a broad comprehension of the

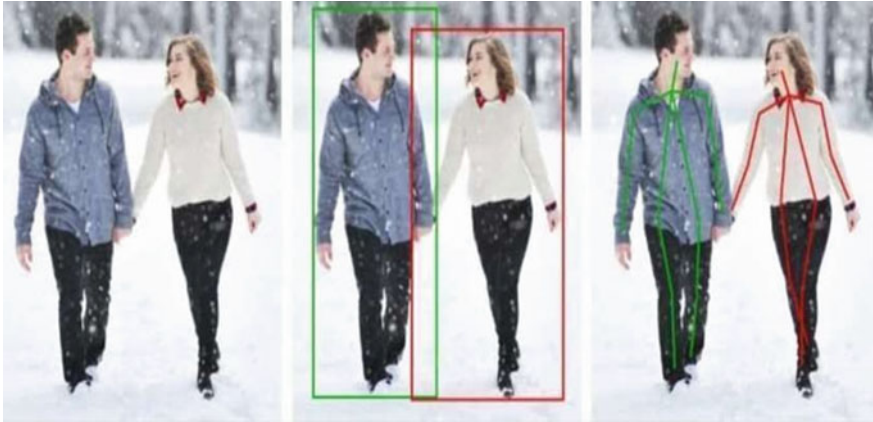


Fig. 6 Top down [7]



Fig. 7 Bottom up [7]

whole observation object. Automatic identification of number plates is an example of a Bottom-Up technique and is used in a computer vision application. This sort of software is a logical addition to a traffic speeding camera. Background knowledge is used in the top-down method to build an understanding from an observation. The background information serves as a reference for selecting parameters that are right for a model (like the approach of Deep Learning techniques). This method may be characterized as the process of breaking an image into sub-components and obtaining knowledge of the fragmented information to give a comprehension of the full image. Deep learning-based methods for human pose estimation are broadly categorized into single person pipelines where we are trying to get the pose of just a single person in

the frame and a multi-person pipeline where there could be multiple people in the frame, and we would like to know the pose of each of them.

1.7 Computer Vision

It deals with modeling and replicating human vision using computer software and hardware and hardware. It is a system that investigates how to re-construct, interrupt, and comprehend a 3D scene based on its 2D pictures in terms of the scene's structure. It is focused with employing computer software & technology to simulate and recreate human vision.

The computer had no idea about the contents of the image whether it contains the face of a person, a person you know, a famous personality whether it is the image of a tree, a mountain or some kind of an animal etc. So that was not very useful mainly because this allowed us to manipulate the data that lies beneath the images, but it does not allow us to provide any contextual information about the images. So, for instance if I wanted to search for an image out of the images that I have on my phone give me an image where there is a photo of me inside. It could be a selfie it could be any kind of image, but I should be present inside that image. Before computer vision our computers, our applications were not able to identify what image contains.

The part of the problem was that image data is very complicated, it may look simple on the outside but because images are made up of several colors, several variations of colors, gradients and they can be they can be manipulated in several ways it can be difficult to understand well what computer vision actually is. So, in cases like this when we're trying to understand what computer vision is, an important idea is to understand that our brains are much better and much more capable of understanding visual data, we do it all the time. When we look at a mobile phone we understand what the phone is, we can understand contextual information about the phone that is how large the screen is, what's the weight, how does it look, what color of the mobile phone is there, what company brand but the computers were not able to do so. With the advent of machine learning and data science now this is very much possible.

Now we can make computers learn what a mobile phone looks like and then when the mobile phone is seen in any of the images or videos, we can very easily understand what it actually means. This is a very narrow understanding of what computer vision is because in computer vision we might want to not just be able to do it with a mobile phone but there's several fields such as if we are trying to analyze the video for traffic we might want to know which of the images or which of the people in the traffic or vehicles in the traffic are moving at a pace that is not suitable for them. So that's what computer vision is basically it allows computers to understand visual data such as data contained inside an image. Computer vision is useful for several reasons. One of them is Contextual understanding-The greatest achievement of computer vision which is contextual understanding of images and visual data. For example, if an image contains a person that is or inside a vehicle that is in an image that a computer vision can understand and depending on how the

computer vision algorithm is set up it could even understand whether the person is driving or is just a passenger that's the contextual understanding. Before this all we had to do we had to process the image on our own and then we could label it on our own but there was no way of it being understood by the computers automatically. Computer vision was helpful in providing contextual understanding. Another good information would be that since contextual understanding is built using computer vision, we can even use it in different areas such as security where one person might be allowed to do some things but not the other things.

Picture classification recognizes an image & can categorize it (a dog, an apple, a person's face). More specifically, it can correctly forecast that a given picture belongs to a particular class. The social network firm, for example, would wish to utilize it to automatically recognize & separate problematic photographs shared by users. Picture classification may be used in object identification to identify a certain type of picture & then detect & tabulate its occurrence in a picture or video. Detecting defects on an assembly line, for example, or recognizing machines in need of two examples. When an item is discovered, object tracking follows or tracks it. This operation is frequently carried out using photos shot in sequence or real-time video streams.

Computer vision is applied to browse, search, & retrieve photographs from massive data warehouses based on the content of the images rather than the metadata tags connected with them. This task can include automatic picture annotation, which can be used in place of manual image labelling. These activities can be used in digital asset management systems to improve search & retrieval accuracy.

The greatest achievement of computer vision is the contextual understanding of images and visual data. Accurate predictions before computer vision predicting which data in visual data was very tiresome very cumbersome and was not very accurate as well, we got around the accuracy of 60% [5] but that was not desirable. So, to get over that hurdle computer vision makes processing videos, live streaming, data images and segregating them classifying them, or making some prediction based on that data very easy, and understanding the images as well becomes very easy and very accurate so accurate that it's now feasible to be used in the real-world applications. Computer vision has significantly reduced the cost of understanding the images and their contents which has mainly happened. You can deploy these applications on several platforms be it a low powered mobile phone be it a high-powered mainframe computer region application that can run on several applications or several platforms so the cost that we sort of invest in creating a computer vision system gets paid overtime because of the wide variety of platforms and the wide number of users that it can be accessed by. When comes to user experience there are several things that a computer vision system does to enhance user experiences. For instance, a very useful use case of computer vision is we have an image or we have some product in front of you that we want to get some information about now before this what we would have to do is we would have to use a textual interface in which we would have to type in the text depending on the image that we have now on the other hand what we could do is now use your camera to point at the product that we want to buy and the computer vision system will take it over from there and will give us all the

information that it can find on the internet based on the product that it has, that's the real advantage of using computer vision. It enhances the user experience to a level that was not possible before.

It shares some similarities with the following fields:

Picture Processing: It deals with image alteration.

Pattern recognition: It deals with numerous strategies for classifying patterns

Photogrammetry: It is the mechanism of acquiring precise measurements from photographs. The way computer vision technology works is similar to human brain functionality. However, how does our brain deal with visual object recognition? One prominent theory holds that our brains use patterns to decode particular items. This notion is employed in the development of computer vision systems.

Pattern recognition is the foundation of today's computer vision algorithms. We train machines on vast amounts of visual data—computers analyze photos, name items on them, & look for patterns in those images.

My Moments for 2018 Master golf event was made using IBM's computer vision. IBM Watson was able to recognize the sights (and sounds) of important shots after reviewing hundreds of hours of Masters film. It compiled crucial moments and sent them to supporters in the form of personalized highlight clips. Users may use Google Translate to aim their smartphone camera at a sign in another language and rapidly receive a translation of the sign in their favorite language. The development of vehicles that drive themselves is reliant on computer vision to interpret visual input from cameras and other sensors in a vehicle. IBM along with its partners such as Verizon to deliver sophisticated AI to the edge & to assist automobile makers in identifying quality flaws before a vehicle leaves the plant.

1.8 Artificial Intelligence

According to the Artificial Intelligence computers can imitate humans in the many ways-by Thinking, Learning, Planning and Understanding. It is a technology that helps the machine to mimic human behavior. It is an ability that the machine will think, behave, act, and perform some tasks like humans. Intelligence means the ability to understand something. It is an area of science that is helping machines in finding human-like answers to hard problems. This typically involves taking features from intelligence of humans and converting them into computer-friendly algorithms. According to the needs, an efficient strategy might be used, which determines how artificial the intelligent behavior seems. AI is a broad phrase that refers to all computer programs that can think in the same manner as humans. AI is a computer program that exhibits features like self-improvement, inference learning, or even fundamental human abilities like recognition of image and language processing. The subfields are machine learning and deep learning. Deep Learning comes under machine learning that use more advanced approaches to solve challenging issues. However, there are

Fig. 8 Artificial intelligence [4]



still some differences that is made between machine learning and artificial intelligence. Deep learning is deterministic, whereas machine learning is probabilistic. It is the field of computer science with virtue of which we have power to make intelligent machines that behave, think and make decision like human. You can create a machine with programmed algorithm which can work with its own intelligence (Fig. 8).

1.9 Project Description

Artificial intelligence has transformed the gym and fitness industry with smart workout ideas and body tracking. The saying goes, “Healthiest is richest!” However, only 20–30% of the world’s population maintains healthy behaviors, with the other 70% doing little physical exercise other than walking [8]. Adults, on the other hand, are unfit and acquire health problems at a young age. Individuals, on the other hand, have recently prioritized their health and well-being. In this project, we will be counting repetitions for 4 exercises: squats, curls, pushups and pullups. The person will select the exercise which he wants to do and then accordingly repetitions.

1.10 Benefits of Our Gym Tracker

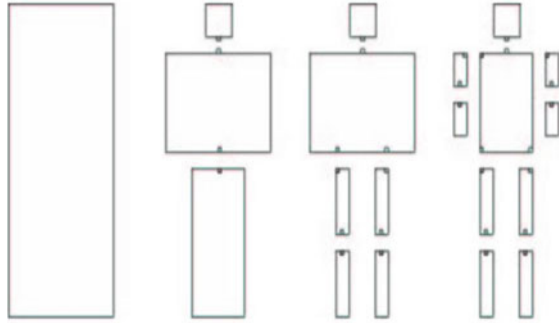
Person will get to know about real time count of exercise to know how many successful repetitions of exercise he has successfully done. Accordingly, he can accordingly correct himself.

Literature Review

2.1 Deep Neural Based Approach (Toshev et al.)

Author proposed a way for estimation of human pose by employing Deep Neural Network (DNN) [9]. In this technique, the posture estimation is found by applying

Fig. 9 Multi scale template model [10]



regression on body joints, The merit of this approach is that it is simple as well as powerful.

2.2 Estimation Using Multi-scale Template Model

In this technique researchers employed different human models, *one having 10 limbs, one having 6 limbs, one having 3 limbs and last one having only limb representing entire body as a whole*. Primary goal of this model was to provide richer description of limbs [10]. However, it doesn't prove effective as arms were usually merged in torso, which made them non-distinguishable (Fig. 9).

2.3 Generative Approaches

In year 1980, O'Rourke & Badler proposed a top-down approach where they assumed human frame as a tree-like connected model [11]. Any other approach with the aid of Fischler & Elschlager, that is known as the pictorial systems, in which they proposed a deformable components model in which the advent of every component is modeled individually and pairwise relationships are modeled by help of spring-like connections between pairs of parts.

2.4 Human Pose Estimation Features with Convolutional Networks

The authors presented a novel architecture for assessing human postures by combining a multilayer convolution network with a altered learning approach [12].

Comprehensive human posture assessment is amongst the most difficult challenges in computer vision, & their novel architecture and learning scheme demonstrated a huge enhancement over existing findings. Their primary contribution was demonstrating, for the first time, that a certain kind of deep learning can outperform all current classical architectures in this test. Higher spatial models improve the overall outcome of a work, although to a far lower extent than expected. Numerous specialists recently contended that kinematic construction and hierarchical data are basic for this space, yet with our unadulterated base up and frail spatial model, we could further develop other more muddled structures that are presently getting the best outcomes.

2.5 Deep Learning Framework Using Motion Features for Human Pose Estimation

In this dissertation, the researchers proposed a novel & effective method for the articulated estimation of human pose in videos that uses convolutional network [13], which includes color and movement characteristics. They proposed a new dataset on the posture of the human body.

2.6 Movenet

Move Net comprises of 3 modules: encoder, mapper, & decoder. The encoder module is prepared to change over the knee joint point profile into its 6-D inert portrayal [14]. The mapper module is prepared to plan the inert portrayal of point profile at zero degrees point of tendency (AOI) to idle portrayal of point profile at the objective AOI. The decoder module is prepared to foresee a point profile from its given idle portrayal.

2.7 Human Pose Estimation Via DNN

DNN-based posture estimation technique was presented. The estimate of posture is viewed as a DNN-based regressor towards body joints [15]. A class of such regressors was developed, resulting in high accuracy amongst present techniques. The technique has merit of thinking about present in an all-encompassing design and has a simple yet wonderful strategy that benefits from recent developments in Deep Learning. A bunch of regressors based on DNN can capture context and reason about posture in the holistic way. As a consequence, we can get superior results on numerous demanding academic datasets. Furthermore, it was demonstrated that a

general convolutional neural network, designed originally for classification tasks, may be used for localization.

2.8 TFPOSE

The researchers suggested a pose estimation technique that solves the problem using regression. Unlike earlier regression-based approaches, which frequently lag behind state-of-the-art methods, they structured the pose estimation issue as a sequence prediction problem that transformers can successfully tackle [4]. Their methodology is straightforward and straightforward, avoiding the disadvantages of heatmap-based pose assessment. Furthermore, by exploiting the attention process of tensor flow, the suggested framework can adaptively look for the features most relevant to the target key points, resolving the feature misalignment issue that plagued earlier regression-based techniques and significantly improving performance.

