EAI/Springer Innovations in Communication and Computing

José Antonio Marmolejo-Saucedo ·
Román Rodríguez-Aguilar ·
Pandian Vasant · Igor Litvinchev ·
Brenda M. Retana-Blanco *Editors*

Computer Science and Engineering in Health Services

6th EAI International Conference Proceedings, COMPSE 2022, Mexico City, July 28, 2022

EAI
RESEARCH MEETS INNOVATION

Springer

EAI/Springer Innovations in Communication and Computing

Series Editor

Imrich Chlamtac, European Alliance for Innovation, Ghent, Belgium

The impact of information technologies is creating a new world yet not fully understood. The extent and speed of economic, life style and social changes already perceived in everyday life is hard to estimate without understanding the technological driving forces behind it. This series presents contributed volumes featuring the latest research and development in the various information engineering technologies that play a key role in this process. The range of topics, focusing primarily on communications and computing engineering include, but are not limited to, wireless networks; mobile communication; design and learning; gaming; interaction; e-health and pervasive healthcare; energy management; smart grids; internet of things; cognitive radio networks; computation; cloud computing; ubiquitous connectivity, and in mode general smart living, smart cities, Internet of Things and more. The series publishes a combination of expanded papers selected from hosted and sponsored European Alliance for Innovation (EAI) conferences that present cutting edge, global research as well as provide new perspectives on traditional related engineering fields. This content, complemented with open calls for contribution of book titles and individual chapters, together maintain Springer's and EAI's high standards of academic excellence. The audience for the books consists of researchers, industry professionals, advanced level students as well as practitioners in related fields of activity include information and communication specialists, security experts, economists, urban planners, doctors, and in general representatives in all those walks of life affected ad contributing to the information revolution.

Indexing: This series is indexed in Scopus, Ei Compendex, and zbMATH.

About EAI - EAI is a grassroots member organization initiated through cooperation between businesses, public, private and government organizations to address the global challenges of Europe's future competitiveness and link the European Research community with its counterparts around the globe. EAI reaches out to hundreds of thousands of individual subscribers on all continents and collaborates with an institutional member base including Fortune 500 companies, government organizations, and educational institutions, provide a free research and innovation platform. Through its open free membership model EAI promotes a new research and innovation culture based on collaboration, connectivity and recognition of excellence by community.

José Antonio Marmolejo-Saucedo •
Román Rodríguez-Aguilar • Pandian Vasant •
Igor Litvinchev • Brenda M. Retana-Blanco
Editors

Computer Science and Engineering in Health Services

6th EAI International Conference Proceedings, COMPSE 2022, Mexico City, July 28, 2022

Springer

EAI
RESEARCH MEETS INNOVATION

Editors
José Antonio Marmolejo-Saucedo
Engineering Department
National Autonomous University of Mexico
Mexico City, Distrito Federal, Mexico

Pandian Vasant
Universiti Teknologi Petronas
Tronoh, Perak, Malaysia

Brenda M. Retana-Blanco
School of Engineering
Universidad Anahuac México
Naucalpan de Juárez, Mexico

Román Rodríguez-Aguilar
Facultad de Ciencias Económicas y Empresariales
Universidad Panamericana
Mexico City, Mexico

Igor Litvinchev
Graduate Program in Systems Engineering
Nuevo Leon State University
San Nicolás de los Garza, Mexico

ISSN 2522-8595 ISSN 2522-8609 (electronic)
EAI/Springer Innovations in Communication and Computing
ISBN 978-3-031-34749-8 ISBN 978-3-031-34750-4 (eBook)
https://doi.org/10.1007/978-3-031-34750-4

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

Paper in this product is recyclable.

Preface

Universidad Anáhuac México has been delighted to introduce the sixth edition of the European Alliance for Innovation (EAI) International Conference on Computer Science and Engineering on Digital Transformation in Organizations: New Challenges in the Post-Covid Era. This very interesting topic has arisen from the COVID-19 pandemic, which has given us the opportunity to search for different technologies to transmit the knowledge of distinguished engineering personalities from anywhere in the world. The plenary conference "Blurred Technology: Roadmaps for Artificial Intelligence and Bioengineering" by Dr. Utku Kose has invited researchers to explore the new possibilities of interaction who was presented by Brenda Retana.

We had the opportunity to prepare an online congress, receiving 16 papers which discussed topics of biomedicine, optimization, logistics, simulation, machine learning, IoT, strategy and artificial intelligence. The papers were sent from universities located in Europe, America and Asia. This modality of communication gives the opportunity to receive a greater number of papers for subsequent editions.

It is a pleasure to thank the support of the committee which is greatly appreciated, especially to Diego Leon, in charge of broadcasting the event. We are also grateful to the Dean of school of engineering MSc. Mario Buenrostro, Erika Pedraza and Ramiro Navarro for their support, and Dr. José Antonio Marmolejo and Veronika Kissova as EAI Conference Manager.

The School of Engineering of the Anáhuac University México strongly believes that this international forum is relevant to develop, discuss and strengthen new trends by engineering researchers members of scientific community for the benefit of others. We hope to contribute to future editions of EAI.

Naucalpan de Juárez, Mexico | Brenda M. Retana-Blanco

Conference Organization

Steering Committee

Imrich Chlamtac	Bruno Kessler Professor, University of Trento, Italy
José Antonio Marmolejo-Saucedo	Universidad Nacional Autónoma de México, México
Igor Litvinchev	Universidad Autonoma de Nuevo Leon, México

Organizing Committee

General Chair

José Antonio Marmolejo	Universidad Nacional Autónoma de México, México
Utku Köse	Suleyman Demirel University, Turkey
Pandian Vasant	Ton Duc Thang University, Vietnam

General Co-Chairs

Brenda María Retana Blanco	Universidad Anáhuac México, México

TPC Chair and Co-Chair

Román Rodríguez-Aguilar	Universidad Panamericana México, México
Igor Litvinchev	Universidad Autonoma de Nuevo Leon, México

Local Chair

Mario Buenrostro Perdomo	Universidad Anáhuac México, México

Workshops Chair

José Ramón Álvarez Bada	Universidad Anáhuac México, México

Publicity & Social Media Chair

Diego Alberto León Jaime	Universidad Anáhuac México, México

Publications Chair

Erika Elideth Pedraza Arroyo	Universidad Anáhuac México, México

Web Chair

Ariadna Lozano Ramos	Universidad Anáhuac Mexico, Mexico

Demos Chair

Patricio Olea Cossio	Universidad Anáhuac México, México

Technical Program Committee

Pandian Vasant	Ton Duc Thang University, Vietnam
Socorro Rangel	Universidade Estadual Paulista, Brazil
Joshua Thomas	UOW Malaysia KDU Penang University College, Malaysia
Jose Antonio Marmolejo	Universidad Nacional Autónoma de México, México
Román Rodríguez	Universidad Panamericana, Mexico
Utku Kose	Suleyman Demirel University, Turkey
Bharat Singh	Big Data Labs, Hamburg, Germany

Ugo Fiore	Federico II University, Italy
Warusia Yassi	Universiti Teknikal Malaysia Melaka
Ouarda Assas	UNIVERSITE Batna 2, Algeria
Abdellah Derghal	LGEA Laboratory
Jorge Luis Rojas Arce	Universidad Nacional Autonoma de Mexico
Goran Klepac	University college professor, Croatia
Igor Litvinchev	Universidad Autónoma de Nuevo León, Mexico
Valentina Emilia Balas	Aurel Vlaicu University of Arad, Romania
Wei Siang Hoh	Malaysia Pahang University
Mario Pavone	Associate Professor in Computer Science, University of Catania, Italy
Rosana Cavalcante de Oliveira	Faculdade de Ciências Exatas e Tecnológicas, Brasil
Etin Indrayani	Institut Pemerintahan Dalam Negeri, Indonesia

Contents

Part I Supply Chain Optimization

Reverse Logistics in Recycling Companies Using a CVRP Approach 3
Carmen Alexia Gonzalez-Lagunas, Tomas Eloy Salais-Fierro, and Jania Astrid Saucedo-Martinez

Storage Location Assignment Problem in a Warehouse: A Literature Review 15
Lucy Medrano-Zarazúa, Jania Astrid Saucedo-Martínez, and Johanna Bolaños-Zuñiga

A Greenfield Analysis for Supply Chains Enhanced with Agent-Based Simulation 39
Eduardo García-Roa, José Antonio Marmolejo-Saucedo, Ana Paula Martínez, and Fernando Elizarrarás

Resilience in Supply Chains: A Strategy Based on Inventory Policies 47
Rodrigo Eyssautier-Alvarado, José Antonio Marmolejo-Saucedo, and Eduardo García-Roa

Part II Computational Intelligence and Computer Sciences

Intelligent Technology in Geometric Design 63
Igor Litvinchev, Andrii Chuhai, Sergey Shekhovtsov, Tatiana Romanova, and Georgiy Yaskov

An Information Architecture for the Engineering and Design of Industrial Electrical Systems 79
Francisco de Asis López-Fuentes

Optimized Packing Soft Convex Polygons 89
Igor Litvinchev, Luis Infante, Tatiana Romanova, Alberto Martinez-Noa, and Luis Gutierrez

Part III Industry 4.0 Applications

E-Commerce on Startup: A Systematic Literature Review 101
Agung Purnomo, Mega Firdaus, Yogi Tri Prasetyo, Elsa Rosyidah, Satria Fadil Persada, Thalea Christy Nathaniela, and Fairuz Iqbal Maulana

A Strategy to Analyze the Metal Packaging Market in the Food Cans Industry Using Agent-Based Simulation 109
Luis F. Chaverra Vargas and José Antonio Marmolejo-Saucedo

Personalised Emotion Detection from Text Using Machine Learning 169
A. V. Bhavya, R. H. Dhanush, J. Sangeetha, and Arun Cyril Jose

A State of the Art of Non-fungible Tokens: A Literature Review 181
Joan Cattori Krähenbühl and José Antonio Marmolejo-Saucedo

Part IV Health 4.0 and Pervasive Health

A Camera-Based Remote Sensor for Physical Therapy in Parkinson's Disease 203
Jorge L. Rojas-Arce, José Antonio Marmolejo-Saucedo, and Luis Jimenez-Angeles

Health 4.0, Prevention, and Health Promotion in Companies: A Systematic Literature Review 217
Sergio Arturo Domínguez-Miranda and Román Rodríguez-Aguilar

Catastrophic Health Spending by COVID-19 in the Mexican Insurance Sector 247
Ulises Domínguez-Gutiérrez and Román Rodríguez-Aguilar

Modeling and Computer Simulation of Nanocomplexation for Cancer Therapy 257
Tatiana Romanova, Anna Grebinyk, Alexander Pankratov, Yuri Stoyan, Alina Nechyporenko, Yuriy Prylutskyy, Igor Grebennik, and Marcus Frohme

Multiple Sensor Fusion for Stress Detection in the Hospital Environment 273
Muhammad Ali Fauzi and Bian Yang

Index 287

About the Editors

José Antonio Marmolejo-Saucedo is a Professor at the National Autonomous University of Mexico, Mexico. His research is on operations research, large-scale optimization techniques, computational techniques and analytical methods for planning, operations and control of electric energy and logistic systems. He is particularly interested in topics related to artificial intelligence, digital twins, the Internet of Things and Industry 4.0. Currently, Prof. Marmolejo, together with his graduate students, leads the laboratory for the development of digital twins in companies and businesses. He received his PhD in Operations Research (Hons) at the National Autonomous University of Mexico. At present, he has the first-highest country-wide distinction granted by the Mexican National System of Research Scientists for scientific merit (SNI Fellow, Level 3). He is a member of the Network for Decision Support and Intelligent Optimization of Complex and Large-Scale Systems and Mexican Society for Operations Research. He has co-authored research articles in science citation index journals, conference proceedings, presentations, books and book chapters. E-mail: jose.marmolejo@fi.unam.edu.