2.9 POSEAUG

Due to the limited 2D-3D posture pairings in training dataset, existing 3D pose estimators perform poorly on new datasets [8]. To solve this issue, the researchers developed PoseAug, an unique auto-augmentation method that learns to supplement the existing training poses with greater variation, hence boosting generalization of trained 2D pose into 3D pose estimator. PoseAug, in particular, is game changing. It learns to modify numerous geometrical elements of a pose (for example, posture, body size, viewpoint, and location) via differentiable operations. With such distinguishing features, the augmentor may be tweaked concurrently with the 3D pose estimator and use the estimation error as feedback to generate increasingly varied & complex poses in real time. PoseAug also built a differentiable module to assure the plausibility of enhanced postures and exhibited a unique methodology for analyzing local joint-angle likelihood. PoseAug's superior designs allowed it to create more diverse yet realistic postures than previous offline augmentation systems, resulting in greater pose estimator generalization. PoseAug is a general-purpose pose estimator that may be used to a variety of 3D pose estimators. Extensive testing shows that PoseAug excels on both intra & cross-scenario datasets. Notably, it obtains 88.6% accuracy on the MPI-INF-3DHP after cross-dataset assessment, outperforming the previous best data augmentation-based solution by 9.1%.

2.10 EMPOSE

Fully immersive AR/VR experiences rely on recreating the user's whole-body posture without limiting their movements [1]. The authors investigated the application of body-worn electromagnetic (EM) field-based sensing for the challenge of 3D human posture reconstruction in this work. They developed a technique for estimating SMPL values from 6 to 12 EM sensors to that end. They used a proprietary wearable system comprised of wireless sensors that measured time-synchronized 6D postures at 120 Hz. They used a recently published hybrid framework, learning gradient descent (LGD), to iteratively estimate SMPL position and shape from our input data in order to offer accurate poses even with minimum user instrumentation. This enables us to use powerful pose priors to deal with the quirks of the input data and obtain reliable position predictions. The proposed technique employs AMASS to generate virtual EM-sensor data, and we demonstrate that it generalizes effectively to a recently recorded actual dataset containing 36 min of motion from 5 patients. We obtain reconstruction errors of 31.8 mm and 13.3°, beating both pure learning and pure optimization approaches.

2.11 DSPNET

Existing human posture estimate strategies often have a large processing burden, which is disadvantageous for equipment with low resources [16]. To overcome this issue, the authors suggested DSPNet, a lower computational cost-based network. To begin, instead of using transposed convolution as a decoder for the network, they created a lightweight up-sampling unit. Reduced computation has resulted in an increase in prediction accuracy. Second, they introduced a unique deep supervision pyramid design to boost MSRA Simple Baseline's multi-scale acquiring capabilities while not increasing the number of parameters. DSPNet achieves almost similar state-of-the-art results with a minimal computing burden, as demonstrated by experimental results on both the MPII and COCO posture estimation benchmarks.

2.12 ADAFUSE

Occlusion is most probably the toughest obstacle for pose estimation outside [5]. To identify occluded joints, typical methods frequently rely on invasive sensors like IMUs. AdaFuse is the adaptive fusion strategy that may enhance the characteristics in blocked views by exploiting them in visible scenes, to making the job truly unconstrained. Its fundamental function is to find the point-to-point co-relation between 2 views, which we do well by examining the sparsity of the heatmap representation. They also made an adaptive fusion weight for each camera view to represent

its feature quality, reducing the possibility that good features would be unintentionally disturbed by poor views. They thoroughly tested the method on three public datasets: Human3.6 M, Total Capture, and CMU Panoptic. On all of them, it outperformed and proved its mettle. They also produced the Occlusion-Person big scale synthetic dataset, which allowed them to do numerical assessment on the occluded joints because it gives occlusion labels for every endpoint in the photos.

2.13 PIFPAF

The authors suggested the unique bottom-up strategy for multi-body posture estimation, which is especially well capable for urban mobility applications such as self-driving automobiles and delivery robots [6]. PifPaf is a novel approach that employs PIF to locate body parts & PAF to connect body parts for construction of entire human postures. Because of their novel field PAF containing fine-grained input & the use of Laplace loss for regressions that integrates a sense of uncertainty, this technique outperformed earlier strategies at lower resolution & in crowded, cluttered, & obstructed environments. Their concept is built on a completely convolutional, single-shot, box-free architecture. On the regular COCO key point problem, they performed equally good as compared to the existing bottom-up technique.

2.14 HigherHRNet

Strategies for estimating body pose via bottom-up strategy have difficulty estimating the right position for tiny people owing to scale variation problems [17]. HigherHRNet is a revolutionary bottom-up body pose estimation strategy for learning representations that are scale-aware using feature pyramids with higher resolution that we propose in this research. The proposed strategy, which uses multi-resolution tactic for training & multi-resolution aggregation for inference, may solve the scale variation challenge in bottom-up multi-body posture evaluation & more precisely find key points, especially for tiny persons. HigherHRNet's feature pyramid is composed of HRNet feature map outputs as well as sampled higher resolution outputs using transposed convolution. It proved its mettle by working great on COCO dataset for medium test-dev by 2.5% AP for medium people, demonstrating its efficacy in dealing with scale variance.

2.15 Sim2Real

While bypassing ethical concerns about privacy and prejudice, synthetic visual data can give almost endless diversity & rich labelling [18] However, existing models that

are trained through synthetic data show the poor generalization to realm datasets for many tasks. The function of 3D pose estimation is an especially intriguing example of this sim2real challenge because, while learning-based systems perform admirably given real-world training data, labelled 3D pose datasets are incredibly difficult to get, which restricts scalability. Researchers demonstrated in this publication that traditional neural-network techniques that perform badly when trained on synthetic RGB photos may perform well when the data is pre-processed to extract clues about the person’s mobility, namely optical flow and the motion of 2D key points.

2.16 ContextPose

Evaluating 3D pose from just 1 photograph is difficult due to the possibility of numerous 3D joint configurations having the same 2D projection [19]. To eliminate ambiguity, latest strategies frequently depend on context modelling methods like PSM or GNN. There has been no proper study that examines and differentiates them. As a result, the authors initially developed a generic formula for context modelling, with PSM and GNN as particular examples. They discovered that the end-to-end training scheme in GNN and the limb length limitations in PSM are two complementing elements for improving outcomes by comparing the two approaches. To combine these benefits, they devised ContextPose, a DNN-based attention technique for enforcing soft limb length limits. On 2 benchmark datasets, the technique effectively lowers the possibility of producing ludicrous 3D pose estimations with wrong limb lengths and proved its mettle. More crucially, introducing limb length limitations into deep networks allows the technique to achieve much improved generalization performance.

2.17 CVFusion

The authors demonstrated a method for recovering absolute 3D body postures from multi-view photos using multi-view geometric priors in our model [7]. It comprises of 2 distinct steps: predicting 2D postures in multi vision photos, and recovering 3D postures from multi-view 2D poses. First, they included a cross-view fusion technique into CNN to jointly evaluate 2D postures from various perspectives. As a result, each view’s 2D posture estimate already benefited from another view. Second, they demonstrated the recursive Pictorial Structural Model for recovering 3D postures from multi-view 2D poses. It steadily increases 3D posture accuracy at a low computational cost. We put our strategy to the test on two available datasets, H36M and Total Capture.

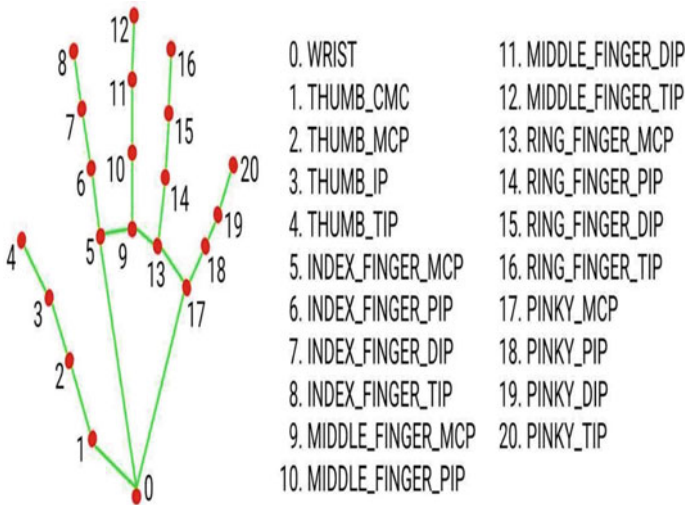


Fig. 10 Hand landmark [11]

3 Methodology

3.1 Mediapipe Based Approach

Mediapipe is a library that is developed by google and offers customizable ML solutions for live media like selfie segmentation, hand tracking, hair segmentation, etc. The basic approach is to find the landmarks of the body of the person. It is helpful for finding landmarks from a single frame. On the contrary, the current approaches rely primarily on powerful desktop environments. The figure below describes the landmarks of hands and face landmarks detected by mediapipe (Figs. 10, 11, 12 and 13).

Similarly, we have face landmarks detected in the image given above.

We have following 32 landmarks of body detectable by mediapipe (Figs. 14, 15 and 16).

Above images show the landmarks of body that can be detected using mediapipe.

3.2 Calculating Angles

After finding the landmark, our next aim is to calculate the angle between these landmarks & on the basis of that judge, whether a pose is correct or not.

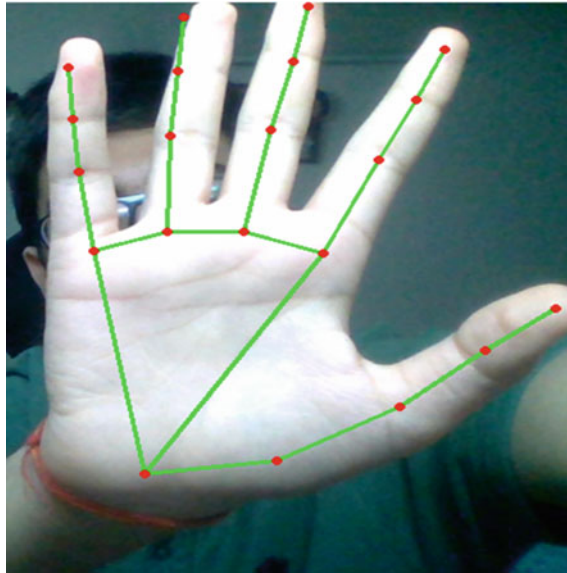


Fig. 11 Hand landmarks live detection

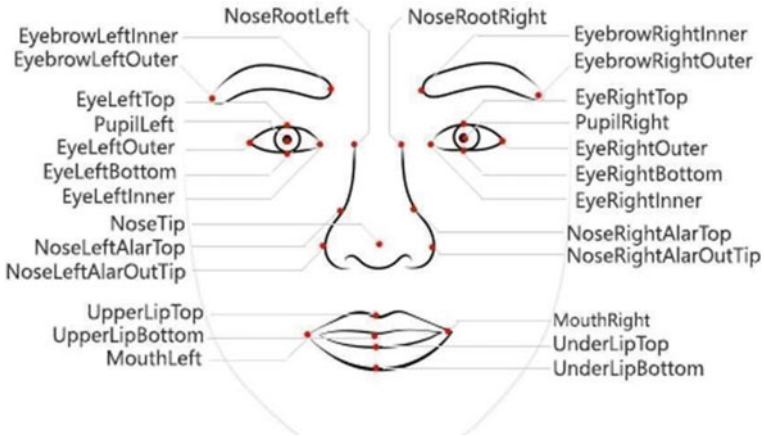


Fig. 12 Face landmarks

3.3 Calculating Repetitions for Curls

For curl we found out the coordinates for the left shoulder, left arm and left wrist. Then we calculated the angle between them with the help of numpy. Now, if this angle is greater than 160, then it is considered that it is start of new rep and if angle

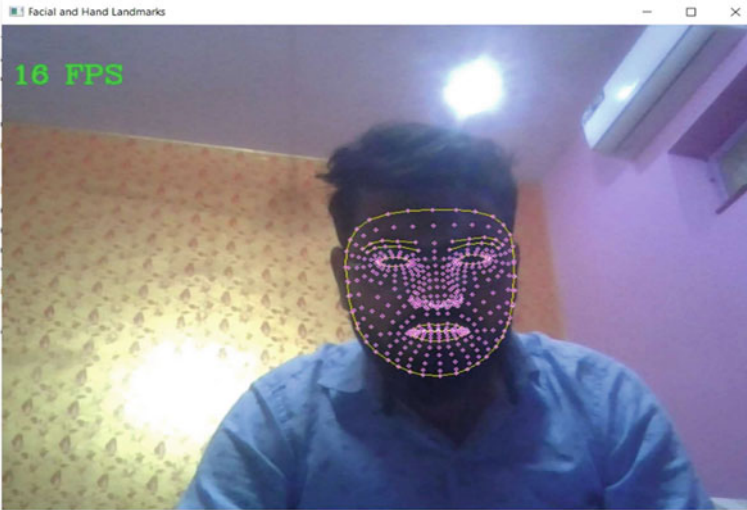


Fig. 13 Face landmarks live detection

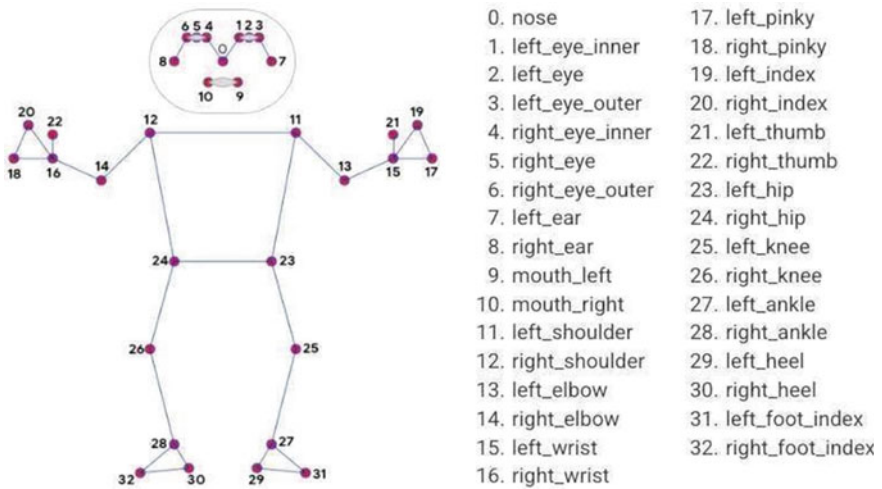


Fig. 14 Body landmarks [4]

becomes less than 30 then it is considered to be end of that particular rep and count is increased by one.



Fig. 15 Body landmarks live detection

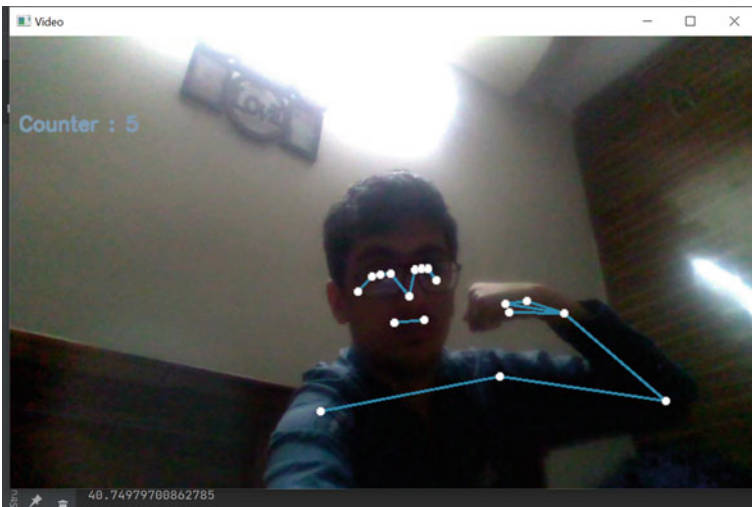


Fig. 16 Curls count live detection

3.4 Calculating Repetitions for Pushups

For Push-Up, we found out the coordinates for left shoulder, left arm and left wrist, right shoulder, right arm and right wrist. Then we calculated the angle between left

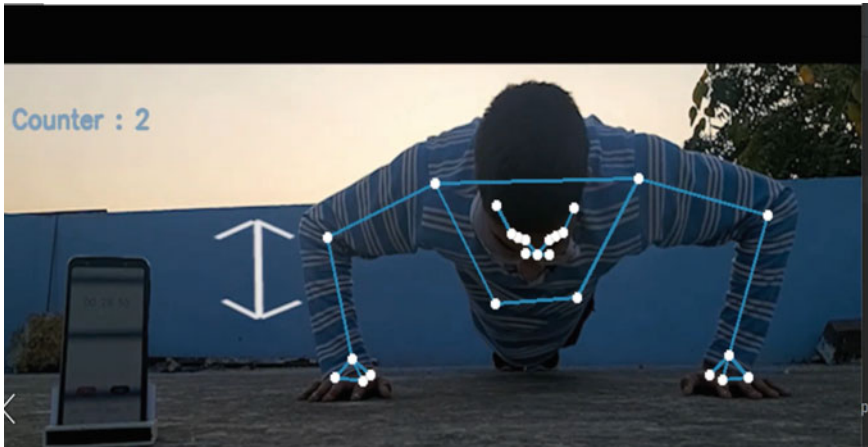


Fig. 17 Pushups count live detection

shoulder, left arm and left wrist with the help of numpy. Similarly, we found angle between right shoulder, right arm and right wrist. Next, we took an average of both these exercises. Now, if this angle is greater than 160, then it is considered that it is start of new rep and if angle becomes less than 70 then it is considered to be end of that particular rep and count is increased by one (Fig. 17).

3.5 Calculating Repetitions for Pullups

For the pullups, we have taken nose's y coordinate into consideration.

If this value becomes less than average y value for shoulders, then it is considered that exercise has started and after that when the nose's y coordinate becomes greater than average y value of shoulder, then repetitions are increased by 1 (Fig. 18).

3.6 Calculating Repetitions for Squats

For squats we found coordinates for left hip, left knee, and left ankle, right hip, right knee and right ankle.

Then, we found the angle between left hip, left knee and left ankle using numpy. Similarly, we found angle between right hip, right knee and right ankle. Then we took average of both these angles. Next, we took an average of both these exercises. Now, if this angle is greater than 160, then it is considered that it is start of new rep and if angle becomes less than 70 then it is considered to be end of that particular rep and count is increased by one (Fig. 19).



Fig. 18 Pullups count [16]



Fig. 19 Squats count [5]

Table 1 Accuracy of exercises

Exercise	Accuracy (%)
Curl	95
Squats	97
Pushups	94
Pullups	95

4 Result

Currently, if we look at play store, we have apps that maintain repetitions count on the basis of data entered by user. It shows that on this particular day, this much exercise has been done by users. But, they cannot count the repetitions with help of AI. Also, we have four exercises in a single application. Using mediapipe library, we have been able to count repetitions of exercises with great accuracy. For the model to work effectively, there need to be a computer with webcam enabled. For finding accuracy we ran our model on about 30–50 videos and checked for how many videos the count was correct. We did this for all the exercises. The accuracy for each exercise is mentioned Table 1.

5 Conclusion

Artificial intelligence technology has become crucial in numbers of businesses, including the fitness industry. Human posture estimation is becoming highly prominent in field of computer science. It can be done with Artificial Intelligence or Machine Learning, in which the system is fed sample data or trained models and can therefore find joints in the human body over video or image. Human posture estimation is currently one of the hot topics amongst researchers. We can utilize this in various ways like drowsiness detection, gym training, etc. We have worked upon gym training aspect of human pose estimation. We leveraged MediaPipe & Python for detecting different poses from a webcam feed. Then rendered the results to the screen using OpenCV. This pose estimation can be employed to track athletes, recognize their activities, count repetitions, and examine the performance of curls, pushups, pullups, and squats. For building an AI Gym tracker, a curl, squats, and pushup counter logic is implemented. We intend to expand the number of activities available in the future, as well as create a multi-person gym tracker that might be more useful for detecting exercises of multiple people in one go. We can also extend it to make it a multipurpose application. We can include drowsiness detection, sign recognition in this to extend the usability.

References

1. Kaufmann M, Zhao Y, Tang C, Tao L, Twigg C, Song J et al (2021) EM-POSE: 3D human pose estimation from sparse electromagnetic trackers. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 11510–11520
2. Lovato C, Castellani U, Giachetti A (2009) Automatic segmentation of scanned human body using curve skeleton analysis. In: International conference on computer vision/computer graphics collaboration techniques and applications. Springer, Berlin, Heidelberg, pp 34–45
3. Poppe R (2007) Vision-based human motion analysis: an overview. *Comput Vis Image Underst* 108(1–2):4–18
4. Mao W, Ge Y, Shen C, Tian Z, Wang X, Wang Z (2021) Tfpote: direct human pose estimation with transformers. [arXiv:2103.15320](https://arxiv.org/abs/2103.15320)
5. Hang Z, Wang C, Qiu W, Qin W, Zeng W (2021) Adafuse: adaptive multiview fusion for accurate human pose estimation in the wild. *Int J Comput Vision* 129(3):703–718
6. Kreiss S, Bertoni L, Alahi A (2019) Pifpaf: composite fields for human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 11977–11986
7. Qiu H, Wang C, Wang J, Wang N, Zeng W (2019) Cross view fusion for 3d human pose estimation. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 4342–4351
8. Gong K, Zhang J, Feng J (2021) Poseaug: a differentiable pose augmentation framework for 3d human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8575–8584
9. Toshev A, Szegedy C (2014) Deeppose: human pose estimation via deep neural networks. In: 2014 IEEE conference on computer vision and pattern recognition, Columbus, OH, pp 1653–1660. <https://doi.org/10.1109/CVPR.2014.214>
10. Human Body Pose Estimation Fernando Flores-Mangas University of Toronto
11. Zheng C, Wu W, Yang T, Zhu S, Chen C, Deep learning-based human pose estimation: a survey. IEEE, Ruixu Liu, JuShen, Senior Member, IEEE, Nasser Kehtarnavaz Fellow, IEEE and Mubarak Shah, Fellow, IEEE
12. Jain A, Tompson J, Andriluka M, Taylor GW, Bregler C (2013) Learning human pose estimation features with convolutional networks. [arXiv:1312.7302](https://arxiv.org/abs/1312.7302)
13. Jain A, Tompson J, LeCun Y, Bregler C (2014) Modeep: A deep learning framework using motion features for human pose estimation. In Asian conference on computer vision. Springer, Cham, pp 302–315
14. Bajpai R, Joshi D (2021) MoveNet: a deep neural network for joint profile prediction across variable walking speeds and slopes. *IEEE Trans Instrum Meas* 70:1–11
15. Roberts TJ, McKenna SJ, Ricketts IW (2004) Human pose estimation using learnt probabilistic region similarities and partial configurations. In: European conference on computer vision. Springer, Berlin, Heidelberg, pp 291–303
16. Zhong F, Li M, Zhang K, Hu J, Liu L (2021) DSPNet: a low computational-cost network for human pose estimation. *Neurocomputing* 423:327–335
17. Cheng B, Xiao B, Wang J, Shi H, Huang TS, Zhang L (2020) Higherhmet: scale-aware representation learning for bottom-up human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5386–5395
18. Doersch C, Zisserman A (2019) Sim2real transfer learning for 3d human pose estimation: motion to the rescue. *Adv Neural Inf Process Syst* 32
19. Ma X, Su J, Wang C, Ci H, Wang Y (2021) Context modeling in 3d human pose estimation: a unified perspective. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 6238–6247

Himanshu Sharma Amity University, Noida, Uttar Pradesh, India

Himanshu Sharma has received Engineering degree from Amity University, Noida. He is currently working in Hanu Software as Cloud Engineer.

Anshul Tickoo Amity University, Noida, Uttar Pradesh, India

Dr. Anshul Tickoo is an Associate Professor at the Amity School of Engineering and Technology, Amity University, Noida. She completed her Ph.D. in Computer Science Engineering from Amity University, Noida. She obtained her Ph.D. degree in Computer Science Engineering from Amity University in 2018. Her research interests include software reliability, soft computing and modeling, and optimization. She has been a member of Society for Reliability Engineering, Quality, and Operations Management (Regd.) since 2012. She has published work in the areas of optimization, soft computing, and software reliability.

Avinash K. Shrivastava International Management Institute, Kolkata, India

Dr. Avinash K. Shrivastava received his Ph.D. Degree from Department of Operational Research, University of Delhi. He is working as an Assistant Professor in the area of Management Information Systems and Analytics at International Management Institute-Kolkata (IMI-K). He has presented papers at conferences of international repute and won accolades for best paper presentations. He has been publishing extensively and serving as a reviewer for various journals of international repute. He is the book series editor of Advances in Emerging Markets and Business Operations published by Routledge, Taylor, and Francis and serves as managing editor of a reputed Journal "International Journal of System Assurance Engineering & Management" by Springer. He is also a life member of the Society for Reliability, Engineering, Quality, and Operations Management (SREQOM).

Umer Khan Amity University, Noida, Uttar Pradesh, India

Umer Khan has received an Engineering degree from Amity University, Noida. He is currently working in ACS as Software Engineer Trainee.

Neural Network Modeling and What-If Scenarios: Applications for Market Development Forecasting



Valentina Kuskova, Dmitry Zaytsev, Gregory Khvatsky, and Anna Sokol

Abstract In this chapter, we demonstrate the use of neural networks for forecasting automotive market sales. Using data from an entire country, with a large set of micro- and macroeconomic variables and other factors, we show how to supplement “black box” neural network calculations with a solid theoretical foundation from social sciences. Doing so allows to not only create exceptionally accurate forecasts, but understand the “black box” weights on different variables that are used in generating predictions. We also demonstrate how individual variables fit into the overall market dynamics, and how political changes play a role in economic outcomes.

1 Introduction

In this study, we outline the use of neural networks for prediction of an automotive market sales. The original goal of the study was to develop a sales forecast for new automobiles on the country level in a variety of time scenarios at the request of one of the major players in the Russian automotive industry. Short-term forecasts (up to three years) needed to be created on a monthly basis; longer-term (ten years and up)—on the annual basis. The original design assumed the use of neural networks as a forecasting method. However, in the early stages, the design was substantially expanded to include a solid theoretical base and resulted in robust methodology that

V. Kuskova

Lucy Family Institute for Data & Society, University of Notre Dame, Notre Dame, IN, USA
e-mail: vkuskova@nd.edu

D. Zaytsev (✉)

University of Notre Dame at Tantur, Jerusalem, Israel

G. Khvatsky · A. Sokol

Lucy Family Institute for Data & Society, Department of Computer Science and Engineering,
University of Notre Dame, Notre Dame, IN, USA
e-mail: gkhvatsk@nd.edu

A. Sokol

e-mail: asokol@nd.edu

allowed for exceptionally high forecast accuracy. This methodology also provides an opportunity to establish what-if scenarios: forecasting the results we can be expected if combinations of input factors or conditions changes.

It is important to understand that as a part of future studies [2, 11], the proposed methodology is subject to the general rules of forecasting. Namely, we cannot predict the future, we can only extrapolate the existing trends into a limited time horizon. This allows decision-makers to develop an early warning system, to understand the influence of factors that generate a certain result, and to design the operating system by accounting for the necessary factors. The proposed methodology is mostly targeted at forecasting market development but can also be expended to the prediction of social, economic, or political development of social entities (companies, industries, countries).

The main goals of market forecasting are to identify key factors which potentially influence the market dynamics; to define the impact/weight of factors underlying market sales; to develop a tool for modelling and forecasting (a group of simulation models). To accomplish this complex and complicated goal, the process is broken down into distinct, but interconnected steps.

The first step is the methodology development and revision. We account for scientific and business approaches to forecasting, theorizing the key contributing factors. The second step is data collection. At the start of the study, all types of data are collected: company-provided data; interviews with company management; additional quantitative data. The third step is data clean-up: converting data into one format; imputing missing data; extrapolating quarterly and annual data to monthly, where necessary.