Román Rodríguez-Aguilar is a Professor at the Faculty of Economic and Business Sciences of the "Universidad Panamericana" in Mexico. His research is on large-scale mathematical optimization, evolutionary computation, statistical learning, computational intelligence, health economics, energy economics, mathematical finance, competition and market regulation. He received his PhD from the School of Economics at the National Polytechnic Institute, Mexico. He also has a Master's degree in Engineering from the School of Engineering at the National University of Mexico (UNAM), a Master's degree in Administration and Public Policy from the School of Government and Public Policy at Monterrey Institute of Technology and Higher Education, a postgraduate in Applied Statistics from the Research Institute in Applied Mathematics and Systems of the UNAM and his degree in Economics from the UNAM. Before joining Panamericana University, he worked as a specialist in economics, statistics, simulation, finance and optimization, occupying different management positions in various public and private entities such as the Ministry of Energy, Ministry of Finance and Ministry of Health. He has co-authored many research articles in science citation index journals, conference proceedings, presentations and book chapters. Professor Rodríguez has supervised many MSc and PhD students. He is a member of the National System of Researchers Level II of CONACYT in Mexico. E-mail: rrodrigueza@up.edu.mx, roman.goldbatch@gmail.com.

Pandian Vasant is a Research Associate at MERLIN Research Centre, TDTU, HCMC, Vietnam, and Editor-in-Chief of *International Journal of Energy Optimization and Engineering* (IJEOE). He holds PhD in Computational Intelligence (UNEM, Costa Rica) and MSc (University Malaysia Sabah, Malaysia, Engineering Mathematics) and BSc (Hons, Second Class Upper) in Mathematics (University of Malaya, Malaysia). E-mail: eic.ijeoe@gmail.com.

Igor Litvinchev received his MSc degree in Applied Mathematics from Moscow Institute of Physics and Technology (Fizteh), Russia, PhD in Systems Theory and Operations Research and DrSci (Habilitation) in Systems Modeling and Optimization from Computing Center, Russian Academy of Sciences, Moscow. He is currently a Professor at Nuevo Leon State University (UANL), Mexico. His research is focused on large-scale systems modeling, optimization, and control with applications to interdisciplinary research. Professor Litvinchev is an author of four books and an editor of eight more books published by Kluwer, Springer and Elsevier. He published more than 90 research papers in leading international journals and served in Program and Organizing Committees for more than 50 international conferences. His research was supported by more than 30 grants from NATO Scientific Affairs Division and European Community; ISF (USA) and RFBR (Russia); CNPq and FAPESP (Brasil); BRFBR (Belarus); CONACYT, PROMEP and PAICYT (Mexico). Professor Litvinchev is a member of the Russian Academy of Natural Sciences and Mexican Academy of Sciences. He is a co-founder of Mexican Society of Operations Research and Mexican Logistics and Supply Chain Association.

Brenda M. Retana-Blanco studied MSc in International Business at The Dublin Institute of Technology in Dublin, Irlanda, Diploma in Business at Macquarie University, Sydney, Australia and a BS in Industrial Engineering at Anahuac University, Mexico.

She is committed, proactive, process-oriented and focused on execution. Brenda is particularly interested in developing academic projects and financial and insurance sectors with the ability to relate and work in multicultural teams in Mexico and overseas.

She is currently Head of Industrial Engineering Department at Universidad Anahuac Mexico. It leads the national and international accreditations of the program. It is also responsible for updating the curriculum of the career, strengthening the knowledge lines of Operations Research, Data Mining, Manufacturing and Quality.

Part I
Supply Chain Optimization

Reverse Logistics in Recycling Companies Using a CVRP Approach

Carmen Alexia Gonzalez-Lagunas, Tomas Eloy Salais-Fierro, and Jania Astrid Saucedo-Martinez

Abstract At present, the generation of waste has increased, due to the high consumption of products and the lack of an efficient collection system as well as its correct disposal of waste, causing it to end up in clandestine dumps or sanitary landfills without these being able to be reused or recycled. In Mexico there are companies that seek to solve this problem, for which they are destined to the collection and correct disposal of MSW, in order to support the companies that generate products so that they can recover them, and these can be reused and used giving them a new use. In order for a greater number of companies to implement a collection process, what is needed is to apply a model where companies have the capacity to cover the generated demand, covering the collection points with a route that minimizes the cost of the process. For this reason, the proposal of this research is to improve a reverse logistics network in its collection process, where the demand points where the MSW are collected are located, the distances between each of the points, logistics costs, using the vehicles destined for collection taking into consideration their capacity. For this reason, it is proposed to apply a model of routing problem with limited CVRP capacity, where routes are generated that cover all collection points, minimize the cost of operation, and take into account the capacity of the vehicles that the company has. The solution must determine the number of vehicles necessary for the process, as well as the order in which each vehicle must visit the generating points, and thus increase the collection of waste, meeting the growing demand.

Keywords Reverse logistics · Residue recollection · Routing · PVR · CVRP · Supply chain

C. A. Gonzalez-Lagunas (✉) · T. E. Salais-Fierro · J. A. Saucedo-Martinez
Universidad Autónoma de Nuevo León, Nuevo Leon, Mexico
e-mail: carmen.gonzalezl@uanl.edu.mx

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_1

1 Introduction

In Mexico, the waste collection and treatment system that exists is deficient, so processes must be implemented so that companies can recover their products, and these can be reused and used by giving them a new use. In recent years, companies have shown greater interest in caring for the environment, in order to protect and safeguard it. Creating policies and strategies that reduce the environmental impact generated by the products they make try to implement a model where their products can be returned for the reuse or use of such products. However, from the model for the product collection and treatment system, the constant problem is the high cost generated by product collection.

In Mexico, 102,895 tons of waste are generated daily, of which only 86,328 tons are collected. Only 9.63% of the waste that is collected is recycled because its materials such as paper, cans, cardboard, and PET, among others, can be used for recycling [1].

This work aims to improve a reverse logistics network, in its collection process, with the application of tools that improve the process and thus increase the amount of waste collected, finding the shortest route; minimizing the operating cost, which covers all the collection points; and solving the problem that companies have, because it is a process that requires a large investment.

The waste collection and reuse process is applicable to the VRP, as well as its variants that have been developed since it is one of the most common to optimize logistics operations. Resulting in an optimal solution and a minimum operating cost of collection. Being defined the capacity, the number of vehicles, demand for collection points and distance between points as restrictions.

2 Theoretical Contextualization

The supply chain is a set of operations that are developed through a series of processes in order to respond to market demands, where stages such as the acquisition of raw material, the transformation of the finished product, and the distribution of the product are involved. Reverse logistics is a stage of the supply chain, which takes place after the delivery of the product to the consumer, where it is sought to return the said product to the supplier. To understand the reverse logistics of waste, its objective must be observed, which consists of the collection, recycling, and treatment of waste, which is generated from the product after its commercialization, so that it can be reused giving a new value either as raw material or spare parts [2].

To illustrate this process, Fig. 1 shows the stages that comprise logistics with the integration of the supply chain and emphasizes that the proposed application model seeks to provide a solution to the collection stage of the supply chain process.

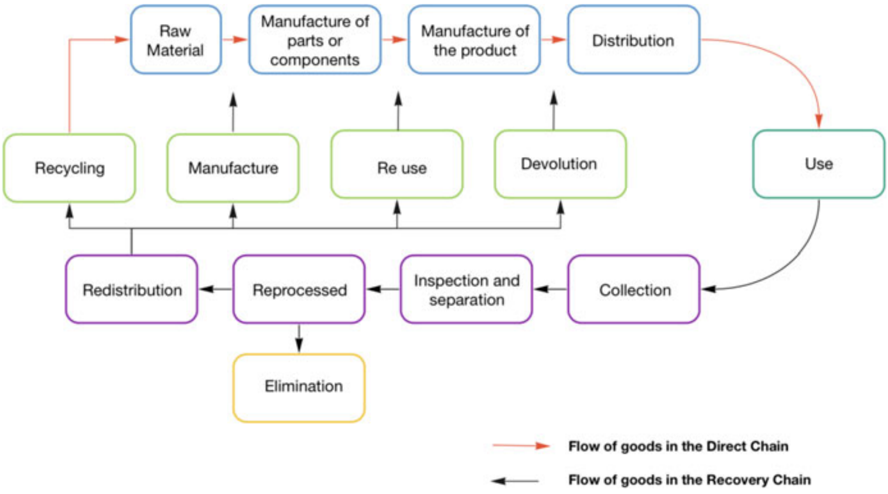


Fig. 1 Reverse logistics process

The reason for implementing reverse logistics ranges from compliance with environmental legislation to economic benefits, which are as follows: decrease in production costs, savings in the purchase of raw materials that are difficult to obtain, customer service and warranties, social responsibility, and generation of competitive advantages.

The researchers [3] comment that reverse logistics provides competitive advantages in a sustainable way. It not only allows quality standards to be controlled but also seeks to identify opportunities to reuse products and reduce costs within companies.

In order to make logistics processes efficient, it is necessary to establish a collection system, taking into account the participation of society, creating strategies that focus on a certain sector of the population, so that within educational institutions at the primary and high school, where a relevant ecological behavior is presented, it can be effective in generating a socio-environmental impact, promoting the culture of recycling in students, thus favoring the collection process [1]. If the model proposed by González-Torre and Adenso-Díaz [7] of the central area is taken as a reference, it is possible to determine the collection points in strategic places, which will be the educational institutions in the case study, where it is possible to cover the elderly demand points, to balance the participation of the student population, so that coverage is greater together with social participation.

There are different models that seek to solve the problems that arise when applying them in a specific case, despite being multiple models, they have some generalities such as the objective, which is the function or goal to be achieved, and how it is going to be achieved, maximizing or minimizing what is required; variables, which are the elements that we are going to find when applying the specific model and that are going to have a direct impact on its application, and it is linked to the variables that they present according to the objective; and restrictions, in some models some values of variables are not allowed, so the model is limited to a possible solution or objective. As a model with these characteristics, its main objective is to find the value that optimizes an objective and can satisfy the restrictions that exist through the application of these models [4].

What is intended is to impact the reverse logistics process in its collection stage by designing routing strategies, time, frequency of visits that optimize costs, and the vehicles used, taking advantage of its capacity to cover the largest number of collection points and increase the collection of MSW, so it is feasible to analyze the vehicle routing problem (VRP).

3 The Vehicle Routing Problem (VRP)

In recent years, the routing problem has become relevant, due to its application and optimization of the process, fulfilling the objective of the companies, since it is an NP-hard problem. It implies that since it is an exact method of solution, it can only be applied to problems of small or medium size where a large number of variables are not involved, satisfying customer demand and minimizing the cost of vehicle travel, where customers are visited once for each vehicle taking into consideration that the capacity of the vehicle is not exceeded by customer demand [5]. The VRP facilitates the set of routes, which can be carried out by a single vehicle or vehicles starting from the company, visiting the assigned clients, and returning to the point of origin. The objective function of the vehicle routing problem is to minimize the total cost of operating the collection process.

According to the type of problem to which this model has been applied, variants of the VRP model have been developed, which are directly related to the characteristics of the problem, taking into account the restrictions, variables, and other elements related to the process. The most used variants are the following:

A. CVRP

The limited capacity routing problem is characterized by the fact that each of the vehicles has a certain capacity where it cannot be loaded above its capacity. Its main objective is to minimize costs and satisfy customer demand, generating routes to travel where the starting point and end point are the same. All the vehicles used have the same capacity and start from the same distribution center. Each of the clients must be visited only once, and the sum of the demand of the clients must not exceed the capacity of the vehicle. Otherwise it would not be viable to visit that point.

B. HFVRP

In the routing problem with a heterogeneous fleet, the vehicles that are considered for the collection process, unlike the CVRP, have different storage capacities. The capacity and types of trucks will depend on the customer demand that define the route they must follow.