The analysis phase also consists of several stages. Paramount to forecast success is the exploratory data analysis: graphical and numerical summaries and displays of data; correlation and factor analysis. Next in the order of importance is evaluating time-series data for trends, finding time-related patterns of all types, from linear to cyclical. Data were detrended for evaluating the influence of exogenous variables on the dependent variable of interest—monthly sales. Extracted trends were later added to the forecasts to account for time-dependent fluctuations. Next stage, the actual modeling, was executed using in-database machine learning using Vertica platform, followed by building the what-if scenarios based on simulation modeling. The final step was forecasting or identifying new data to collect.

As a result of the iterative process, we created unique forecasts not just for the overall market, but also for smaller segments based on car sizes and prices (e.g., budget, volume, and premium segments and subsegments within). Finally, at the request of the industry client, we created an interactive dashboard to use with the forecasting tool.

We start this paper with short dive into the history and theory of neural networks for time series analysis and market prediction, to motivate our choice of the final model. We then proceed with the description of data as proper selection of variables and the cleanup process were paramount to the model accuracy. We then describe the modeling approach. Next, we explore the theoretical foundation that was used to guide the exploratory data analysis and variable selection. In fact, these two

steps—modeling and theory-based variable selection—were performed iteratively multiple times, until the proper model convergence was achieved. In conclusion, we summarize the capacity of developed methodology based on neural network modeling enhanced by theoretical development to building forecasts.

2 Neural Networks

The present study heavily relies on the use of neural networks for time series analysis. This warrants a deeper dive into the history of how neural networks were used for similar tasks, as well as into the theoretical foundations of neural networks use.

The concept of an artificial neural network has a history almost as long as the history of computing itself, dating at least to the late 1950s [17]. However, the art and science of processing data using artificial neural networks has undergone a period of explosive growth in the 1990s and 2000s. This growth was in part fueled by the rapid advances in computing technology that has made more and more raw computational power available to the public. For example, per the Web of Science citation database, only 2459 articles related to neural networks (both natural and artificial) were published in 1990. In 2020, this number was 74,584, almost 1.5 orders of magnitude larger. The field of time series analysis with neural networks shows a similar pattern of exponential growth in the last 3 decades. Only 18 articles were published in 1990, and as much as 3401 in 2021.

At the same time, this simple quantitative analysis highlights the fact that the field of time series analysis with neural networks is considerably smaller than the general field of neural network research. This can suggest that there's still many advances to be made in this field. This conclusion is echoed in the literature [15], where the authors suggest multiple possible ways of improving the models used for time series analysis. Most notably, the authors highlight the need for further development in domains of multivariate data processing. Another suggested development that is crucial for this paper is developing models that perform equally well for short- and long-term time series analysis.

The purpose of this section of the paper is thus twofold. At first, we will try to provide a comprehensive overview of the use of neural network for time series analysis in general and for market research. Then we will look at how, on the theoretical level, these models can be applied to the task at hand.

Initial research into the field of time series analysis started with the use of relatively simple feed forward neural networks [19]. The architecture of these models is relatively simple, as they form a directed acyclic graph of neurons organized into layers, neurons on one layer connected only to the neurons on the next layer. Each neuron, in turn is represented as a function from multiple input to one output. This function consists of a linear combination of a neuron's inputs with certain weights and a nonlinear activation function. There are several functions that can be used for this purpose, and one of the most common is the logistic function:

$$f(x) = \frac{1}{1 + e^{-x}}$$

The process of training a neural network, is, in essence, a process of setting these weights in a way that minimizes the error made by the model. These models are commonly trained by using backpropagation, a technique that allows for efficient optimization of such models. Backpropagation is possible because the logistic function has a continuous derivative.

The main advantage of these models is their relative simplicity and ease of building and training. However, feedforward neural networks as described in [19] are fundamentally limited in the amount of data they can process at once, as they can only see a “window” of a fixed size. This makes these models unsuitable for almost all practical applications [6]. A one possible way to solve this problem of data availability is by use of a recurrent neural network. Recurrent neural networks are neural networks that allow feedback from their outputs back into their inputs. Since networks of this type have an innate ability to store information about the data they have seen in the past, they are uniquely suited for time series analysis. However, due to their recursive nature, networks of this type cannot be directly trained using backpropagation [6]. Another problem of classic recurrent neural network models is that even when training with specialized algorithms, the training process is unstable, and may lead to zero or infinite weights in neurons.

The way to solve this problem is to design a specialized neural network architecture. One such architecture is Long Short-Term Memory (LSTM), first proposed in 1997 [10]. LSTM achieves its purpose by combining a specialized training algorithm with an architecture that helps enforce the stability of the training process. The peephole LSTM, which has been shown to be effective in time-series analysis, is represented in figure [8].¹

The intuition behind LSTM models is that their ability to have cell state (and to forget it at will) allows them to retain a greater amount of information from previous iterations than classical recurrent neural networks. This, in turn, helps alleviate the vanishing and exploding gradient problems while at the same time allowing them to capture long-term dependencies in the data.

Because of their ability to remember and reconstruct both short- and long-term processes happening in temporal data, LSTM networks are uniquely suited for time-series analysis and forecasting. LSTM networks have demonstrated good performance when predicting time series that are difficult to learn using other methods. However, it has also been shown that if a time series can be predicted using a feedforward network with a window of a fixed size, then LSTM will in general perform worse than a fixed window based approach [7]. Gers et al. [7] also propose a hybrid approach to neural-network based prediction, where at first a fixed window-based model is trained to capture the general structure of the time series, and then use an LSTM model to reduce the residual error.

¹ Image used under the Creative Commons Attribution-Share Alike 4.0 International License.

Another approach for time-series analysis using recurrent neural networks suggested in the literature is the use of even more specialized training algorithms for recurrent neural networks. For example, a special boosting-based algorithm was designed to train a classic recurrent neural network to model and predict a chaotic time series [1].

In general, neural network-based models have better performance than models based on linear analysis, such as ARIMA or the Box-Jenkins model. For example, [9] provides a direct comparison between neural network prediction and Box-Jenkins ARIMA. The results of the study show that the proposed classic recurrent neural network had similar performance to Box-Jenkins ARIMA, a classical model used for time series analysis. It should be noted, however, that some literature [19] shows that a fixed-window model can perform better than a Box-Jenkins method. The same study also suggests that performance of neural networks-based models can be affected by many parameters among which is the nature of a particular time series. We can thus conclude that the field where neural networks are applied for time series analysis and prediction is of paramount importance.

Of a particular interest to this paper, is, of course, the field of market and economic analytics. Neural networks have a wide range of applications in that field, especially given the fact that market analytics and time series analysis are deeply interconnected.

Machine learning has a long history of applications for financial modeling and analytics. For example, in the early 2000s, a use of a support vector machine model was suggested for company performance analytics [5]. Support vector machine is a well-known machine learning model first proposed in the early 1960s, this method has been successfully applied in a variety of academic and commercial domains. Parallel to these developments, from the early 1990s onwards, the use of artificial neural networks for application in finance has started to grow. A study from 1992 [18] indicates that neural networks can be successfully applied for company performance and failure predictions.

In roughly the same time period, the use of neural network for financial market forecasting was growing in popularity. Kimoto et al. have tested applications of simple feedforward neural networks for stock prices prediction [14]. In their study, the neural network model has shown better performance than classical statistical models such as linear regression. It should be noted that the authors have pointed out that special preprocessing and normalization are required when training neural networks on highly irregular financial data. Another study [13] also points out the fact the neural networks can be susceptible to noise and trends in the data, as well as different variables being on different scales. The authors suggest using normalized data, that is then smoothed using a moving average, and filtered to create a more uniform distribution of the data. Another important part of building a well-performing model is selecting the right variables to train the model on.

For our project, taking the results from [7] into account, we decided to start the modeling process with a fixed-window based feed-forward network, the results from which have turned out to be satisfactory. Also taking into account the susceptibility of neural networks to certain types of noise in the data, we decided to employ an ensemble of a large number of individually trained neural networks. The neural

networks were both different in terms of their structure (i.e., number of nodes on different layers) as well as their input variables and the size of their windows. The results from multiple neural networks were then averaged to produce the final value of the forecast. This approach allowed us to both increase the overall accuracy of our forecast, and, at the same time, make the model more robust to noise in the data. This approach of randomized inputs also helped to prevent the model overfitting to the data, allowing it to generalize and be applicable even after sudden changes in the market.

3 Data

At the onset of the project, massive amounts of data were collected on a wide variety of variables that could be related to the one dependent variable—sales of new automobiles in Russia (measured as counts). Through the iterative steps of exploratory data analysis, many variables were eliminated as either irrelevant or redundant. The final dataset included more than one hundred and fifty independent variables or predictors.

The variables can be separated into four groups. The first group was consumer and dealer data, which included consumer sentiment, consumer's behavior, and the companies' KPIs related to consumer attraction and retention. Several variables from the last category were aggregated into a customer satisfaction index. Consumer sentiments were aggregated into consumer confidence index. Which reflected their inflationary expectations.

The second group of variables was marketing data which consisted of wider advertisement metrics ("above the line" marketing, or the ATL, consisting of measures of TV, radio, billboard, and similar advertising efforts); and incentives, dealer pricing, and other "below the line" (BTL) measures targeted at specific audiences. Advertisement and ATL variables also included indirect ads in numbers and minutes of advertising through cinema and TV. Incentives, dealer pricing, and other BTL were measured as the difference between retail price and transaction price.

The third group of data was consumer demographics (average family income, age, sex, occupation, family composition, etc.), consumer attributes (hobbies and interests, number of cars in the household, potential usage of cars, etc.) and automotive market data. Major correlates of the automotive market data consisted of gasoline prices and variables from related industries (such as prices of metals and rubber).

The fourth group of data reflected the overall Russian economy: macroeconomic, financial factors, and industry correlates (energy and precious metal prices). Key indicators of economic health were the balance of imports, exports, hard currency, and gold reserves. Financial factors were represented by the Russian Central Bank key rate and hard currency exchange rate against the Russian ruble.

All data were monthly series. Although all variables were included in the models, some had much more significant impact on sales than others. All data were analyzed in current time and with lags (influence of prior period variables on current period

sales). It was found that after 12 months, lagged effects were no longer significant, and final models contained only a few variables that were lagged at longer periods.

Of course, not all variables were equally important to forecast market development. Identification of key variables became a major part of the analysis, since approximately 20% of variables explained more than 70% in historical sales. Most variables retained in the final models were macroeconomic indicators such as consumer-related economic indicators (CPI, Oil, USD, Key rate, etc.); indicators of economic health (imports/exports, reserves (currency/gold), industrial production, etc.); and energy and metal prices (gasoline, natural gas, silver, gold, etc.). Consumer-related variables were found to have very little impact in the final models, leading to the search of theoretical explanation in the automotive market dependencies.

As we describe in Sect. 4, automotive market turned out to be strongly tied to major political cycles in Russia. It appears obvious in retrospect that markets for major consumer goods, such as cars, are tied to geopolitical events and major business cycles more so than to individual consumer variables. However, previous studies [3, 4, 12, 16, 20] focused mostly on the individual-level variables, leaving very little ground for building macro-level extensions to existing theory. This challenge was overcome by the modeling process, described in detail next.

4 Modeling

Vertica² platform is a unified analytics warehouse, allowing to both store the massive amounts of data and employ advanced analytics at the same time. While the actual amount of data in this project were not sufficient to constitute “big data,” Vertica contained all the tools that were necessary for the analysis at hand. Therefore, it was the tool of choice for this project.

The first Vertica module that was employed was the time series analytics. Built-in Vertica functions used in this analysis, included data normalization into time slices, aggregate functions that optimized the values returned for each time slice, and missing data imputation. Data were fully cleaned and prepared for the analysis using in-database functions. Time-series modeling included forecast error decomposition and analysis of impulse responses functions.

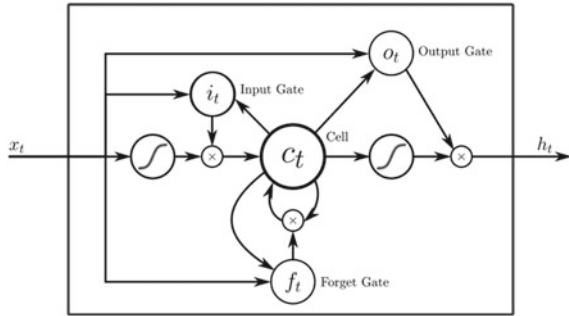
Machine learning module within Vertica allowed to perform a multitude of necessary functions for exploratory data analysis (normalization, outlier detection, joining separate groups of data into one dataset). Vertica allows for integration of existing R libraries directly into its routines, which allowed to combine existing Vertica machine learning modules with R-based forecast libraries. This made the neural network more flexible and more easily adaptable to the data. Results of all computations are provided in Table 1; they represent the output from the neural networks-based forecasting procedure.

² <https://www.vertica.com/>.

Table 1 Partial supporting numbers for Fig. 1, demonstrating forecast accuracy

	July	August	September	October
Actual sales	132,786	136,731	146,722	148,210
Forecast	135,531	137,682	146,758	147,715
Difference	2,745	951	36	495

Fig. 1 Peephole long short-term memory. Adapted from [8]



Starting learning algorithm was Naïve Bayes since many of the variables in the model were nearly independent (based on their pairwise correlations). This algorithm was used to classify variables into different clusters, each of which was then evaluated against the dependent variable.

Original forecasts were built using Support Vector Machine (SVM) for regression. Even though dependent variable belongs to the counts class, given the range of data (anywhere between 90 and 200 K vehicles per month) it was analyzed as continuous. Therefore, SVM for time-series prediction was appropriate in this case. More detailed forecasts (including segment-level forecasts) were created via R forecast libraries and TensorFlow incorporated into Vertica.

Final models were nonlinear autoregressive models built with feed-forward neural networks with a single hidden layer and mostly lagged inputs (independent variables). The number of nodes in the hidden layer was usually half the number of input nodes (including exogenous predictors) plus one, though, for some models, this number was manually adjusted for better performance. For example, for a typical model with 28 independent variables, starting number of nodes would be 15.

Manual adjustment process was iterative, controlled by the resulting model fit. Seasonality was accounted for with several seasonality adjustments (from 1 to 12 months), including monthly and quarterly factor variables. The number of fitted networks varied from 30 to 500, usually with random starting weights. Further in the process, results of previous models were used as weights in the new models. Networks were averaged for building forecasts, and multi-step forecasts were computed recursively.

Below is an example set of node loadings, obtained from one of the procedures, which allows to see how the calculations are performed by the neural network.

$$Y_1[t] = 4529.49 - 1337.4\sqrt[3]{\text{Oil}_{t-10}} + 1.00258 * N_2 \quad (1)$$

$$N_2[t] = 1927.66 - 0.0184439 * \text{Sales}_{t-9} + 1.0017 * N_3 \quad (2)$$

$$N_3[t] = 6535.22 - 1553.82\sqrt[3]{\text{USD}_{t-9}} + 0.997137 * N_4 \quad (3)$$

$$N_4[t] = -13762 + 3153.47\sqrt[3]{\text{USD}_{t-2}} + 1.01063 * N_5 \quad (4)$$

$$N_5[t] = 9915.74 - 190.761\sqrt[3]{\text{Sales}_{t-10}} + 0.994558 * N_6 \quad (5)$$

$$N_6[t] = 33119.5 - 7957.16\sqrt[3]{\text{USD}_{t-11}} + 0.987494 * N_7 \quad (6)$$

$$N_7[t] = -516638 + 4238.55 * \text{BCI}_{t-3} + 0.98867 * \text{cycle} \quad (7)$$

where

- Y dependent variable (car sales)
- t period (in months), starting with time 1, which is a date specified based on the length of the training dataset
- $t - i$ lagged period for i months; e.g., $t - 9$ means a period 9 months ago
- N neural network node number
- Oil price of crude oil, USD, per barrel
- Sales care sales in the indicated period
- USD price of US dollar in rubles
- BCI business confidence index (in percentages)
- Cycle calculated cyclical variable, usually a regular cycle of a specified length.

In the interests of demonstration, we show a model with seven nodes, with dependent variable, $Y_1[t]$, being Node 1 (Eq. 1). This value, $Y_1[t]$, is the forecast of sales at time t . As is apparent from the calculations, calculations roll up to Node 1, with each node taking calculations from previous node and making adjustments to it. So, the first node in the above sequence is Node 7 (Eq. 7), which calculates the value of node at Time t as a linear function of a constant, *business confidence index* three periods ago (Time $t - 3$), and a *cycle* variable. Node 6 (Eq. 6) then is a polynomial regression, with a constant, a cubic root of the *USD price in rubles* 11 periods ago (Time $t - 11$), and a value of Node 7. Each subsequent node takes adjustments to previous nodes, though, as is apparent from the example, such adjustments are usually centered around 1, with variations in the decimals. However, the values of independent variables differ, since no two nodes take the same variable in the same time period. Final calculation (Eq. 1) shows that the forecast, $Y_1[t]$, is a polynomial regression with the cubic root of the oil price ten periods ago and an adjustment aggregation of the value of all previous nodes, contained in N_2 . In the final set of models, these $Y_1[t]$ for each month were the *Forecast* numbers, shown in Table 1.

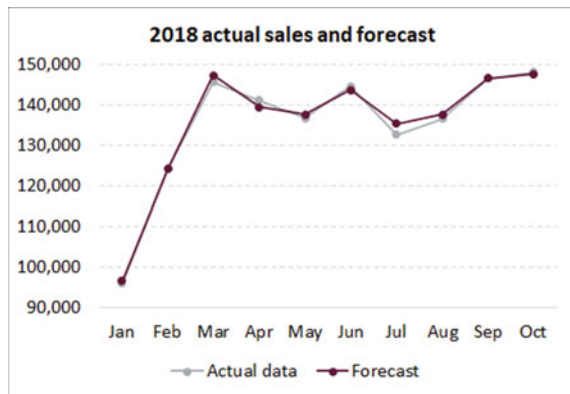
Even though the neural network is usually a “black box” with little information on why certain inputs were used in a certain way, analyzing the structure of node weights allows to draw conclusions about relative importance of the variables that play a role in generating the forecast. We describe the relative importance of each variable in Part 5.2, *Factor Impact Analysis*.

Each of the obtained models were evaluated on the test data, and models with forecast accuracy (using MAPE) under 98% were discarded. Forecast accuracy of the final set of models exceeded 99%. Models were examined for the possibility of overfitting. Given the complexity of the variable structure, overfitting was deemed to be highly unlikely.

Obtained forecasts were evaluated by the client six months after the forecast was turned as a final report. Short-term forecast accuracy of the models against actual future values that were predicted by the models exceeded 99.3%. In fact, in one of the months of the forecast, the difference between the predicted sales and the actual sales comprised just 36 automobiles out of more than 146 K total. Forecast against actual sales are presented in Fig. 2; supporting numbers for some of the data are shown in Table 1. Full data are not provided, because some of them are still protected by the NDA.

It is important to note that this short-term forecast accuracy was achieved because of the iterative comparison of the obtained preliminary results with theoretical background against which the forecast was built. As mentioned, it became apparent early in the modeling process that macroeconomic factors played a major role in forecasting automotive market sales. Consumer-related variables had little to no impact on the outcome. Therefore, theory-based interpretation of the results, and incorporating socio-economic factors into the model, resulted in a much more accurate final forecast.

Fig. 2 Actual sales against the forecast built for the year 2018



5 Interpretation

There are two main groups of drivers of Russian automotive market: economic and socio-political. Even though we did not directly measure the socio-political variables, we found a strong connection between socio-political and automotive market cycles (Fig. 3). This allowed us to conclude that socio-political drivers of Russian automotive market are intangible, and like all intangibles, they do find reflection in financial data. Moreover, important aspect to consider when forecasting Russian automotive market development are the new electoral cycle of 2021–2024, and the corresponding political cycle (expected changes in the Russian parliament—State Duma, and of the President).

The role of sanctions is another aspect of socio-political drivers of Russian automotive market. They mostly have psychological effect as actual economic limitations are minimal. What is more pronounced is high currency volatility and its effect on foreign multinational corporations (MNCs). Russian counter—sanctions also have an effect mostly on individual MNCs and do not have large-scale consequences on automotive market.

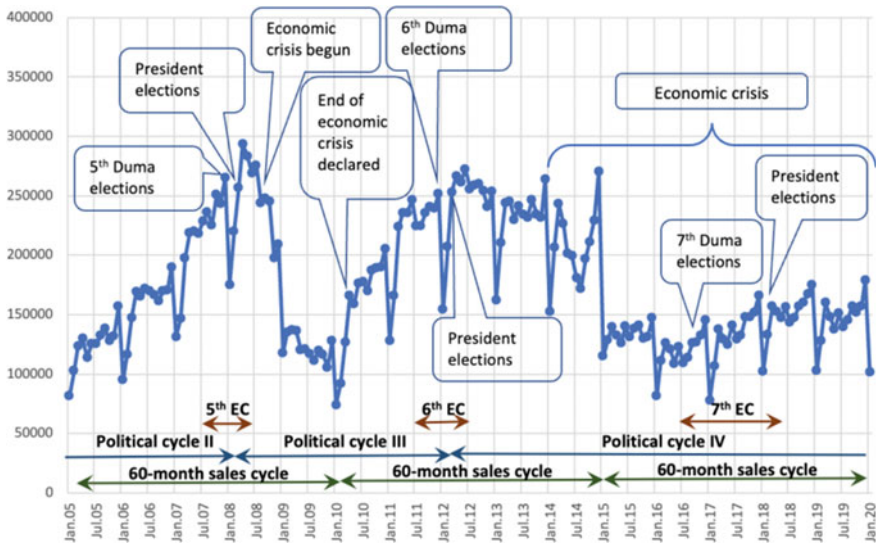


Fig. 3 Political, electoral, and economic cycles in Russia 2005–2020

5.1 Political and Electoral Cycles: Relationships with Economic Cycles

Political cycle changes are almost always accompanied by some political/economic volatility. There appears to be a relatively regular automotive sales cycle of approximately 60 months. Even accounting for the possibility of reaction to crises, the cycle appears to be relatively stable and correlates highly with average time of car ownership (4.9 years). Variety of explanations are possible for the observed market cyclicity, including political (Fig. 2).

Political regimes in Russia are personalistic. That is why a change of political cycle coincides with a change of the head of the state or the President. In the current Russian history, there were seven electoral cycles and four political cycles (Table 2).

The first political cycle (1991–1999) can be traced from the first Russian presidential elections on June 12, 1991. These elections were held in Russia, at the time—still one of the fifteen republics within the former USSR. However, Boris Yeltsin won elections and shortly become equally or even more popular than then-current USSR president Michael Gorbachev. But the first Russian electoral cycle (1991–1993) did not finish until the new constitution of the Russian Federation (RF) was adopted and elections in the new parliament—the State Duma (lower chamber) and the Council of Federation (the higher chamber)—were held on December 12, 1993. As a result of the second electoral cycle (1995–1996), then-current President Boris Yeltsin weakened his position in the parliament, but was nonetheless reelected as the President and stayed in power until the end of 1999.

The results of the third electoral cycle (1999–2000) marked the beginning of the new, second political cycle (2000–2007)—the rule of the new President Vladimir Putin. He won elections in 2000 and became the second President of the RF. The fourth electoral cycle (2003–2004) strengthened Vladimir Putin’s power. Pro-presidential party “United Russia” took constitutional majority in the parliament (more than two thirds’ seats) for the first time. Vladimir Putin won elections with a large margin over the other candidates (71%).

The design of the third political cycle (2008–2011) was made as result of the fifth electoral cycle (2007–2008). According to the constitution Vladimir Putin could not run in the presidential elections for the third time in a row. Therefore, his successor, Dmitriy Medvedev, won the election with the same result as Putin did in 2004 (70%) after the success of the “United Russia” in the State Duma election in 2007 (they took constitutional majority again). During this political cycle, Russia experienced its first economic crisis (2008–2010) as a part of global financial crisis of 2007–2008 after a prolonged period of growth in the early 2000s. This economic crisis resulted in the first downward turn of the automotive market sales after a period of expansive growth.

The fourth political cycle (2012–2021) corresponded to the sixth electoral cycle (2011–2012). In September 2011, Vladimir Putin declared that he would run in presidential elections of 2012. This and the massive fraud during 2011 State Duma

Table 2 Election calendar and cycles

Year, month	Elections and cycles		
	Elections	Electoral cycles	Political cycles
1991, June 12	Russian presidential elections	1st	1st
1993, Dec. 12	1st convocation the State Duma (lower chamber of Russian parliament) elections	1st	1st
1995, Dec. 17	2nd convocation the State Duma (lower chamber of Russian parliament) elections	2nd	1st
1996, June 16, July 3	Russian presidential elections	2nd	1st
1999, Dec. 19	3rd convocation the State Duma (lower chamber of Russian parliament) elections	3rd	From 1st to 2nd
2000, March 26	Russian presidential elections	3rd	
2003, Dec. 7	4th convocation the State Duma (lower chamber of Russian parliament) elections	4th	2nd
2004, March 14	Russian presidential elections	4th	2nd
2007, Dec. 2	5th convocation the State Duma (lower chamber of Russian parliament) elections	5th	From 2nd to 3rd
2008, March 2	Russian presidential elections	5th	
2011, Dec. 21	6th convocation the State Duma (lower chamber of Russian parliament) elections	6th	From 3rd to 4th
2012, March 4	Russian presidential elections	6th	
2016, Oct. 5	7th convocation the State Duma (lower chamber of Russian parliament) elections	7th	4th
2018, March 18	Russian presidential elections	7th	4th
2021, Sept. 19	8th convocation the State Duma (lower chamber of Russian parliament) elections	8th	From 4 to 5th
2024, March 17	Russian presidential elections	8th	

elections caused mass protests in Russia in 2011–2012. Political crisis was accompanied by an economic crisis, which started in 2014. This crisis also found a reflection in another car sales market downturn after a period of growth that followed the first economic crisis and almost matched the highs of the first expansive cycle.

The seventh electoral cycle (2016–2018) took place against the background of two processes. First, it was the end of the acute phase of the economic crisis and its transition to the long-term stagnation. Second, the ongoing “rally around the flag” after the annexation of the Crimea gained support of the great majority of

Russians and strengthen the position of Vladimir Putin. However, this third automotive market cycle was much lower in magnitude than the previous two. Apparently, the crisis recovery was strong enough, and the “rally around the flag” had no economic consequences, to assure full automotive market recovery.

Today, Russia is on the eve of the eighth electoral cycle (2021–2024), which, with high probability, will lead to the change of the Russian president and give a beginning to the next, the fifth political cycle in Russia.

From Fig. 3 it is apparent that the period of change of political cycle is associated with the peak of sales within the 60 month automotive sales cycle. Before the regime change, automotive sales grow at a fast rate. Peak values of sales hold for a while, the first months of the new presidency, but then the trend inevitably changes to a downward one. Downward trend is usually accompanied by the economic crisis—or is a reflection of it. Therefore, before electoral cycle 2021–2024, which must give a lead to regime change and the fifth political cycle, we expected a growth of the automotive sales (the first predictive model we have made in 2018). As we elaborate in the next paragraph, this long-term forecast was also very accurate, very much like the short-term forecast.

We also expected the high uncertainty in automotive sales during the electoral cycle 2021–2024, because this period is a milestone of larger changes. Russian history knew 10–15 years periods of liberalization, which changes with 20–25 years periods of autocratization.³ After 22 years of the new authoritarian regime of Vladimir Putin a new period of liberalization is expected to proceed.

5.2 Cause-Effects: Factor Impact Analysis

Special consideration should be given to the factors that have made the forecast accurate both short- and long-term. As mentioned, consumer-related variables had very little impact on the forecast. Figure 3 presents breakdown of the factors that had an impact on automotive sales; Fig. 4 further breaks down the “other factors” that account for 46% of the variance in the sales.