C. DCVRP

In the vehicle routing problem with capacity and distance restriction, like the limited capacity model, the vehicles will have the same capacities. Unlike the CVRP, there will be a restriction of maximum distance to travel, which cannot exceed the capacity limit.

D. VRPTW

The routing problem with time window or VRPTW contains an additional restriction where the time used to serve each client is considered. The collection vehicles will be at each point to visit each client at a certain time. This problem has other generalities where the time of arrival at the clients, the route used to reach the clients, the arrival of cargo at the warehouse, and the total service used by the drivers or vehicles are considered.

E. MDVRP

The multiple-depot routing problem is designed to find a set of routes where there are several depots located at different points where each of these depots has its own fleet of vehicles. It is not possible to separate the clients from groups that are close to the deposit since if this were the case we would be talking about solving the problem as a VRP.

F. VRPPD

In the variant vehicle routing problem with collection and delivery, not only will the customer be delivered, but at the same time as the product is delivered, the product will be collected to be returned to the manufacturer. The capacity of the vehicle has a special emphasis because both the delivery and output of the products during the journey cannot exceed the capacity of the truck.

4 Solution Methods

Over time, attempts have been made to implement different criteria to obtain optimal solutions to the problems posed. VRP problems in recent years have been developing and implementing different solution techniques; however due to the complexity of the problem posed and since the cases in which the problem has a large number of variables, it is not possible to solve with exact techniques so the processing time increases, and it is necessary to implement heuristic and metaheuristic techniques.

A. Exact Method

This type of method in its application usually provides the optimal solution, but its main disadvantage is that its execution requires a long time to process, giving as a consequence that in medium- or large-size problems it is not feasible to obtain a result. It has been necessary to develop variants that solve the problems found when applying this tool. The exhaustive search method and branch and event in its application have managed to improve the time in which the information is processed.

B. Heuristic Method

The heuristic method tries to solve problems that have a large number of variables to verify that the solution given by the model is the ideal one and compare it with known optimal results of an exact model applied. Within this type of method, the nearest neighbor and local search method stand out.

C. Metaheuristic Method

This type of model is used in problems where the number of variables is very extensive, so applying this model significantly reduces the processing time compared to an exact method. The solutions thrown turn out to be of better quality than the heuristics. Within this model, techniques have been developed where the results have been improved, such as the taboo search, ant colony, and GRASP.

In order to identify the collection points, the route that the vehicles must take, the starting point, and the end point, a database is required where the information captured geographically can be displayed, so the model will be supported by a system of geographic information.

5 Geographic Information Systems (GIS)

A geographic information system works as a database with the support of computer tools, that is, the software application where the information captured refers to a geographic system. It is a software that allows users to interactively consult where they efficiently integrate and analyze any geographic information by connecting maps with a database. The cartographic location can be obtained from the database record, and the thematic layers can also be separated to work with them in a simple way, facilitating the possibility of relating the information and obtaining results. To capture the information, there are two ways: vector data capture and raster data capture.

6 Statement of the Problem

According to the models used in waste collection problems, and the research carried out, it is considered that the problem posed by the ISIPLAST recycling company is a CVRP (limited capacity routing problem). This model consists of defining the number of vehicles of equal capacity and establishing a route where the cost is minimized, taking the company as the starting point and arrival point, with the restriction that the demand for the points must not exceed the capacity of the vehicle, adding the total cost of all the arcs belonging to the traveled cycle.

The problem includes the location of 30 collection points, the cost of transportation, and the amount of waste to collect at each point. For the collection process, there are 22 vehicles with the capacity to transport 10 tons of waste. The 30 collection points are educational institutions located within a range no greater than 30 kilometers from the collection company. The geographic information system allowed locating the points on a cartographic plane, determining in the same way the geodesic distances between the points to model the problem of routing with limited capacity (Fig. 2).

The objective of taking into consideration educational institutions as collection points in a nearby radius is due to the fact that research such as in [6] showed savings in the collection process by only having to cover the collection points that remain in a certain radius of coverage.

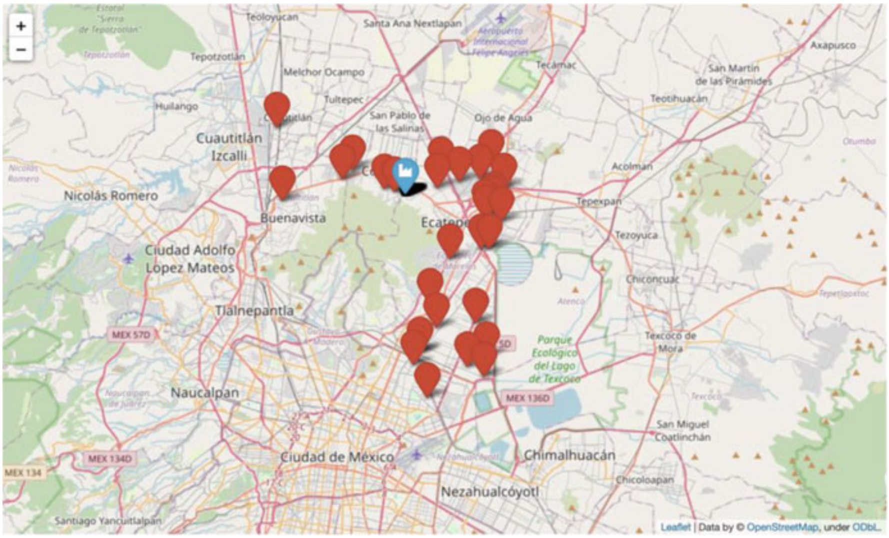


Fig. 2 Map of MSW collection points

7 Mathematical Formulation of the Model

From the previous analysis, and once the areas where the process can be optimized have been identified, the tools defined by their adaptability will be applied to the reverse logistics problem of waste collection, and it is the vehicle routing problem with limited CVRP capacity.

Its objective function is to minimize the operating cost of the collection process, optimizing the route considering the vehicle capacity does not exceed when defining the points. The starting point is the collection company, where educational institutions are visited only once, and finally the collection truck returns, with the collection of waste, to the company. In the same way, the application of this solution method favors the reduction of time, distance, and resources.

In the literature review, a variable number of formulations for the CVRP are shown, which can be divided into three groups: the first, in binary variables that are associated with each link of the graph in order to determine if each vehicle is within the optimal solution or not and the variables that are associated with the arcs and that represent the waste flow within the traveled paths.

Formula Indices

1. $G = (V, X)$ a graph showing the location and path of the collection point.
2. $V = \{0, 1 \ldots \ldots, n\}$ is the set of nodes. The node that represents the company is 0, and the nodes that represent the collection points range from 1 to n.
3. X is the set of vertices that connect to each node $X_{ij} = (i, j)$.
4. $K = \{1, 2, \ldots, |K|\}$: vehicles.
5. q_{i}: demand of each point.
6. Q: maximum vehicle capacity (all vehicles have the same Q capacity).
7. $C_{ij:}$ cost between each point i and j (distance between i and j).

Decision Variable Definition

- Whether or not a vehicle k has visited the route $e_{ij} \times K_{ij} = \{1$ took the point/0 did not take the point.

The objective function is to minimize the sum of costs associated with the collection process, by tracing the shortest route of all the vehicles used so that the desired result can be obtained, and it is necessary to apply some conditions to the formula such as the following:

- Only one vehicle can visit each collection point.
- Return to the point of origin.
- Vehicle capacity K cannot be exceeded.

The CVRP to be formulated as an integer linear programming (IP) problem is represented as follows:

$$\min \sum_{k \in K} \sum_{(i,j) \in E} c_{ij} x_{ij}^{k} \tag{1}$$

$$\sum_{\kappa \in K} \sum_{i \in V, i \neq j} x_{ij}^{k} = 1 \qquad \forall j \in V \backslash \{0\} \tag{2}$$

$$\sum_{j \in V \backslash \{0\}} x_{0j}^{k} = 1 \qquad \forall k \in K \tag{3}$$

$$\sum_{i \in V, i \neq j} x_{ij}^{k} - \sum_{i \in V} x_{ji}^{k} = 0 \qquad \forall j \in V, \forall k \in K \tag{4}$$

$$\sum_{i \in V} \sum_{j \in V \backslash \{0\}, i \neq j} q_j x_{ij}^{k} \leq Q \qquad \forall k \in K \tag{5}$$

$$\sum_{k \in K} \sum_{(i,j) \in S, i \neq j} x_{ij}^{k} \leq |S| - 1 \qquad S \subseteq V \backslash \{0\} \tag{6}$$

$$x_{ij}^{k} \in \{0, 1\} \qquad \forall k \in K, \forall (i, j) \in E \tag{7}$$

Parameters and Restrictions

1. The objective function is to minimize the cost of the trip (sum of the trip distances) of all the vehicles used.
2. Only one visit per vehicle to each collection point.
3. Departure from the company.
4. Constraint that the number of vehicles entering and leaving a point is the same.
5. The maximum capacity of the vehicle must not be exceeded.
6. Remove the created subtours.

Decision Variables

7. Variable constraint: This variable takes the value of 1; if the arc (i, j) is used by the vehicle K, otherwise it takes the value of 0.

8 Solution Methodology

The CVRP problem in urban solid waste collection is an NP-hard problem, so its solution could be complex. So, taking into account the findings in the literature review, and considering that the size of the problem does not require much processing time, it is possible to find its solution using exact techniques. Therefore, the methodology is applied to obtain the collection network with the routes obtained to satisfy the customer's demands considering the restrictions of the proposed model. To graphically represent the model, the formula will be applied in Python on the Anaconda platform with the Solver IBM® ILOG® CPLEX® Optimization Studio in its version 22.1.0 with the support of the SIG Leaflet to represent the location of the points and routes that are obtained from the applied model.

9 Results

The routes that were found are the ones that generate the least cost, thus benefiting the increase in waste collection for the company. The results obtained were evaluated and analyzed, as well as the trucks that were used according to their capacity, maximum travel distance, and time to determine the different routes and attend to all the determined collection points, with the restriction that they must return to the same point of origin, respecting the maximum capacity of the vehicle and the allotted time. In order to better develop the results obtained, the following table is shown.

Route	Vehicle	Harvest	Stops	Time	Distance
Route A	ISI 1	9.3	12	2:37	60.60
Route B	ISI 2	7.5	7	1:40	35.44
Route C	ISI 3	9.9	8	3:00	45.62
Route D	ISI 4	4.3	7	1:47	35.04

From the results obtained, 4 routes were found, which were assigned 4 vehicles of the 22 available for collection, where each one has a capacity of 10 tons. Now, of the four routes, the total distance traveled was 176.7 km, which were distributed among the four routes, comparing the results with the previous process that the company had, where the distance traveled only covered 26 collection points of the 30 traveling a total distance of 235 km resulting in an improvement of 24% by covering all points and traveling less distance.

In the same way, the amount of waste collected was a total of 31 tons compared to its previous process. The company only collected the amount of 26 tons. So in the same way, a 19% improvement was obtained in the amount collected, thus benefiting the company by having a greater amount of material collected for treatment.

For each individual route, there was no problem regarding the time assigned to the collection process, which is 3 hours, because in each of the four routes made, this time limit did not exceed, with route C having the longest time used to carry out the route; however, it was the route that collected the greatest amount of waste.

The total cost generated from the collection process was $26,786.5. If we consider that before the total cost generated was $34,737.64, a saving of $7951.14 was obtained. So the objective function of the model improved the collection cost by 22.8%.