As is apparent from Figs. 4 and 5, the most important factor affecting the automotive sales is the market health variables (imports, exports, gasoline, Russian natural gas, oil). As Russian economy is the resource-based economy, “market health” variables, apparently, reflected all the intangibles related to the socio-political cycles. Other important variables included prices of metals, gasoline, energy, and a variety of industrial indicators—again, economy-based measures. Consumer-based factors, such as sentiment and confidence, played a minor role.

³ 1917–1927: revolution, Civil War, and Lenin’s new economic policy with some elements of market economy; 1928–1953: Stalin’s dictatorship; 1954–1964: Khrushchev’s de-Stalinization and liberalization; 1965–1985: Beshnev’s and his followers dictatorship; 1986–1999: Gorbachev’s and Yeltsin socio-political and economic reforms of liberalization and democratization in the USSR and Russia.

Fig. 4 Factors that account for the variance in the automotive market sales

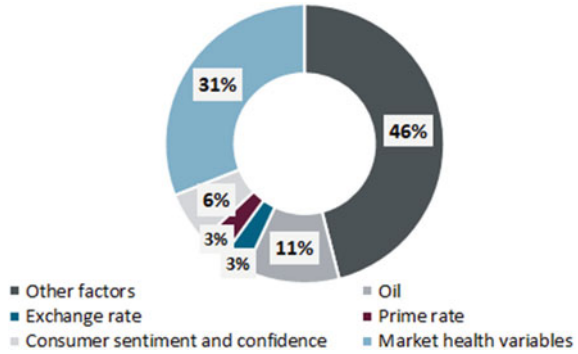
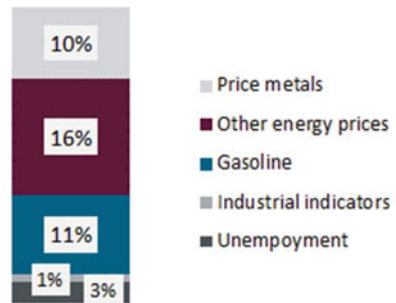


Fig. 5 “Other factors” that impact automotive market sales



5.3 Automotive Sales Forecast: Long-Term Accuracy

This study was conducted in the year 2018 but is presented after most NDA clauses have expired. However, later presentation of the results to the academic community allows to re-evaluate the long-term accuracy of the forecast against the actual sales. Despite the COVID-19-related market downturn, the model has held up. In fact, as can be easily verified by the most recent data available, the forecast made by our model in 2018 was so accurate that it remained well within the predicted boundaries. Covid-19-related crisis has brought the sales lower than they could have been, but still within the margins outlined by the original forecast.

It is clear from the Fig. 6 that 2021 marks the end of the current realistic forecast and coincides with end of Political Cycle IV. Economic uncertainty is manifested by reduction in automotive sales in 2021. Three divergently different forecast scenarios are driven by possible different political developments at the end of the fourth political cycle. Exchange rate forecast was a part of the forecast modeling; it varied from a low of 50 to a high of 88 rubles per dollar.

Period of uncertainty was predicted by the authors in 2018, long before COVID-19 pandemic and political crisis in Russia due to the pandemic, possible war in Ukraine, and other burning political issues of today. It seems that Russia was doomed to slide

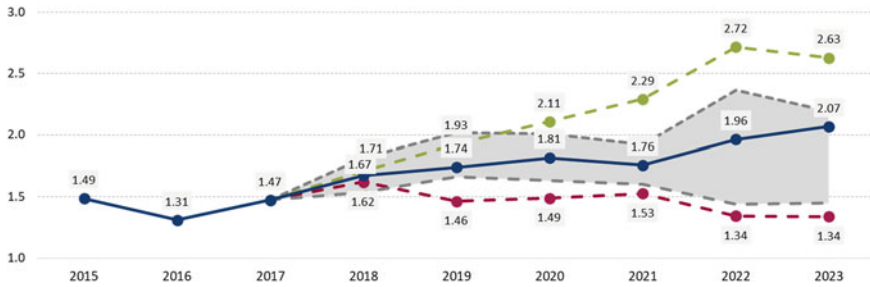


Fig. 6 Forecast till 2023

down into the period of uncertainty as more tectonic changes were preprogrammed by the previous development.

6 Conclusion

In this article, we presented the innovations to the sales forecasting approach. First, we have relied on solid methodology: using a combination of time-series analysis methods with neural network methodology. Second, we have relied on iterative theoretical interpretation of the obtained results to include the appropriate predictor variables in the model. Third, we have evaluated forecast for accuracy both short- and long-term, demonstrating that our approach withstood the test of time and showed excellent results with respect to forecast accuracy.

References

1. Assaad M, Boné R, Cardot H (2006) Predicting chaotic time series by boosted recurrent neural networks. In: King I, Wang J, Chan L-W, Wang D (eds) *Neural information processing*. Springer, Berlin, Heidelberg, pp 831–840
2. Burt G, van der Heijden K (2003) First steps: towards purposeful activities in scenario thinking and future studies. *Futures* 35:1011–1026. [https://doi.org/10.1016/S0016-3287\(03\)00065-X](https://doi.org/10.1016/S0016-3287(03)00065-X)
3. Chung M (2015) From expansion to mature: turning point of the Korean automotive market. In: Jetin B (ed) *Global automobile demand: major trends in mature economies*, vol 1. Palgrave Macmillan UK, London, pp 179–201
4. Dobson R, Larson KE (1979) Psychological and socioeconomic correlates of automobile size. *Transp Res Rec* 723:7–11
5. Fan A, Palaniswami M (2000) Selecting bankruptcy predictors using a support vector machine approach. In: *Proceedings of the IEEE-INNS-ENNS international joint conference on neural networks. IJCNN 2000. Neural computing: new challenges and perspectives for the New Millennium*, vol 6, pp 354–359
6. Gent CR, Sheppard CP (1992) Predicting time series by a fully connected neural network trained by back propagation. *Comput Control Eng J* 3:109–112. <https://doi.org/10.1049/cce:19920031>

7. Gers FA, Eck D, Schmidhuber J (2002) Applying LSTM to time series predictable through time-window approaches. In: Tagliaferri R, Marinaro M (eds) *Neural nets WIRN Vietri-01*. Springer, London, pp 193–200
8. Graves A, Mohamed A, Hinton G (2013) Speech Recognition with deep recurrent neural networks. arXiv:13035778 [cs]
9. Ho SL, Xie M, Goh TN (2002) A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Comput Ind Eng* 42:371–375. [https://doi.org/10.1016/S0360-8352\(02\)00036-0](https://doi.org/10.1016/S0360-8352(02)00036-0)
10. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
11. Höjer M, Mattsson L-G (2000) Determinism and backcasting in future studies. *Futures* 32:613–634. [https://doi.org/10.1016/S0016-3287\(00\)00012-4](https://doi.org/10.1016/S0016-3287(00)00012-4)
12. Jónsson EG (2018) Forecast modelling for the Icelandic automotive market. Thesis
13. Kaastra I, Boyd M (1996) Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10:215–236. [https://doi.org/10.1016/0925-2312\(95\)00039-9](https://doi.org/10.1016/0925-2312(95)00039-9)
14. Kimoto T, Asakawa K, Yoda M, Takeoka M (1990) Stock market prediction system with modular neural networks. In: 1990 IJCNN international joint conference on neural networks. <https://doi.org/10.1109/IJCNN.1990.137535>
15. Langkvist M, Karlsson L, Loutfi A (2014) A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recogn Lett* 42:11–24. <https://doi.org/10.1016/j.patrec.2014.01.008>
16. Philippson A, Aumeboonsuke V (2014) Study of the weak form market efficiency of the automotive industry in nine industrialized markets. In: *The proceedings of the 3rd ICADA 2014*. National Institute of Development Administration (NIDA), Bangkok, p 299
17. Rosenblatt F (1957) *The perceptron—A perceiving and recognizing automaton*. Cornell Aeronautical Laboratory, Ithaca, New York
18. Tam KY, Kiang MY (1992) Managerial applications of neural networks: the case of bank failure predictions. *Manag Sci* 38:926–947. <https://doi.org/10.1287/mnsc.38.7.926>
19. Tang Z, Fishwick PA (1993) Feedforward neural nets as models for time series forecasting. *ORSA J Comput* 5:374–385. <https://doi.org/10.1287/ijoc.5.4.374>
20. Ülengin F, Önsel Ş, Aktas E et al (2014) A decision support methodology to enhance the competitiveness of the Turkish automotive industry. *Eur J Oper Res* 234:789–801. <https://doi.org/10.1016/j.ejor.2013.09.044>

Valentina Kuskova University of Notre Dame, Indiana, USA

Dr. Valentina Kuskova is a Professor of Practice in the Lucy Family Institute for Data & Society and Managing Director AnalytiXIN site in Indianapolis with the University of Notre Dame. Her research interests include network modeling, time-series analysis, and nonparametric methods (especially for efficiency/effectiveness analysis). She is passionate about using science for the benefit of industry and society.

Dmitry Zaytsev University of Notre Dame, Indiana, USA

Dr. Dmitry Zaytsev is a visiting researcher at the University of Notre Dame at Tantur. Dr. Zaytsev has a wide range of research interests in mass protests and social movements; power and efficiency of nation states; education, STI, and environmental policies; democracy and democratization studies; social impact of think tanks, policy advisors, and intellectuals, etc. In aggregate, these topics present the main focus of Dr. Zaytsev's research—forecasting and managing social changes with computational social sciences.

Gregory Khvatsky University of Notre Dame, Indiana, USA

Gregory Khvatsky is a Ph.D. student in Computer Science at the University of Notre Dame. His main research interests is with analysis of complex unstructured and semi-structured datasets and with applications to computational social science. He is also interested in ways that data science can be applied to enact positive social change.

Anna Sokol University of Notre Dame, Indiana, USA

Anna Sokol is a Ph.D. student in Computer Science at the University of Notre Dame. She focuses on quantitative methods; quantifying the political impacts of public behavior and climate change, and economic growth in relation to democracy, human rights, and conflict risks. She is also interested in the design of international agreements for improving communication between scientists and policymakers.

Mental Health Studies: A Review



Rachel Wesley and Hoang Pham

Abstract Mental health is a predominant issue in the United States. Conversations and research around mental health have increased in the past few years. This review examines and categorizes research that has been conducted regarding social media and its impact on mental health into three main categories: (1) Data Collection, Self-Reporting Surveys, and Reviews; (2) Data Collection with Machine Learning Applications; and (3) Modeling of Mental Health Using Survey Data and Machine Learning. Summary tables of risk factors studied, machine learning techniques, and social media sites studied or where data was gathered from are included.

Keywords Mental health · Social media · Machine learning · Modeling · Well-being

1 Introduction

Mental health is defined as emotional, psychological, and social well-being [1]. These are not independent of one another. Social well-being is associated with the development and maintenance of meaningful relationships with others, from which value and a sense of belonging within one's social group is derived [33]. Having a positive mental state enables one to have a stable emotional state, thoughts, and behaviors. It can also allow an individual to realize his or her own potential, work productively, cope with the normal stresses of life, and make positive contributions to society [5]. Being mentally healthy predominantly means having positive characteristics such as a feeling of purpose, contentment, maintaining fulfilling relationships, and participating in life to the fullest [5]. On the other hand, negative mental health may make it difficult to manage thoughts, feelings, or actions in response to daily stressors [7].

R. Wesley · H. Pham (✉)

Department of Industrial and Systems Engineering, Rutgers University, Piscataway 08854, NJ, USA

e-mail: hopham@soe.rutgers.edu

Focusing on mental health can improve physical health, increase productivity, and mitigate stress.

Mental health impacts every person even if they are unaware of it. According to Anwar [1], more than 40 million adults in the United States have an anxiety disorder, but less than 37% of people actually seek mental health treatment for their symptoms. Other common mental health conditions include depression, panic disorders, Post-Traumatic Stress Disorder (PTSD), Obsessive Compulsive Disorder (OCD), and eating disorders. The CDC estimates 1 in 5 experience at least one mental health condition every year [1]. Mental health is not static; it can shift at any time due to biological factors, life experiences, or a predisposition due to family history [21]. There are already ongoing studies about this issue and it is expected to see an increase of interest in the coming decade.

The federal and state governments are responsible for addressing mental health. The federal government is responsible for regulation, consumer protection, supplying funding for services and research. Since the federal government is a major funding source for mental health services, it establishes and enforces minimum standards for states to follow. States are free to expand beyond the standards set by the federal government [20]. The U.S. House of Representatives passed a \$1.5 trillion spending package in March 2022 for Fiscal Year 22 (FY22) that includes investments to expand mental health care [22]. Specifically, \$2.14 billion of the total amount was allocated to the National Institute for Mental Health (NIMH). This is a \$37 million increase that includes \$20 million to expand research on the impact of the COVID-19 pandemic on mental health. Currently there is research that explores the impact of COVID-19 on different groups like the general population, children, elderly, health care professionals, people with preexisting mental health conditions, people with no preexisting mental health conditions, people who were infected with COVID-19, and homeless or refugees [37]. While there is much research and discussion that exists already regarding mental health and the COVID-19 pandemic [13, 26, 27, 37], there is much left to be uncovered.

The COVID-19 pandemic was an unprecedented event that saw long periods of quarantine and social isolation. While social isolation and loneliness were an issue already, the COVID-19 pandemic only exasperated these problems. Preliminary surveys suggest that within the first month of COVID-19, loneliness increased by 20–30% while emotional distress tripled [13]. At the time of publication, Rajkumar [27] reported there were 720,000 cases and 33,000 confirmed deaths and this widespread outbreak is associated with negative mental health consequences. As of July 6, 2022, the World Health Organization (WHO) Coronavirus Dashboard reports 548,990,094 confirmed cases and 6,341,637 deaths worldwide [38]. This is a huge increase in the number of people impacted by the COVID-19 pandemic as well as the number of people who have potentially developed one or more mental health issues stemming from it. Rajkumar [27] suggested that symptoms of anxiety, depression, and self-reported stress are psychological reactions to the pandemic. Panchal et al. [26] reported an increase on various mental health issues like difficulty sleeping, difficulty eating, increases in alcohol consumption or substance use, and worsening chronic conditions due to worry and stress over the coronavirus.