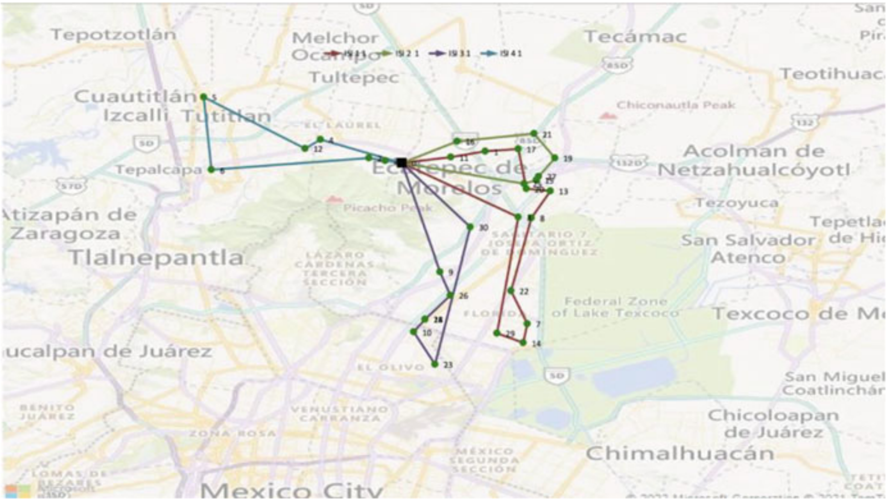


Fig. 3 GIS route trace

To represent the route layout with the support of the geographic information system, the following layout was obtained (Fig. 3):

10 Conclusion

The results obtained in this research work, when applying the proposed CVRP model, are the routes that minimize the cost of operation, which must be carried out by the collection vehicles, in the defined coverage area. Covering the amount of waste required by the company, taking into account the capacity of the vehicle. In addition, new raw material is generated that serves as an additional profit for the company and contributes to the care of the environment. It was possible to apply the routing problem with limited capacity while still observing the number of collection points. However it was observed that when considering a greater number of points, the solution time of the model is solved in a significantly greater time, which gives the guideline for future research to improve the model by applying some method of heuristic or metaheuristic solution which reduces the processing time.

The cost and profitability associated with the implementation of this type of tools should be considered from a financial approach in their cost-benefit relationship, so that they can quantify the benefits not only economic but also social, environmental, and competitive advantages for the company. It was demonstrated that the reverse logistics model improved between 19% and 24% in its collection stage.

References

1. Secretaria del Medio Ambiente: Inventario de Residuos Solidos 2017. [Online]. Available: https://www.sedema.cdmx.gob.mx/storage/app/media/IRS-2016.pdf (2017)
2. Chingal Huaca, D.S.: Logística Inversa aplicada en la gestión de residuos generados por la empresa FlexoFilm, fabricante de fundas plásticas, en la ciudad de Ibarra, p. 552. Universidad Politécnica Estatal del Carchi, México (2019)
3. Lee, C.K.M., Lam, J.S.L.: Managing reverse logistics to enhance sustainability of industrial marketing. Ind. Mark. Manag. **41**(4), 589–598 (2012)
4. Winston, W.L., Goldberg, J.B.: Operations research: applications and algorithms, vol. 3. Thomson Brooks/Cole Belmont (2018)
5. Sami, F., Saoussen, K., Wissem, I.: A DSS based on GIS and Tabu search for solving the CVRP: the Tunisian case. The Egyptian Journal of Remote Sensing and Space Science. **17**(1), 105–110 (2018)
6. Penna, A.F.: Problema del Agente Viajero. XIKUA Boletín Científico de la Escuela Superior de Tlahuelipan. **2**(3) (2014)
7. González-Torre, P.L., Adenso-Díaz, B.: Influence of distance on the motivation and frequency of household recycling. Waste Manag. **25**(1), 15–23 (2005)

Storage Location Assignment Problem in a Warehouse: A Literature Review

Lucy Medrano-Zarazúa, Jania Astrid Saucedo-Martínez, and Johanna Bolaños-Zuñiga

Abstract Warehouse management is one of the many companies' operations that is a main part of the supply chain. Storage takes up between 2 and 5% of the total cost of sales in an organization (Hwang and Cho (Comput Ind Eng 51(2):335–342, 2006)). The storage allocation of products is an activity inside the warehouses that help managing a good flow of the products. It is a complex issue since it depends on many parameters such as the number of spaces, the quantity of products, and the number of rack levels that are in the warehouse, and the resources to move, among other things. When optimizing the flow of operations in the warehouse, good decisions have to be made, and, in this work, through a systematic literature review, different academic papers published between 2005 and 2021 on the SLAP are analyzed and classified according to the solution methodology, objectives, and other characteristics, so that the most used methods in the literature are obtained for this type of problem as well as the parameters that are most taken into account to solve it.

Keywords Storage location assignment problem (SLAP) · Warehouse management · Solution methods

1 Introduction

The supply chain can be defined as a network that manages the relationships between companies and clients, from suppliers to the final customer, using key and cross-functional business processes adding value to the final product or service that companies offer [78].

L. Medrano-Zarazúa (✉) · J. A. Saucedo-Martínez · J. Bolaños-Zuñiga
Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Mexico
e-mail: johana.bolanoszn@uanl.edu.mx
http://www.uanl.mx

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_2

One of the main processes that connects activities on the supply chain is distribution since it is responsible for making the goods and services of a company available to the final consumer and it ensures a quick response to the demand of costumers. Companies have warehouses in their distribution networks.

The impact that storage has on the cost and efficiency of companies is a critical element for the supply chain. Different studies carried out on inventory management show that storage operations in a company represent approximately 23% of logistics costs in the case of countries like the United States and up to 39% in the case of Europe [35].

The warehouse is responsible for receiving, storing, picking, packing, and shipping products between suppliers and customers. The order picking operation is among the activities that generate the most costs within warehouses, with approximately 55% responsibility of operational costs [66].

The storage location assignment problem (SLAP) is an important step when we are designing a warehouse layout, and this step influences in the arranging, picking process, sorting, routing, and sequencing of requests [19]. The problem is to decide the corresponding areas in which the products should be allocated. It can be done using different techniques to establish a specific position for the products. Academic research generally seek to optimize the space used, distances, and cycle times of order picking (OP).

SLAP is considered an NP-hard due to the impact in operations which depends on the order selection method, design and dimensions of the storage system, material handling, type of products, demand behavior, turn over, and spaces within the warehouse.

An adequate allocation in the warehouses must adapt to the flexibility of the demand, which causes a good flow of the materials to optimize the time in the collection of orders and at the same time that it ensures a cost reduction [57].

Figure 1 shows essential operational processes that affect storage systems [37].

Through a systematic review, in this chapter we present state-of-the-art methodologies used to solve the SLAP and the main criteria that are taken into account for the decision-making process.

2 Methodology

The methodology used in the development of this research is based on the content of the qualitative analysis. Lacey and Luff [46] segmented the study into four phases for a better understanding of each of the sections. In this study, the methodology we use refers to qualitative content analysis. Table 1 shows the description of each phase.

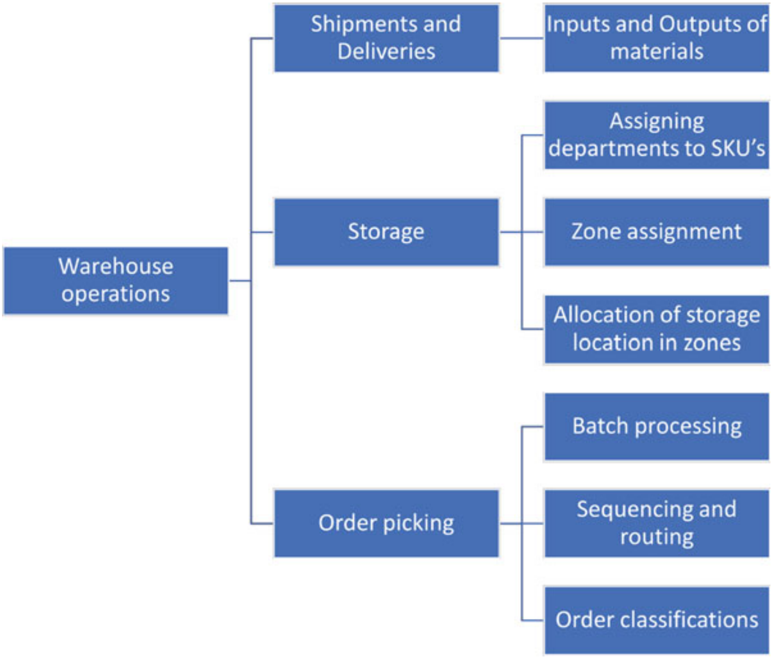


Fig. 1 Operational decisions on the warehouse. *Source: Own elaboration based on [37]*

Table 1 Methodology stages

Stage	Description	Objective
Data collection	Includes papers that search the solution to SLAP	Collect papers that relate to different methods related to SLAP
Sampling strategy	This strategy was established considering a "discrimination" sampling [58] and descriptive analysis	A more evolved bibliographic review of the period 2005–2021
Selection of categories	It maximizes the opportunities to verify the constructed arguments and the relationships among categories	Find the parameters used in the literature for the assigning optimization or reduction of OP time/distance
Evaluation material	This phase is carried out with the compilation of all the material in a thorough manner	Know the methodologies that are state-of-the-art methodologies to solve the problem

2.1 Data Collection

The information collected for the analysis and preparation of this chapter covers the period from 2005 to 2021, considering papers published in journals and thesis works. The investigation was carried out from January 2021 to March 2022.

The keywords that were included in the search process are "storage assignment," "product allocation," "slotting," "layout problem," and "storage location problem."

The material collection phase of the bibliographic review that was presented for the chapter was limited to journals with an exclusive academic format of high impact and those that had a good relevance score.

2.2 Sampling Strategy

Articles with a relevant title to this research were sorted out which a total of 70 were checked, and the number was reduced to 45 papers selected for their relevance and quality. The rest of the papers were discarded due to lack of information on the methodology used, and the combination of the SLAP with other problems in the warehouse "that created a problem for the order picking" or the problem was the allocation design of the machines on the warehouse layout and not about the storage assignment.

Figure 2 shows a graph of the papers collected, and it is possible to see that as of 2015 there was a decrease in the number of papers about the SLAP, and however we can see that it took relevance once again in the previous year.