Social media plays a large role in daily life. Devices with screens, such as computers and smartphones, are widely accessible to the general population. A rise in social media has taken place over the last few years in conjunction with this increased accessibility of devices. The COVID-19 pandemic increased the reliance on technology for daily tasks as well as communication with others through social media. Currently, there has been interest and research focus specifically on the impact of social media on children, adolescents, and young adults. The Department of Health and Human Services (HHS) announced in March 2022 nearly 35 million in funding to strengthen and expand mental health services and suicide prevention programs for America's children and young adults [31]. This announcement is part of a new Administration-wide initiative to tackle the nation's mental health crisis.

2 Current Research

The current research done regarding mental health and social media has been mostly from a psychology standpoint. There have been many publications that took data from social media sites or conducted self-reporting surveys to collect data regarding one or more mental health issues. Some studies analyzed this data to make observations regarding the connection between mental health and social media. Other researchers used this data in conjunction with various machine learning techniques to make predictions. A small subset used mathematical modeling techniques along with machine learning to make predictions. Depression and anxiety were the two most commonly studied risk factors in relation to mental health. According to Nesi [24], cyberbullying has been consistently associated with higher rates of suicidal behavior and self harm, which were the third and fourth most studied risk factors in this review, as well as internalizing and externalizing problems. A list of various risk factors can be found in Table 1.

2.1 Data Collection, Self-Reporting Surveys, and Reviews

As interest in mental health research has increased, there has been an increase in review studies as well. Naslund et al. [23] summarized challenges and benefits of social media and mental health by analyzing recent peer-reviewed publications from Medline and Google Scholar. The challenges were categorized as: (1) Impact on symptoms; (2) Facing hostile interactions; and (3) Consequences in daily life. The studies included in this review showed an increased risk of exposure to harm, social isolation, depressive symptoms, bullying, increased loneliness, worsening of mental health symptoms, cyberbullying, increased anxiety, risk of suicide, and privacy concerns. Karim et al. [15] categorized sixteen papers into two outcomes of mental health: anxiety and depression. The main risk factors for anxiety and depression found from the study were time spent, activity type, and addiction to social media.

79 studies focusing on adolescents ranging between 13 and 19 years of age were categorized by Schønning et al. [32] and it was found that depression was the most studied aspect and Facebook was the most studied social network site. From this review, about three-fourths of the studies focused on social media and an aspect of pathology and 94% of the papers in this study were quantitative. Meier and Reinecke [19] also found that depressive symptoms were the most common meta-analyzed indicator in internalizing psychopathology. This study also found that there may be a small negative association between social media and mental health, but results depend on various mental health indicators and computer-mediated-communication approaches. Findings by McCrae et al. [18] corroborated the finding of Schønning et al. [32], Meier and Reinecke [19] by concluding there may be a statistically significant correlation between social media use and depressive symptoms in young people. Chancellor and De Choudhury [6] also found that depression was the most studied symptom and suicidality was the second most studied symptom out of the 75 papers summarized that were published between 2013 and 2018. That literature review focuses on study design, methods, and research design of publications that predict mental health status using social media data. Almost all of the papers cited by [6] frame their contributions as predicting mental health and use algorithms from machine learning and statistical models. After evaluating 501 articles and using inclusion/exclusion criteria, Sadagheyani and Tatari [30] assessed 50 cases for their results regarding mental health and social media. Some negative effects were anxiety, depression, loneliness, poor sleep quality, poor mental health indicators, thoughts of self-harm and suicide, increased levels of psychological distress, cyberbullying, body image dissatisfaction, fear of missing out (FOMO), and decreased life satisfaction.

There have been numerous studies on mental health that involved self-reporting surveys. A large focus of surveys have targeted adolescent-aged persons [14, 25] or college-aged persons [3, 8, 11, 34]. O'Reilly et al. [25] surveyed adolescents ranging in age from 11 to 18 years and found they are concerned about the risks that social media has towards mental wellbeing. Three major findings were that social media was believed to cause mood and anxiety disorders for some adolescents, social media is a platform for cyberbullying, and social media is addicting. Jones et al. [14] used data from the 2021 Adolescent Behaviors Experiences Survey conducted by the CDC from January to June 2021. This survey assessed United States high school students' (grade 9–12) mental health and suicidality during the COVID-19 pandemic with a large sample size of 7,705 participants. This survey found that more than one in three high school students (37.1%) experienced poor mental health during the COVID-19 pandemic. For many students, mental health was impacted by remote learning, social isolation, financial and familial hardships, worry of illness, and reduced access to healthcare. Son et al. [34] studied the impact of COVID-19 on college students' mental health and wellbeing at Texas A&M University in the United States. It was found that 71% saw an increased level of stress and anxiety due to the pandemic. Stressors like fear and worry about health, difficulty in concentrating, disruptions to sleep patterns, decreased social interactions, and increased concerns regarding academics contributed to increased levels of stress, anxiety, and depression. Davila et al. [8] conducted an initial survey and followup survey with college-age young

adults enrolled in an introductory psychology or abnormal psychology course at an unspecified university to establish if there was any relationship between social media and depression or rumination/co-rumination. It was found that depressive symptoms were associated with quality, and not quantity, of the social networking interactions. Some evidence suggested that depressive rumination and co-rumination impacted social networking usage and quality. In a similar research direction, Feinstein et al. [11] surveyed students at Stony Brook University in the United States to test a model whether negative social comparisons on Facebook put individuals at risk for rumination and depressive symptoms. The results of this model concluded that this is true and contributes to linking social networking use to negative mental health outcomes. A survey was conducted at an unnamed university in the south-eastern region of the United States by Berryman et al. [3] and it examined time spent on social media, importance of social media, and tendencies to engage in vaguebooking. Factors considered were general mental health symptoms, suicide ideation, loneliness, social anxiety and decreased empathy. There were also regression analyses carried out for total mental health symptoms, suicidal thoughts, social anxiety, loneliness, and general anxiety. The results from this study indicated that social media use was not predictive of negative mental health, but vaguebooking was predictive of suicidal ideation.

2.2 Data Collection with Machine Learning Applications

Studying mental health using machine learning has allowed for classifying content being posted as well as making various types of predictions. The data utilized in these studies came from social media sites like Reddit [10, 12, 29, 36], Twitter [9, 28], and Instagram [17] as well as from self-reporting surveys [35]. It is expected to see more publications using these and other techniques in the coming years. A summary of the machine learning techniques can be found in Table 2 and a list of the platforms studied or where data was gathered from can be found in Table 3.

Using data directly from the social media sites studied allows researchers to train and obtain more accurate classification algorithms. Gkotsis et al. [12], De Choudhury et al. [10], and Thorstad and Wolff [36] used public data from Reddit to recognize and classify mental illness posts. Using a neural network and deep learning approach, Gkotsis et al. [12] was able to automatically recognize a mental illness post with an accuracy of 91.08% and select the correct theme with a weighted accuracy average of 71.37%. De Choudhury et al. [10] was able to classify content from Reddit into categories of mental health concern or suicide ideation. A machine learning prediction framework was made specifically to see if users would go from the MentalHealth to SuicideWatch subreddit by looking at linguistic structure, interpersonal awareness, and interaction categories. By taking language samples from Reddit and analyzing posts on mental illness subreddits and non-clinical subreddits, Thorstad and Wolff [36] was able to use an L2-penalized logistic regression model to train a program to distinguish different mental illnesses like Attention-Deficit/Hyperactivity Disor-

Table 1 Risk factors studied

Factors	Source(s)
Depression	Gkotsis et al. [12], Meier and Reinecke [19], De Choudhury et al. [10], Reece et al. [28], Davila et al. [8], Sadagheyani and Tatari [30], O'Reilly et al. [25], Bagroy et al. [2], Chancellor and De Choudhury [6], Manikonda and De Choudhury [17], Feinstein et al. [11], De Choudhury et al. [9] Thorstad and Wolff [36], McCrae et al. [18], Braghieri et al. [4], Naslund et al. [23], Karim et al. [15], Berryman et al. [3], Schønning et al. [32], Kim [16], Nesi [24], Richter et al. [29], Jones et al. [14], Son et al. [34]
Anxiety	Gkotsis et al. [12], Meier and Reinecke [19], Sadagheyani and Tatari [30], O'Reilly et al. [25], Chancellor and De Choudhury [6], Manikonda and De Choudhury [17], Thorstad and Wolff [36], Naslund et al. [23], Karim et al. [15], Berryman et al. [3], Schønning et al. [32], Kim [16], Richter et al. [29], Jones et al. [14], Son et al. [34], Bagroy et al. [2]
Suicide or suicide ideation	Gkotsis et al. [12], De Choudhury et al. [10], Sadagheyani and Tatari [30], O'Reilly et al. [25], Chancellor and De Choudhury [6], Manikonda and De Choudhury [17], Naslund et al. [23], Berryman et al. [3], Kim [16], Nesi [24], Jones et al. [14], Son et al. [34], Bagroy et al. [2]
Self harm	Gkotsis et al. [12], Sadagheyani and Tatari [30], Chancellor and De Choudhury [6], Manikonda and De Choudhury [17], Schønning et al. [32], Nesi [24], Bagroy et al. [2]
Cyberbullying	Sadagheyani and Tatari [30], O'Reilly et al. [25] Naslund et al. [23], Schønning et al. [32], Kim [16], Nesi [24]
Eating disorders	De Choudhury et al. [10], Son et al. [34], Chancellor and De Choudhury [6], Manikonda and De Choudhury [17], Schønning et al. [32], Nesi [24]
Stress	Meier and Reinecke [19], O'Reilly et al. [25], Chancellor and De Choudhury [6] Jones et al. [14], Son et al. [34], Bagroy et al. [2]

Table 1 (continued)

Factors	Source(s)
Bipolar disorder	Gkotsis et al. [12], De Choudhury et al. [10], Chancellor and De Choudhury [6], Manikonda and De Choudhury [17], Thorstad and Wolff [36], Bagroy et al. [2]
Loneliness	Meier and Reinecke [19], Sadagheyani and Tatari [30], Naslund et al. [23], Berryman et al. [3], Kim [16]
Schizophrenia	Gkotsis et al. [12], Naslund et al. [23], Chancellor and De Choudhury [6], Manikonda and De Choudhury [17],
Body image disorders	Meier and Reinecke [19], Nesi [24] Sadagheyani and Tatari [30], Schønning et al. [32]
Post-traumatic stress disorder (PTSD)	Reece et al. [28], Bagroy et al. [2], Chancellor and De Choudhury [6], Manikonda and De Choudhury [17]
Sleep or poor sleep quality	Sadagheyani and Tatari [30], Schønning et al. [32] Nesi [24], Son et al. [34]
Addiction	Gkotsis et al. [12], De Choudhury et al. [10], O'Reilly et al. [25]
Self-esteem	De Choudhury et al. [10], O'Reilly et al. [25], Schønning et al. [32]
Severe or stigmatized illness	De Choudhury et al. [10], Chancellor and De Choudhury [6]
Rumination/co-rumination	Davila et al. [8], Feinstein et al. [11], Richter et al. [29]
Alcoholism or alcohol use	Gkotsis et al. [12], Schønning et al. [32]
Fear of missing out (FOMO)	Sadagheyani and Tatari [30], Schønning et al. [32]
Borderline personality disorder (BPD)	De Choudhury et al. [10], Chancellor and De Choudhury [6]
Panic attacks	Chancellor and De Choudhury [6], Manikonda and De Choudhury [17]

der (ADHD), anxiety, bipolar disorder, and depression. Twitter data of users who reported being diagnosed with clinical depression was used by De Choudhury et al. [9]. Through these users' previous postings over a year preceding the depression diagnosis various aspects were studied. It was found that by using different cues

Table 1 (continued)

Factors	Source(s)
Drug use or substance abuse	Schønning et al. [32], Kim [16]
Worry	Richter et al. [29], Son et al. [34]
Opiates	Gkotsis et al. [12]
Autism	Gkotsis et al. [12]
Hopelessness	De Choudhury et al. [10]
Impulsiveness	De Choudhury et al. [10]
Psychological distress	Sadagheyani and Tatari [30]
Decreased life satisfaction	Sadagheyani and Tatari [30]
Obsessive-compulsive disorder (OCD)	Manikonda and De Choudhury [17]
Attention-deficit/hyperactivity Disorder (ADHD)	Thorstad and Wolff [36]
Decreased empathy	Berryman et al. [3]
Somatization	Berryman et al. [3]
Vaguebooking	Berryman et al. [3]
Internalizing or externalizing problems	Nesi [24]
Social exclusion	Nesi [24]
Concentration issues	Son et al. [34]

from social media like engagement level, emotion, language styles, and mentions of antidepressants, one could see useful signals for characterizing the onset of depression in individuals. Supervised learning, principal component analysis (PCA), and support vector machines (SVM) were used to train a model to predict depression. This was seen through decreases in social activity, raised negative affect, and increased relational and medicinal concerns. Manikonda and De Choudhury [17] did a quantitative analysis of images shared on Instagram regarding a variety of mental health issues using automated computer vision techniques and clustering. It was found that people are using Instagram to express discontent around mental health issues, seek support, and disclose information about their emotional distress.