The investigations consulted come from 14 countries, where different model proposals have been developed with different parameters for a decision-making. Figure 3 shows the ranking of the countries that have worked on the SLAP. It can be shown that the highest concentration of works are in China were the most used

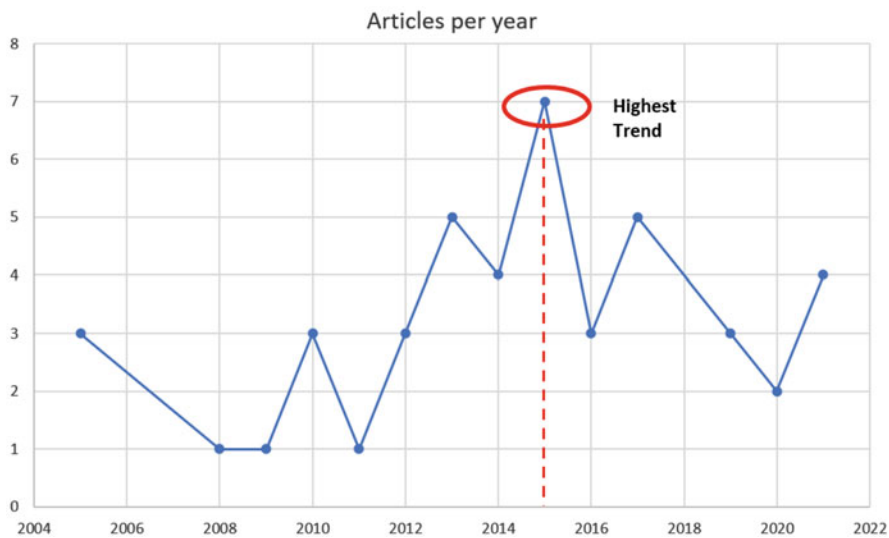


Fig. 2 Distribution of analyzed papers on SLAP

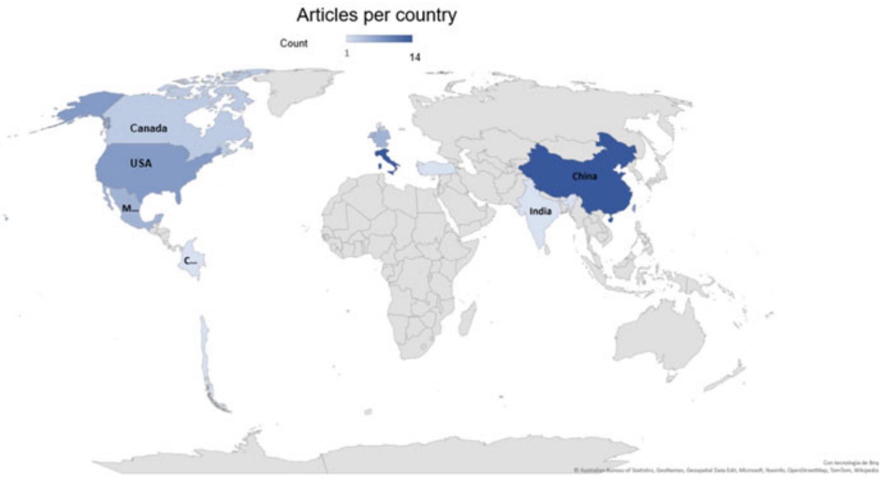


Fig. 3 Distribution of papers by country

methods are the exacts, heuristics, and metaheuristics. They do not include many simulations, but there is a trend to work on genetic algorithms due to their advanced technology and software.

2.3 Category Selection

Three category assignments were made for the papers, and each paper can belong to multiple categories and work with multiple methods inside the same category. The selection of categories was defined in the research process and made according to the scenarios considered in the SLAP problem, dividing them into "Methods used" (Fig. 4), "Performance measures" that were taken into account (Fig. 5), and "Constraints" on the models (Fig. 6).

2.4 Material Evaluation

Exact, heuristic, and metaheuristic methods can be further diversified according to the methodology and solution method chosen for them.

Among the exact methods in the literature, we find that they are divided into different classes depending on the type of problem they solve: linear programming, which is further divided into continuous and integer, multi-objective, and nonlinear programming, which is further divided into constrained and unrestricted. Later in this chapter, these models will be mentioned and a brief explanation of each one will

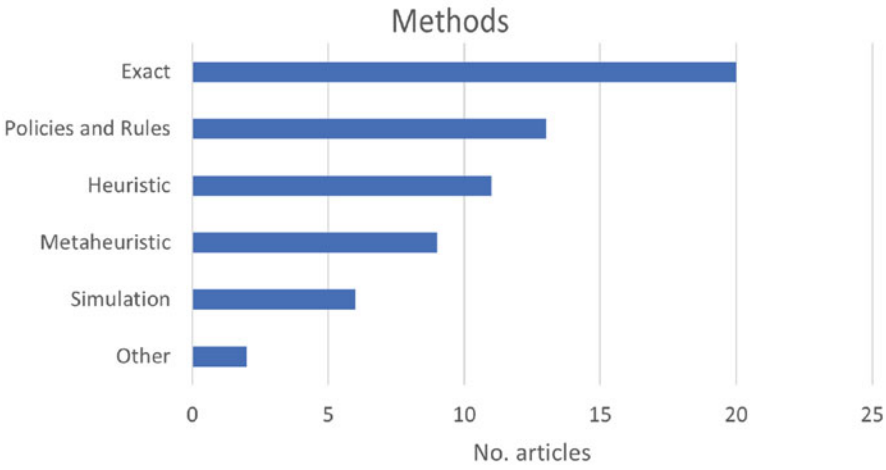


Fig. 4 Distribution of methods used

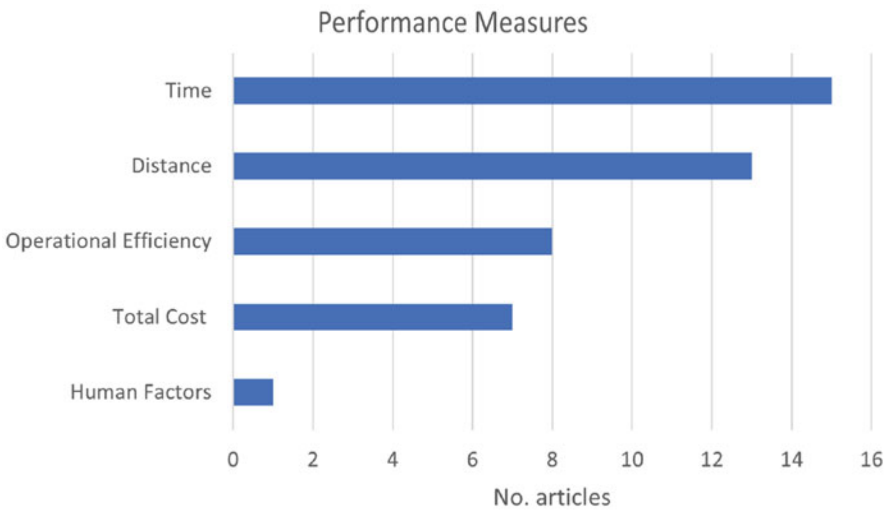


Fig. 5 Distribution of performance measures

be given, whereas in the metaheuristics we find as solution methods for the SLAP the Tabu search, local search, genetic algorithm, iterative local search, simulated annealing, ant colony, and artificial neural networks (Fig. 7).

Figure 8 shows the graph of assignment of papers to categories in percentage.

The evaluation of the selected scientific papers was carried out through a spreadsheet, in which the most relevant information of each paper was organized by name of the publication, authors, objective, methodology, constraints, and conclusions.

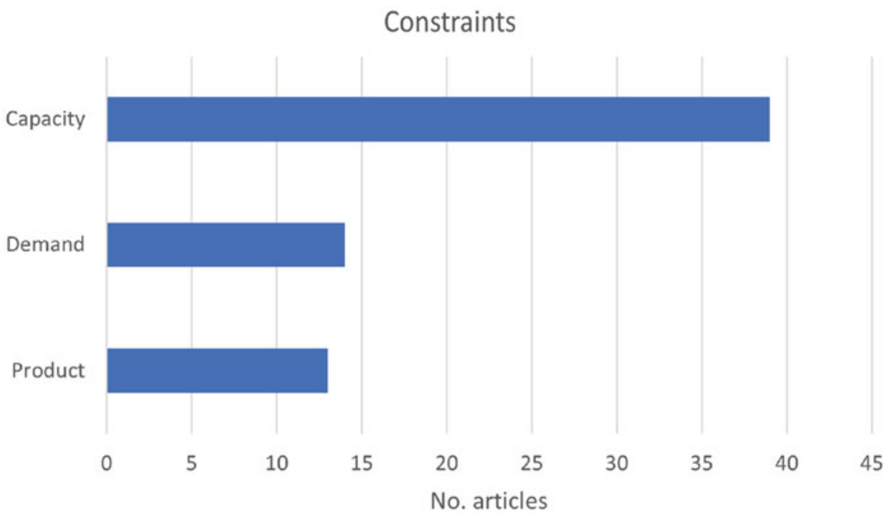


Fig. 6 Distribution of constraints on the models

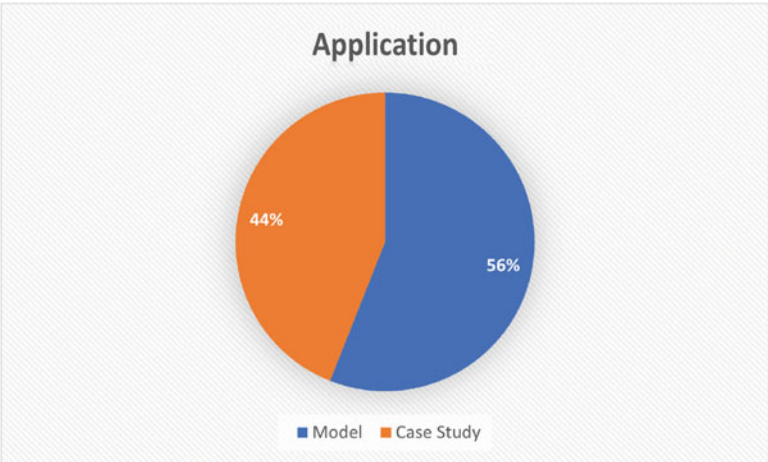


Fig. 7 Application types

Exact methods are the most used in the literature. Policies and rules are widely used by the industry since they are easy to implement and do not require much analysis time, and however it is well known that their solutions are not optimal. Researchers often use them in their papers to compare the results with other methods' results, and moreover simulation methods are often used to review the impact of these policies on their storage management system.

Next, we created a comparative table with the criteria that are taken into account by each model (Table 2):

Table 2 Performance Measures

	Methods						Performance measures					Constraints		
Author(s)	EX	HEU	METAH	SIM	P&R	OT	DIS	TIM	CT	OE	HF	CPT	DEM	PRO
[1]		✓					✓					✓		
[3]	✓						✓					✓	✓	
[4]				✓				✓				✓		
[5]			✓				✓					✓	✓	✓
[7]	✓								✓	✓		✓		✓
[8]	✓									✓		✓	✓	
[13]					✓					✓		✓		✓
[14]	✓							✓				✓		
[15]	✓							✓				✓	✓	✓
[16]	✓								✓			✓		
[17]	✓									✓		✓		✓
[20]	✓	✓	✓		✓					✓		✓		✓
[21]	✓					✓	✓					✓		
[24]			✓				✓					✓	✓	✓
[29]	✓						✓					✓		✓
[31]			✓		✓				✓			✓		
[33]				✓	✓		✓	✓				✓	✓	
[35]	✓	✓		✓	✓			✓				✓		✓
[36]	✓	✓		✓			✓	✓				✓		
[38]	✓						✓					✓		
[39]	✓		✓						✓			✓	✓	✓
[40]		✓	✓					✓				✓	✓	
[41]					✓			✓				✓		
[45]	✓		✓								✓	✓	✓	✓
[47]	✓						✓					✓		
[48]	✓	✓		✓	✓			✓				✓		
[53]					✓					✓		✓		
[54]		✓				✓	✓	✓				✓		✓
[55]			✓	✓				✓				✓		
[57]	✓							✓	✓			✓	✓	
[59]	✓	✓							✓			✓	✓	
[62]					✓		✓					✓		
[67]			✓					✓				✓		✓
[69]		✓						✓		✓		✓	✓	
[70]		✓						✓				✓		✓
[71]		✓			✓			✓				✓		
[72]				✓	✓			✓				✓	✓	
[74]	✓		✓				✓					✓		
[75]	✓		✓				✓					✓		
[77]					✓					✓		✓	✓	
[78]		✓	✓				✓	✓				✓	✓	
[79]	✓				✓				✓			✓		
[80]	✓							✓				✓	✓	
[81]	✓						✓					✓		✓
[82]	✓					✓			✓	✓		✓	✓	

EX Exact, *HEU* Heuristic, *METAH* Mataheuristic, *SIM* Simlation, *OT* Others, *DIS* Distance, *TIM* Time, *CT* Costs, *OE* Operational efficiency, *FH* Human factors, *CPT* Capacity, *DEM* Demand, *PRO* Product

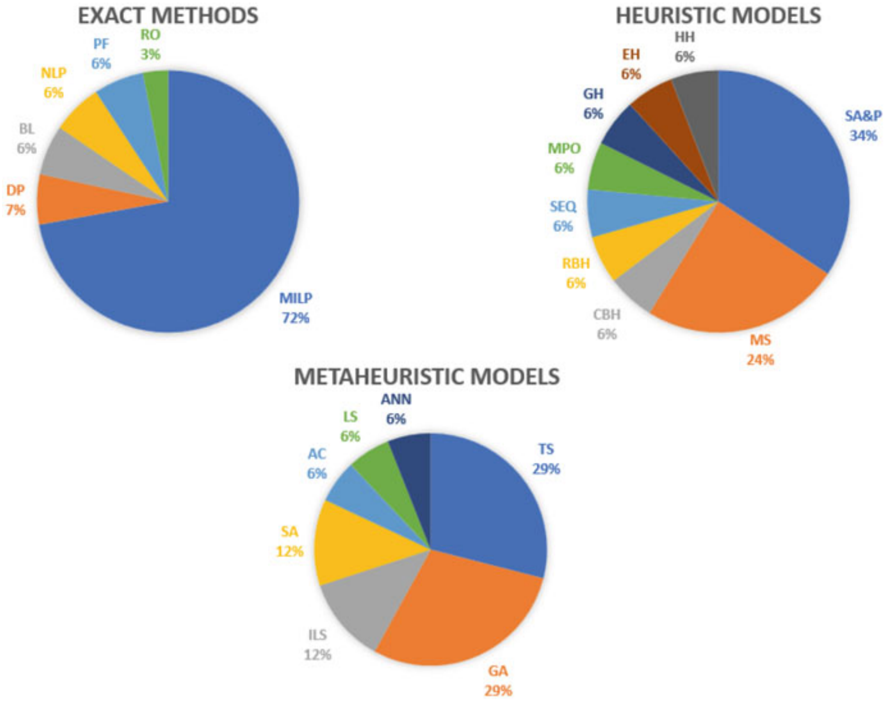


Fig. 8 Detailed distribution of methods used

Table 3 Industry sector of the case studies

Sector	Items
Industrial and manufacturing	[3, 52, 73]
Automatic storage system	[30, 47, 78]
Hardware	[8]
Food	[16, 21, 36, 82]
Gifts	[35]
Furniture and home	[8]
Distribution	[15, 24]

Of the 45 papers analyzed, we were able to find that around 44% of these have a case study in which the model seeks to be implemented. In the remaining investigations, it is more a theorical approach where different classical designs of warehouses found on the literature with different industrial sectors are studied.

Some of the sectors that are mentioned in the literature are those shown in Table 3.

3 Data Discussion and Analysis

The analysis of the information obtained consists of a systematic review of the scientific papers assigned to the categories established in the methodology.

3.1 Exact Models

The advantage of this type of model is that in its application it usually provides the optimal solutions. Nevertheless, it needs to be a relatively "small" problem to be able to compile all the information, which is not feasible in medium or large problems [32].

3.1.1 Mixed Integer Linear Programming (MILP)

A system of linear equations and inequalities [26].

Yang et al. [74] and Ramtin and Pazour [59] address the location assignment and programming problem of automated storage/retrieval systems also known as AS/RS by developing related integer linear programming models to formulate the problem. Both single and multiple cycles of operation are considered. The objective is to minimize the total travel time by creating order picking sequencing routes, taking into account the team distance and speed and how location assignment affects order sequencing.

In the papers by Pérez [57], Bolanos-Zuñiga et al. [52], Guerriero et al. [39], Larco et al. [47], Zhang et al. [80], and Ene and Öztürk [30], a linear programming model is introduced to minimize the product handling through time, distance, or movements of materials (sometimes based on cost) from the entry or output (I/O) areas in the warehouse.

Pérez [57] and Martínez [52] use the idea of a solution that combines different policies for the layout, along with mixed integer linear programming. This model finds an optimal solution by minimizing the entry and exit times of products with forklifts and including the cost reduction for using forklifts, and they have great flexibility to make changes since it distributes the products throughout the entire plant.

Guerriero et al. [39] seek to reduce costs by adding a cost each time a material is moved to the input or output (I/O) doors. In addition, discounts and penalties if some standards applied to the model are met. The objective is to minimize lead times, stock, and total costs of operation on the warehouse while ensuring higher levels of service.

In the work of [80], they work with a multi-objective problem, where it tries to minimize 4 costs: the first cost is for reserving a location for a product with a storage policy, the next one is the cost related to the product cycles from the production zone to the storage areas, then we have the cost of transporting products from picked position to the shipping (exit) point, and the last cost is for the planning, the setup, and the inventory maintenance.

Larco et al. [47] also develop a linear programming model to minimize input and output cycle times, but they added a second objective function where they use a probability parameter, to know if a place will be chosen for a product to be removed from that space depending on how comfortable it is for the worker to retrieve it, taking into account the height.

Using an integer linear programming model, [16] address the problem of industrial buildings located in cities with high seismic activity, creating a model that includes storage and removal of merchandise, seeking the minimizing of travel time for the order picking problem but also the prevention of a collapse of the racks for earthquakes events.

Bolanos-Zuñiga et al. [15] develop a mathematical model based on linear programming that involves both the process of selecting the location of materials in the warehouse and the process of minimizing routes for the preparation of orders.

3.1.2 Dynamic Programming

This programming method solves a global problem, breaking it down into stages that are linked by recursive calculations to generate an optimal feasible relationship for the entire problem. Each calculation is linked, and each decision made in different stages can affect future decisions [42].

This happens when the subdivision of a problem leads to a huge number of problems, problems whose partial solutions overlap, groups of problems of very different complexity, works in stages to solve each problem and optimize each solution in a sequential way.

Dijkstra and Roodbergen [29] seek to relate the impact of the storage location allocation problem with the order picking routing problem, presenting an exact model to calculate the distance traveled on a route that was made for random stocking strategy, making 4 route models to finally determine which is the best storage location for my products.

Bodnar and Lysgaard [14] work on the space allocation and aisle positioning problem (better known as SAAPP for its acronym in English), in which two issues are sought to be optimized at the same time. We have to determinate the quantity of products that need to be assigned to the front area, or hot area, which is where the products have more rotation; the second is the SLAP. The objective is to minimize the number of replenishments made of the products found in the warehouse and improve the handling of materials.

Boysen and Stephan [17] through a dynamic programming try to minimize the effort of the result of picking up a product from a location in an order preparation environment in different instances, this becomes such a complex problem that the author ends up suggesting an heuristic.

Viveros et al. [69] try to find a way to optimize accommodation in slotting families, through a dynamic programming model that sequentially seeks to fix different subproblems in different stages to finally reach a general objective that is to reduce time in the OP process.

3.1.3 Binary Programming

Binary programming is linear programming, the difference is all constraints are based on decision-making, that is, to do a task or not to do it [2].

Fumi et al. [35] combined a multiple method model using an exact method of linear programming and a dedicated storage policy to determinate the assignment of SKU's on the warehouse to reduce space waste.

Chiang et al. [21] develop a binary model from data mining-based storage allocation where they work with a correlation index, so it is more likely that when a type of SKU appears repeatedly in different orders for different clients they will be assigned closer. Therefore, the model can reduce the number of times a collector enters the aisles on a collection run, greatly improving collection efficiency.

3.1.4 Nonlinear Programming

Nonlinear programming is a process where the function to be maximized, or any of the constraints, is different from a first degree or linear equation, where the variables are raised to the power of 1 [9].

Yang et al. [75] work the same as [74] and [59] but with a nonlinear programming approach.

The work of [39] that is presented in the linear programming literature also presents a variant where the model has constraints that are not linear, and therefore it can be proposed in both ways.

3.1.5 Pareto Boundaries

The pareto optimization is another option we have when trying to solve a multi-objective model.

The concept of Pareto optimality defines any situation in which it is not possible to benefit one situation without harming another. Thus, the Pareto optimum is that equilibrium point where you cannot give or ask without affecting other variables. It seeks a solution that in an independent way are not the optimal solutions but simultaneously are superior to the rest of the mutually solutions in the search [12]. It can be seen graphically in Fig. 9.

The work by Battini et al. [8] is an innovative paper that includes the human factor of expenditure due to the workload on pickers, applying a multi-objective

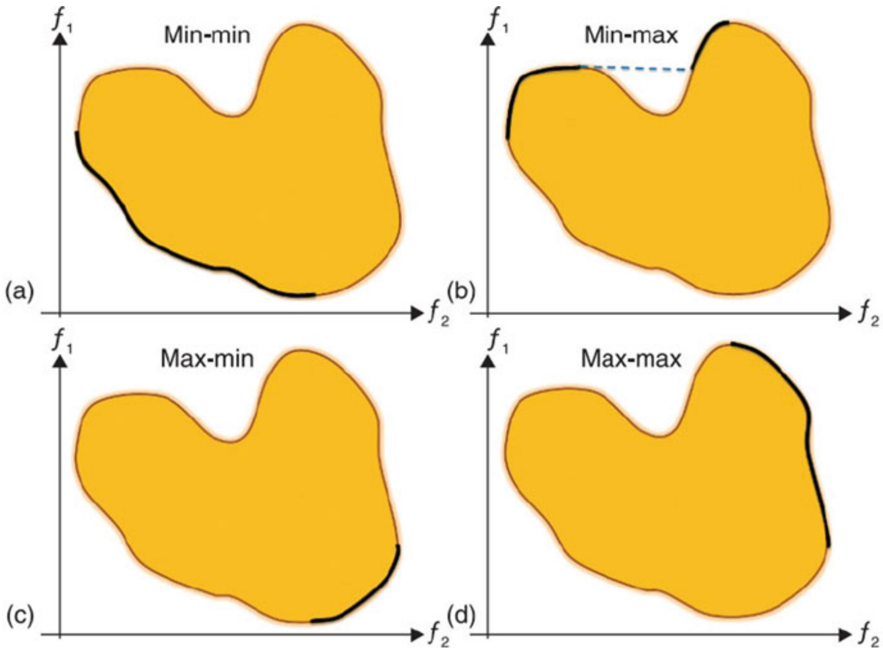


Fig. 9 Pareto optimal solutions. *Source: [12]*

model that seeks to minimize time travels for the OP problem and the energy expenditure on workers.

Ene and Öztürk [30] carry out the work in phases. We have the SLAP solved with a class-based storage policy whose objective is minimize the space used on the warehouse, and then we have another objective where it tries to minimize the travel costs with batching and the vehicle routing problem (VRP).

3.1.6 Robust Optimization

Robust optimization is a method of incorporating uncertainty into mathematical programming models [10]. The key idea is to protect the solutions against the worst case scenarios of the uncertain parameters.

Ang et al. [3] use a robust optimization model to minimize the total cost of storage in a case study in which the demand is uncertain.

3.1.7 Branch and Bound Algorithm (B&B)

The general idea of this algorithm is to decompose the problem into smaller relaxed problems and use their values to find lower bounds for the main problem.

It linearizes the integer programming model, solving it as if it were a linear programming model and then generates bounds if at least one decision variable (integer) adopts a fractional value [25].

Gu et al. [38], through a B&B algorithm, seek to solve the SLAP with the difference that it also seeks to allocate spaces to inventory that is not yet in the warehouse, but due to historical demand data it is known that we will need them in the future. It is based on a relaxed nonlinear modeling, but it searches for upper and lower limits to not extend the problem and for the optimal solution to be found quickly.

3.2 Heuristic Models

A heuristic model is a set of steps that must be performed to identify a high-quality solution to a given problem in the shortest possible time. Heuristics are approximation models that do not guarantee optimal solutions but that are still very close to the real optimal solution and therefore are very well received in the scientific field even if they do not analyze all the information when it is solving a problem, but it shortens the path to find a good resolution [65].