Predictions made using machine learning allow researchers to obtain more accuracy and possibly new insights. Srividya et al. [35] applied several machine learning algorithms like SVM, decision trees, naïve Bayes classifier, K-nearest neighbor classifier, and logistic regression to identify the state of mental health and predict the onset of mental illness in a target group of two different age ranges of 18–21 and 22–26 years old. The initial responses from the target groups were from a questionnaire and a clustering technique was applied and validated to build classifiers to predict the mental health of a person. Advanced machine learning techniques developed by Richter et al. [29] were used to analyze the results of a comprehensive behavioral test battery that detected and quantified cognitive emotional biases. The bagged decision tree classification algorithm had an accuracy of 71.44% for predicting symptomatic patients with high depression, anxiety, or both, and had an accuracy of 70.78% for

Table 2 Machine learning techniques utilized

Source	Technique(s)
Bagroy et al. [2]	Random forests, ada boost, logistic regression, support vector machines (SVM)
Berryman et al. [3]	Regression analysis
Braghieri et al. [4]	Regression analysis
Chancellor and De Choudhury [6]	SVM, logistic regression, feature engineering, random forests, decision trees, deep learning
De Choudhury et al. [9]	Principal component analysis (PCA), SVM supervised learning
Manikonda and De Choudhury [17]	Automated computer vision techniques, clustering
Gkotsis et al. [12]	Deep learning, neural network (NN), convoluted neural network (CNN), SVM
Reece et al. [28]	Random forest, feature extraction, time series analysis
Richter et al. [29]	Random forest, bagged decision tree classification
Srividya et al. [35]	SVM, decision trees, Naïve Bayes classifier, K-nearest neighbor classifier, logistic regression, Density based clustering, bagging, DBSCAN
Thorstad and Wolff [36]	Logistic regression model, clustering analysis, DBSCAN

the non-symptomatic control group of patients. The prediction algorithm had an accuracy of 68.07% in predicting participants with high depression and an accuracy of 74.18% in predicting participants with high anxiety. Computational models were created by Reece et al. [28] using Twitter data and depression history to predict the onset of depression and PTSD. Supervised learning algorithms extracted predictive features measuring affect, linguistic style, and context. These models successfully distinguished between depressed and healthy content. This same process was applied to people with PTSD. The results were consistent even before a diagnosis of depression was issued and a state-space temporal analysis suggests that depression may be predicted from Twitter data several months before a diagnosis. The overall findings strongly support the claim that computational methods can effectively screen Twitter data for indicators of depression and PTSD.

Table 3 Social media platforms data gathered from or studied

Platform	Source(s)
Twitter	Reece et al. [28], Chancellor and De Choudhury [6] De Choudhury et al. [9], McCrae et al. [18], Naslund et al. [23] Karim et al. [15], Schønning et al. [32], Meier and Reinecke [19]
Facebook	Chancellor and De Choudhury [6], McCrae et al. [18], Braghieri et al. [4], Karim et al. [15], Schønning et al. [32] Meier and Reinecke [19]
Reddit	Bagroy et al. [2], Gkotsis et al. [12], De Choudhury et al. [10], Reece et al. [28], Chancellor and De Choudhury [6], Thorstad and Wolff [36],
Instagram	Chancellor and De Choudhury [6], Manikonda and De Choudhury [17], Karim et al. [15], Schønning et al. [32], Meier and Reinecke [19]
Snapchat	McCrae et al. [18], Karim et al. [15], Schønning et al. [32], Meier and Reinecke [19]
WhatsApp	McCrae et al. [18], Schønning et al. [32], Meier and Reinecke [19]
Tumblr	Chancellor and De Choudhury [6], Schønning et al. [32]
MySpace	McCrae et al. [18], Schønning et al. [32]
Youtube	Naslund et al. [23], Schønning et al. [32]
Linkedin	Karim et al. [15]
Skype	Schønning et al. [32]
Weibo	Chancellor and De Choudhury [6]
Reachout	Chancellor and De Choudhury [6]
LiveJournal	Chancellor and De Choudhury [6]
Mixi	Chancellor and De Choudhury [6]
TOYBO	Chancellor and De Choudhury [6]
PTT	Chancellor and De Choudhury [6]
Flickr	Chancellor and De Choudhury [6]
Ping	Schønning et al. [32]
Bebo	Schønning et al. [32]
Hyves	Schønning et al. [32]
Kik	Schønning et al. [32]
Ask	Schønning et al. [32]
Qzone	Schønning et al. [32]

2.3 Modeling of Mental Health Using Survey Data and Machine Learning

There are published studies that use survey data in conjunction with machine learning techniques to formulate mathematical models regarding mental health. This category is smaller as of the writing of this review, but it is expected to see more discussions and publications in the coming years. Kim [16] used data from the Korean Youth Panel Survey to create hierarchical linear models that are estimated to probe the psychological effects of time spent online. This analysis, which was associated with self-reported mental problems and suicidal thoughts, showed that online social networking is negatively associated with the psychological status of Korean students. Again targeting college-age persons, using data from Reddit, Bagroy et al. [2] created a Mental Wellbeing Index (MWI) to evaluate the mental status of over 100 universities in the United States. The relationship between various attributes of the universities and the MWI were shown using machine learning. The actual factors being predicted in this model with 97% accuracy can be found in Table 1. Braghieri et al. [4] looked at when Facebook was introduced at 775 colleges in the United States in conjunction with seventeen successive reports from the National College Health Assessment (NCHA) survey at the time of Facebook's expansion. A two-way fixed-effect (TWFE) was estimated as a baseline. This equation was then expanded upon to model other factors. An event study-version estimation of the model with distance to and from the introduction to Facebook was created. To estimate the effects of the exposure length of Facebook, an intensity level equation was formulated. Some major findings from this study were that the introduction of Facebook at a college increased symptoms of poor mental health (especially depression), students who were susceptible to mental illness used mental healthcare services more, and after Facebook's introduction in conjunction with poor mental health, academic performance suffered. These findings corroborate the hypothesis that social media has had a negative impact on mental health of teenagers and young adults.

3 Concluding Remarks

This review summarizes some of the current research published regarding mental health and social media. Most research is either inconclusive, suggests that there may be small negative correlations, or that there is no correlation at all between mental health and social media. Further research is needed on this subject and there is expected to be a significant amount in the next 10–20 years. Social media as a whole has positive and negative aspects, which are addressed in some of these studies and published papers [23, 24, 30]. Which platforms are popular is always changing and so are the mental challenges people face. That means the impact of social media is also changing over time. This research is interesting and important as the mental health crisis has been continuously worsening.

References

1. Anwar B (2021) Why is mental health important? <https://www.talkspace.com/blog/why-is-mental-health-important/>
2. Bagroy S, Kumaraguru P, De Choudhury M (2017) A social media based index of mental well-being in college campuses. In: Proceedings of the 2017 CHI conference on human factors in computing systems. ACM, Denver Colorado, USA pp 1634–1646
3. Berryman C, Ferguson CJ, Negy C (2018) Social media use and mental health among young adults. *Psychiatr Q* 89(2):307–314
4. Braghieri L, Levy R, Makarin A (2022) Social media and mental health. SSRN
5. CBHS Health (n.d.) What is positive mental health, and how can we foster it? <https://www.cbhs.com.au/mind-and-body/blog/what-is-positive-mental-health-and-how-can-we-foster-it>
6. Chancellor S, De Choudhury M (2020) Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digit Med* 3(1):1–11 (Nature Publishing Group)
7. Change Your Mind (n.d.) What is poor mental health? <https://www.changeyourmindni.org/about-mental-health/what-is-poor-mental-health>
8. Davila J, Hershenberg R, Feinstein BA, Gorman K, Bhatia V, Starr LR (2012) Frequency and quality of social networking among young adults: associations with depressive symptoms, rumination, and corumination. *Psychol Pop Media Cult* 1(2):72–86
9. De Choudhury M, Gamon M, Counts S, Horvitz E (2013) Predicting depression via social media. In: Proceedings of the international AAAI conference on web and social media, vol 7, no 1, pp 128–137
10. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M (2016) Discovering shifts to suicidal ideation from mental health content in social media. In: Proceedings of the 2016 CHI conference on human factors in computing systems. ACM, San Jose California, USA, pp 2098–2110
11. Feinstein BA, Hershenberg R, Bhatia V, Latack JA, Meuwly N, Davila J (2013) Negative social comparison on Facebook and depressive symptoms: rumination as a mechanism. *Psychol Pop Media Cult* 2(3):161–170
12. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, Dutta R (2017) Characterisation of mental health conditions in social media using informed deep learning. *Sci Rep* 7(1):45141 (Nature Publishing Group)
13. Holt-Lunstad J (2020) The double pandemic of social isolation and COVID-19: cross-sector policy must address both | health affairs forefront. <https://www.healthaffairs.org/doi/10.1377/forefront.20200609.53823/full/>
14. Jones SE, Ethier KA, Hertz M, DeGue S, Le VD, Thornton J, Lim C, Dittus PJ, Geda S (2022) Mental health, suicidality, and connectedness among high school students during the COVID-19 pandemic—Adolescent behaviors and experiences survey, United States, January–June 2021. *MMWR Suppl* 71(3):16–21
15. Karim F, Oyewande AA, Abdalla LF, Chaudhry Ehsanullah R, Khan S (2020) Social media use and its connection to mental health: a systematic review. *Cureus* 12(6):e8627
16. Kim HH-S (2017) The impact of online social networking on adolescent psychological well-being (WB): a population-level analysis of Korean school-aged children. *Int J Adolesc Youth* 22(3):364–376 (Routledge). <https://doi.org/10.1080/02673843.2016.1197135>
17. Manikonda L, De Choudhury M (2017) Modeling and understanding visual attributes of mental health disclosures in social media. In: Proceedings of the 2017 CHI conference on human factors in computing systems. ACM, Denver Colorado, USA, pp 170–181
18. McCrae N, Gettings S, Purssell E (2017) Social media and depressive symptoms in childhood and adolescence: a systematic review. *Adolesc Res Rev* 2(4):315–330
19. Meier A, Reinecke L (2021) Computer-mediated communication, social media, and mental health: a conceptual and empirical meta-review. *Commun Res* 48(8):1182–1209
20. Mental Health America (n.d.) The Federal and State role in mental health. <https://www.mhanational.org/issues/federal-and-state-role-mental-health>

21. MentalHealth.gov (2022) What is mental health? | MentalHealth.gov. <https://www.mentalhealth.gov/basics/what-is-mental-health>
22. NAMI (2022) NAMI applauds key mental health investments made in FY 2022 federal budget | NAMI: national alliance on mental illness. <https://www.nami.org/Press-Media/Press-Releases/2022/NAMI-Applaud-Key-Mental-Health-Investments-Made-in-FY-2022-Federal-Budget>
23. Naslund JA, Bondre A, Torous J, Aschbrenner KA (2020) Social media and mental health: benefits, risks, and opportunities for research and practice. *J Technol Behav Sci* 5(3):245–257
24. Nesi J (2020) The impact of social media on youth mental health: challenges and opportunities. *North Carolina Med J Policy Forum* 81(2):116–121
25. O'Reilly M, Dogra N, Whiteman N, Hughes J, Eruyar S, Reilly P (2018) Is social media bad for mental health and wellbeing? Exploring the perspectives of adolescents. *Clin Child Psychol Psychiatr* 23(4):601–613
26. Panchal N, Kamal R, Garfield R (2021) The implications of COVID-19 for mental health and substance use. <https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/>
27. Rajkumar RP (2020) COVID-19 and mental health: a review of the existing literature. *Asian J Psychiatr* 52:102066
28. Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ (2017) Forecasting the onset and course of mental illness with Twitter data. *Sci Rep* 7(1):13006 (Nature Publishing Group)
29. Richter T, Fishbain B, Markus A, Richter-Levin G, Okon-Singer H (2020) Using machine learning-based analysis for behavioral differentiation between anxiety and depression. *Sci Rep* 10(1):16381
30. Sadagheyani HE, Tatari F (2021) Investigating the role of social media on mental health. *Mental Health Soc Incl* 25(1):41–51
31. SAMSHA (2022) HHS announces nearly \$35 million to strengthen mental health support for children and young adults. <https://www.samhsa.gov/newsroom/press-announcements/20220309/hhs-announces-35-million-strengthen-mental-health>
32. Schønning V, Hjetland GJ, Aarø LE, Skogen JC (2020) Social media use and mental health and well-being among adolescents—A scoping review. *Front Psychol* 11
33. Sinclair J (n.d.) What is social well-being? Definition, types, and how to achieve it. <https://www.betterup.com/blog/what-is-social-well-being-definition-types-and-how-to-achieve-it>
34. Son C, Hegde S, Smith A, Wang X, Sasangohar F (2020) Effects of COVID-19 on college students' mental health in the United States: interview survey study. *J Med Internet Res* 22(9):e21279
35. Srividya M, Mohanavalli S, Bhalaji N (2018) Behavioral modeling for mental health using machine learning algorithms. *J Med Syst* 42(5):88
36. Thorstad R, Wolff P (2019) Predicting future mental illness from social media: a big-data approach. *Behav Res Methods* 51(4):1586–1600
37. Tsamakis K, Tsiptsios D, Ouranidis A, Mueller C, Schizas D, Terniotis C, Nikolakakis N, Tyros G, Kypourouopoulos S, Lazaris A, Spandidos D, Smyrnis N, Rizos E (2021) COVID-19 and its consequences on mental health (Review). *Exp Ther Med* 21(3):244
38. WHO (2022) WHO coronavirus (COVID-19) dashboard. <https://covid19.who.int>

Rachel Wesley Rutgers University, New Jersey, USA

Rachel Wesley received her B.S. degree in General Engineering, her B.A. degree in Mathematics from Swarthmore College, Swarthmore, PA, USA, in 2019, and her M.S. degree in Applied Mathematics from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2021. Her research interests include mathematical modeling, system reliability, and optimization.

Hoang Pham Rutgers University, New Jersey, USA

Dr. Hoang Pham is a Distinguished Professor and former Chairman (2007–2013) of the Department of Industrial and Systems Engineering at Rutgers University. His research areas include reliability modeling and prediction, software reliability, and statistical inference. He is editor-in-chief of the *International Journal of Reliability*, and editor of *Springer Series in Reliability Engineering*. Dr. Pham is the author or coauthor of 7 books and has published over 200 journal articles, 100 conference papers, and edited 17 books including *Springer Handbook in Engineering Statistics and Handbook in Reliability Engineering*. He is a Fellow of the IEEE, AAIA, and IISE.