The heuristics that we can find among the literature for the SLAP are distributed in:

- **Algorithms and Procedures (SA&P):** An algorithm follows sequential steps that starts when we input computational information and it transforms it on output valued information [22].
- **Multi-stage (MS) Procedures**: Multi-stage of many single-stage processes, where each one is linked to the previous stage [56]. It is similar to dynamic programming.
- **Hierarchical Procedures (HH)**: It constructs hierarchical criteria giving each decision-making a value to compare with other decisions [6].
- **Sequencing Procedures (SEQ)**: A planning of processes that generates a sequence required to manufacture parts or products to reduce costs or time of production [64].
- **Rollout Heuristics (RBH)**: The starting point is a given policy whose performance is evaluated in some way. Based on the evaluation, an improved policy is obtained. The deployment policy is guaranteed to improve the performance of the base policy [11].
- **2-Option Exchange Heuristic (EH)**: It is an algorithm of the local search family for the traveling agent problem. These algorithms start from an initial solution and iteratively look for opportunities for improvement. This initial solution can be any type of solution as long as it is feasible [68].
- **Greedy Heuristic (GH)**: It seeks an approximate global optimal solution, through local optimal solutions at each stage [76].

– **Multi-product Optimization (MPO) Heuristic**: This is where multiple products travel to different destinations, so they can be transformed and then shipped to the final consumers [61].

In [78], it works with a heuristic model in which they try to reduce congestion between corridors since they work with robots. A model that can work with the correlation of products is used so that in this way those with the highest correlation are separated, and through heuristics of a greedy algorithm and a simulated annealing algorithm (metaheuristic), it seeks to reduce congestion, distance, and time.

As mentioned above [59], deal with the location assignment and programming problem of automated storage/retrieval systems (AS/RS). The objective is to minimize the total travel time by creating routes for the sequencing of order collection through a mixed integer linear programming model, but it also develops a heuristic model that seeks to combine a policy of layout forms and demand curves with the model, through an algorithm.

Wutthisirisart et al. [70] use a sequencing and placement model the objective is to reduce the total routing distance for the OP problem.

Guerriero et al. [40] use a compatibility index to reduce the handling of products on a multilevel warehouse using a heuristic model.

Boysen and Stephan [17] introduced an algorithm using greedy heuristics and a dynamic programming to optimize the routing distance of the OP problem.

Accorsi et al. [1] work sequential hierarchical stages that are linked to defined the decisions for the assignment storage location.

Ming-Huang Chiang et al. [54] propose a new measure of correlation, to represent relationships between products in a distribution center. It introduces the Modified Class-Based Heuristic and the Association Seed Based Heuristic.

Le-Duc* and De Koster [48] propose a 2-option swap heuristic to solve the SLAP. This heuristic swaps proximity classes between aisles, from aisles far from the warehouse to aisles closer to the warehouse to minimize order picking time.

3.3 Metaheuristic Methods

The metaheuristic is an optimization approach that works with cases where data is not complete but the approach to the solution is still good [44].

Metaheuristics are clever strategies to design or improve one or more heuristic simultaneously procedures with a high performance [63].

In the literature of SLAP, there are the following types of metaheuristics:

– **Tabu Search (TS)**: It is a mathematical optimization method, belonging to the class of local search techniques. Tabu search increases the performance of the local search method by using memory structures: once a potential solution is determined, it is marked as “tabu” so that the algorithm does not revisit that potential solution [83].

- **Genetic Algorithm (GA)**: These algorithms make a population of data evolve by subjecting it to random actions similar to those that act in biological evolution (genetic mutations), as well as a selection according to some criterion, based on which it is decided which data is the fittest, which one survives, and which one is discarded [27].
- **Iterative Local Search (ILS)**: The algorithm selects an optimal solution from the adjacent solution space as the current solution and iterates until a local optimal solution is reached. The local search starts from an initial solution and then searches with a neighbor for another solution. If there is a better solution, it moves to that solution and continues searching, otherwise it returns to the previous solution [23].
- **Simulated Annealing (SA)**: It is a method to approximate the global minimum of a route geometry function generated over a large search space that has nonlinearity and discontinuity [51].
- **Ant Colony**: It is a machine learning method inspired by the behavior of food search in ants because, in nature, it is observed that real ants are able to find the shortest path between a food source and their nest with no visual information [60]. The ants tend to follow the others when a path is found thanks to the pheromones, the shorter the path, the longer the pheromone will remain, which creates better paths.
- **Local Search (LS)**: A local search algorithm starts from a randomly chosen instance and moves from one instance to the next to find the best solution. The search is guided by some task-related cost function that estimates the distance between the current assignment and a solution [28].
- **Artificial Neural Networks (ANNs)**: It is an algorithm that is based on how the human brain works. The training procedure for an ANN consists of iteratively supplying the network with a sequence of input patterns and adjusting the weights of the connections based on the outputs obtained [49].

Pan et al. [55] applied AG to solve SLAP and the space that a product can occupy in the racks also trying to balance the workload of pickers dividing the warehouse in zones.

Cruz-Domínguez and Santos-Mayorga [24] developed an artificial neural network and a genetic algorithm for the assignment location of inventory on warehouses applying different resources to minimize distances on the OP routing.

Guerriero et al. [39] developed an ILS-based model to efficiently solve storage location allocation with multiple tiers and compatibility constraints. Furthermore, [45] solved SLAP using simulated annealing, using also a correlation index to minimize the objective function that was the distance travel for order picking.

Chen et al. [20] used Tabu search in an AS/RS system to optimize the distances on the OP routing problem.

Tu et al. [67] using a genetic algorithm show a case where its objective is to balance the workloads in the warehouse due to the constant demand as it is an e-commerce business. In addition, it seeks a solution on what to do in cases where the product runs out of the shelfs, trying to mitigate emergency replenishments to avoid delays in product picking time.

3.4 Rules and Policies

Its objective is to establish the guidelines so that the objectives of a certain project are met. For the SLAP problem, there are already established policies and rules that solve the problem in an accessible time, not in the optimal way, but in a simple way. These policies and rules can be classified into two ways:

- Random storage strategy: The products to enter are placed in the first closest empty space found.
- Dedicated storage strategy: Using correlation rules for the products, they are placed in a zone according to one of the indexes marked below:

  (1) Class-based rule: Homogeneous products are arranged according to a criterion, for example, the useful life of the product.
  (2) Index-based classification: This classification can be ascending or descending values depending on the index that is going to be taken into consideration, there are many types of indexes in the literature such as the popularity-based index, based on the closing date of orders index, turnover index, or the space that a product occupies in the warehouse (Cube per Order Index), and these define that the most demanded products should be assigned to the closest I/O zone.
  (3) Assignment of location by correlation: The products are related in a characteristic, it can be weight, type, or they are found in different client orders multiple times, etc. These correlated products are placed in continuous spaces or in the same corridor to reduce OP distances. The data that creates the correlation must be known, and [34] and [18] illustrate some examples.

Bindi et al. [13] use the correlated assignment location policy testing different rules of clustering.

Meneghetti and Monti [53] develop an index that measures the energy efficiency to save costs on an AS/RS automation system.

Sharma and Shah [62] apply the class-based rule arranging the products by class ad volume.

Zaerpour et al. [79] measure the performance impact of a warehouse using travel time and random stocking strategy class-based and turnover-based index policies.

Bindi et al. [13] apply an index correlation rule in a food sector industry to cluster products and minimize the distance for the OP problem.

Yu and De Koster [77] using different warehouse configurations develop a turnover-based index, comparing them to a class-based rule calculating the total travel time.

Xiao and Zheng [71] have a case study where the warehouse has multiple blocks and aisles, and they study the product information to create a correlation policy that reduces distances on the OP process using information from the manufacturing area.

Manzini et al. [50] propose a simulation that tries to solve the VRP by using a clustering method also, through a correlation index that can also be applied to the OP routing problem process.

Zhang et al. [82] execute a random storage policy joint by an MILP that seeks to reduce the total costs of production and processes in the warehouse but to maximize the space utilized, and later a heuristic model is modeled to implement it in a real-life full-scale model.

3.5 Simulation

They are usually applied to observe different scenarios, on the SLAP problem is used to compare different configuration systems on warehouses by applying different policies in each one of them.

Franzke et al. [33] used a simulation to evaluate the impact that their assignment location had on OP process of their warehouse. Furthermore, [36] had an AS/RS system where they used a different kind of policies to compare them and found the one that minimized their travel routing distance the most. [72] also used a simulation for the same purposes and evaluated different policies to found the most beneficial one to the case study warehouse.

There are works that compare methods to demonstrate that the policies and rules for the accommodation of the products in the warehouse are not the most optimal, just as [4] does, where through a model they seek to apply the formula of gravity in a simulation to reduce the preparation time of orders.

4 Conclusions

This chapter presented a review of the literature on the storage location assignment problem. Forty-five papers from 2005 to 2021 were divided into 3 categories: solution methods, performance measures, and constraints used on the models. We obtained information on how methods work and when it is convenient to work with them. We classified this based on performance measures and constraints or considerations.

We found that different classical methods were applied for the SLAP including exact methods, heuristic, metaheuristic methods, and storage policies, and, in last years, other methods have been used more frequently, including simulation.

Our review of literature gives an update on how the methods work with the new technologies, and we observe a new tendency to work with metaheuristics and simulation thanks to a more advanced software that has been evolving within the years.

The performance measures that were used on the papers analyzed were classified as follows: operational efficiency, distance, time, human factors, and total costs. The constraints that were applied on the models were related to the physical capacity of the warehouses and spaces, product characteristics, and the behavior of the demand. Among the most outstanding models are undoubtedly those of mixed integer linear programming, yielding optimal solutions for those cases where the amount of data is not large. It is also observed that some works first try to implement an MILP, but due to the number of data they have to resort to a heuristic which is generally based on the algorithm of the initial exact method.

A discussion is that there must be a joint effort of researchers to find out how social relationships and decisions affect the product allocation problem in the store and how to link both.

Likewise, for the decision of the location of products in the warehouse, there must be a correlation between this decision, the design, and the operations that are carried out within the warehouse. We must create dynamic policies that study the complexity that we have in the specific case of study, such as the size of aisles, the characteristics of racks, the security measures that must be taken inside the warehouse, the volumes of orders, and the policy of routing that is inside this.

Acknowledgment Supported by PAICYT 547-IT-2022.

References

1. Accorsi, R., Manzini, R., Bortolini, M.: A hierarchical procedure for storage allocation and assignment within an order-picking system. a case study. Int. J. Logist. Res. Appl. **15**(6), 351–364 (2012).
2. Alvarán, G., D.F.E., Hincapié, J.A.G., Correa, J.H.R.: Aplicación de la programación entera binaria para resolver el problema simple de balanceo de linea de ensamble: un caso de estudio. Scientia et technica **2**(50), 84–89 (2012)
3. Ang, M., Lim, Y.F., Sim, M.: Robust storage assignment in unit-load warehouses. Manag. Sci. **58**(11), 2114–2130 (2012). https://doi.org/10.1287/mnsc.1120.1543
4. Ansari, M., Smith, J.S.: Gravity clustering: A correlated storage location assignment problem approach. In: 2020 Winter Simulation Conference (WSC), pp. 1288–1299. IEEE, Piscataway (2020)
5. Bahrami, B., Piri, H., Aghezzaf, E.-H.: Class-based storage location assignment: an overview of the literature. In: 16th International Conference on Informatics in Control. Automation and Robotics (ICINCO 2019), pp. 390–397. Scitepress, Setúbal (2019)
6. BalaSundareshwaran, A., Abdul Rahaman, S., Balasubramani, K., Kumaraswamy, K., Muthuvairavasamy, R.: Habitat risk assessment along coastal Tamil Nadu, India—An integrated methodology for mitigating coastal hazards. In: Coastal Zone Management, pp. 515–542 (2018). ISBN 978-0-12-814350-6. https://doi.org/10.1016/B978-0-12-814350-6.00023-9
7. Ballesteros-Riveros, F.A., Arango-Serna, M.D., Adarme-Jaimes, W., Zapata-Cortes, J.A.: Storage allocation optimization model in a Colombian company. Dyna **86**(209), 255–260 (2019)

8. Battini, D., Glock, C.H., Grosse, E.H., Persona, A., Sgarbossa, F.: Human energy expenditure in order picking storage assignment. Comput. Ind. Eng. **94**(C), 147–157 (2016). ISSN 0360-8352. https://doi.org/10.1016/j.cie.2016.01.020
9. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming: Theory and Algorithms, 3 edn. Wiley-Interscience, Hoboken (2006)
10. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust optimization. In: Robust optimization. Princeton University Press, Princeton (2009)
11. Bertsekas, D.P.: Rollout algorithms for discrete optimization: A survey. In: Handbook of Combinatorial Optimization, Zu, D., Pardalos, P. (Eds.). Springer, Berlin (2010)
12. Besharati Givi, M.K., Asadi, P., Bag, S., Yaduwanshi, D., Pal, S., Heidarzadeh, A., Mudani, S., Kazemi-Choobi, K., Hanifian, H., Braga, D., Silva-Magalhães, A., Moreira, P., Infante, V., Vidal, C., Akinlabi, E., Akinlabi, S., Arora, H., Mukherjee, S., Grewal, H., Zolghadr, P.: Advances in Friction-Stir Welding and Processing. Elsevier, Amsterdam (2014). ISBN 978-08-57094-54-4
13. Bindi, F., Manzini, R., Pareschi, A., Regattieri, A.: Similarity-based storage allocation rules in an order picking system: an application to the food service industry. Int. J. Logist. Res. Appl. **12**(4), 233–247 (2009)
14. Bodnar, P., Lysgaard, J.: A dynamic programming algorithm for the space allocation and aisle positioning problem. J. Operat. Res. Soc. **65**(9), 1315–1324 (2014). https://doi.org/10.1057/jors.2013.64
15. Bolanos-Zuñiga, J., Saucedo, J., Salais, T., Marmolejo, J.: Optimization of the storage location assignment and the picker-routing problem by using mathematical programming. Appl. Sci. **10**, 534 (2020). https://doi.org/10.3390/app10020534
16. Bortolini, M., Botti, L., Cascini, A., Gamberi, M., Mora, C., Pilati, F.: Unit-load storage assignment strategy for warehouses in seismic areas. Comput. Ind. Eng. **87**(C), 481–490 (2015). ISSN 0360-8352. https://doi.org/10.1016/j.cie.2015.05.023
17. Boysen, N., Stephan, K.: The deterministic product location problem under a pick-by-order policy. Discrete Appl. Math. **161**(18), 2862–2875 (2013). ISSN 0166-218X. https://doi.org/10.1016/j.dam.2013.07.002. https://www.sciencedirect.com/science/article/pii/S0166218X13003120
18. Brynzér, H., Johansson, M.I.: Storage location assignment: using the product structure to reduce order picking times. Int. J. Prod. Econ. **46**(1), 595–603 (1996)
19. Charris, E., Rojas-Reyes, J., Montoya-Torres, J.: The storage location assignment problem: a literature review. Int. J. Ind. Eng. Comput. **10**, 07 (2018). https://doi.org/10.5267/j.ijiec.2018.8.001
20. Chen, L., Langevin, A., Riopel, D.: The storage location assignment and interleaving problem in an automated storage/retrieval system with shared storage. Int. J. Prod. Res. **48**(4), 991–1011 (2010)
21. Chiang, D.M.-H., Lin, C.-P., Chen, M.-C.: The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres. Enterp. Inf. Syst. **5**(2), 219–234 (2011). https://doi.org/10.1080/17517575.2010.537784
22. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms. MIT Press, Cambridge (2022)
23. Cruz-Aceves, I., Cervantes-Sanchez, F., Hernandez-Aguirre, A., Hernández-González, M.A., Solorio-Meza, S.: Binary descriptor design for the automatic detection of coronary arteries using metaheuristics. In: Advanced Machine Vision Paradigms for Medical Image Analysis, pp. 169–188. Elsevier, Amsterdam (2021)
24. Cruz-Domínguez, O., Santos-Mayorga, R.: Artificial intelligence applied to assigned merchandise location in retail sales systems. South African J. Ind. Eng. **27**(1), 112–124 (2016)
25. Curtin, K.M.: Operations research. In: Kempf-Leonard, K. (ed.) Encyclopedia of Social Measurement, pp. 925–931. Elsevier, New York (2005). ISBN 978-0-12-369398-3. https://doi.org/10.1016/B0-12-369398-5/00534-X. https://www.sciencedirect.com/science/article/pii/B012369398500534X

26. Dantzig, G.B.: Reminiscences about the origins of linear programming. Oper. Res. Lett. **1**(2), 43–48 (1982)
27. Dawn Thompson, J.: Statistical alignment approaches. Statist. Bioinformat. **1**, 43–51 (2016)
28. Dechter, R.: chapter 7-stochastic greedy local search. In: Constraint Processing, The Morgan Kaufmann Series in Artificial Intelligence, pp. 191–208. Morgan Kaufmann, Burlington (2003)
29. Dijkstra, A., Roodbergen, K.J.: Exact route-length formulas and a storage location assignment heuristic for picker-to-parts warehouses. Transport. Res. Part E Logist. Transport. Rev. **102**, 04 (2017). https://doi.org/10.1016/j.tre.2017.04.003
30. Ene, S., Öztürk, N.: Storage location assignment and order picking optimization in the automotive industry. Int. J. Adv. Manufact. Technol. **60**, 1–11 (2011). https://doi.org/10.1007/s00170-011-3593-y
31. Ene, S., Küçükoğlu, İ., Aksoy, A., Öztürk, N.: A genetic algorithm for minimizing energy consumption in warehouses. Energy **114**, 973–980 (2016)
32. Flórez, J.M.C.: Métodos exactos y heurísticos en la solución de problemas de redes de transporte en las cadenas de suministros. Lupa Empresarial **1**, 44–63 (2017)
33. Franzke, T., Grosse, E.H.,, Glock, C.H., Elbert, R.: An investigation of the effects of storage assignment and picker routing on the occurrence of picker blocking in manual picker-to-parts warehouses. Int. J. Logist. Manag. **1**, 7–17 (2017)
34. Frazelle, E.H.: Stock location assignment and order picking productivity. PhD Thesis, Georgia Institute of Technology (1989)
35. Fumi, A., Scarabotti, L., Schiraldi, M.M.: Minimizing warehouse space with a dedicated storage policy. Int. J. Eng. Business Manag. **5**, 21 (2013). https://doi.org/10.5772/56756
36. Gagliardi, J.-P., Renaud, J., Ruiz, A.: On sequencing policies for unit-load automated storage and retrieval systems. Int. J. Prod. Res. **52**(4), 1090–1099 (2014)
37. Gu, J., Goetschalckx, M., McGinnis, L.F.: Research on warehouse operation: a comprehensive review. Eur. J. Operat. Res. **177**(1), 1–21 (2007). ISSN 0377-2217. https://doi.org/10.1016/j.ejor.2006.02.025. https://www.sciencedirect.com/science/article/pii/S0377221706001056
38. Gu, J., Goetschalckx, M., McGinnis, L.F.: Solving the forward-reserve allocation problem in warehouse order picking systems. J. Operat. Res. Soc. **61**(6), 1013–1021 (2010). https://doi.org/10.1057/jors.2009.39
39. Guerriero, F., Musmanno, R., Pisacane, O., Rende, F.: A mathematical model for the multi-levels product allocation problem in a warehouse with compatibility constraints. Appl. Math. Modell. **37**(6), 4385–4398 (2013). ISSN 0307-904X. https://doi.org/10.1016/j.apm.2012.09.015. https://www.sciencedirect.com/science/article/pii/S0307904X12005124
40. Guerriero, F., Pisacane, O., Rende, F.: Comparing heuristics for the product allocation problem in multi-level warehouses under compatibility constraints. Appl. Math. Modell. **39**(23–24), 7375–7389 (2015)
41. Ho, Y.-C., Liu, C.-F.: A design methodology for converting a regular warehouse into a zone-picking warehouse. J. Chinese Instit. Ind. Eng. **22**(4), 332–345 (2005)
42. Howard, R.A.: Dynamic programming. Manag. Sci. **12**(5), 317–348 (1966)
43. Hwang, H.S., Cho, G.S.: A performance evaluation model for order picking warehouse design. Comput. Ind. Eng. **51**(2), 335–342 (2006). ISSN 0360-8352. https://doi.org/10.1016/j.cie.2005.10.002. https://www.sciencedirect.com/science/article/pii/S0360835206000994. Special Issue: Logistics and Supply Chain Management
44. Jana, M., Biswas, S.: Intelligent and smart enabling technologies in advanced applications: recent trends. Recent Trends Comput. Intell. Enabled Res. **1**, 355–365 (2021)
45. Kim, B.S., Smith, J.S.: Slotting methodology using correlated improvement for a zone-based carton picking distribution system. Comput. Ind. Eng. **62**(1), 286–295 (2012)
46. Lacey, A., Luff, D.: Qualitative Data Analysis. Trent Focus, Sheffield (2001)
47. Larco, J., De Koster, R., Roodbergen, K.J., Dul, J.: Managing warehouse efficiency and worker discomfort through enhanced storage assignment decisions. Int. J. Prod. Res. **55**, 1–16 (2017). https://doi.org/10.1080/00207543.2016.1165880
48. Le-Duc*, T., De Koster, R.M.B.M.: Travel distance estimation and storage zone optimization in a 2-block class-based storage strategy warehouse. Int. J. Prod. Res. **43**(17), 3561–3581 (2005)

49. Malekian, A., Chitsaz, N.: Concepts, procedures, and applications of artificial neural network models in streamflow forecasting. In: Advances in Streamflow Forecasting, pp. 115–147. Elsevier, Amsterdam (2021)
50. Manzini, R., Bindi, F., Pareschi, A.: The threshold value of group similarity in the formation of cellular manufacturing systems. Int. J. Prod. Res. **48**(10), 3029–3060 (2009)
51. Martınez-Alfaro, H., Gomez-Garcıa, S.: Mobile robot path planning and tracking using simulated annealing and fuzzy logic control. Expert Syst. Appl. **15**(3–4), 421–429 (1998)
52. Martínez, J.A.S.: VerificaciÓn y empleo computacional de un modelo matemÁtico utilizado para un layout en un centro de distribuciÓn. Repositorio UANL (2005)
53. Meneghetti, A., Monti, L.: Multiple-weight unit load storage assignment strategies for energy-efficient automated warehouses. Int. J. Logist. Res. Appl. **17**(4), 304–322 (2014)
54. Ming-Huang Chiang, D., Lin, C.-P., Chen, M.-C.: Data mining based storage assignment heuristics for travel distance reduction. Expert Syst. **31**(1), 81–90 (2014)
55. Pan, J.C.-H., Shih, P.-H., Wu, M.-H., Lin, J.-H.: A storage assignment heuristic method based on genetic algorithm for a pick-and-pass warehousing system. Comput. Ind. Eng. **81**, 1–13 (2015)
56. Park, C., Tsui, K.-L.: A profile monitoring of the multi-stage process. Qual. Technol. Quantit. Manag. **16**(4), 407–423 (2019)
57. Pérez, M.M.: Optimización en la disposición del espacio de almacenaje en un centro de distribución. Repositorio UANL (2005)
58. Quintana Peña, A.: Metodología de investigación científica cualitativa (2006). http://biblioteca.udgvirtual.udg.mx/jspui/handle/123456789/2724
59. Ramtin, F., Pazour, J.A.: Product allocation problem for an AS/RS with multiple in-the-aisle pick positions. IIE Trans. **47**(12), 1379–1396 (2015). https://doi.org/10.1080/0740817X.2015.1027458
60. Roy, K., Kar, S., Das, R.N.: Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. Academic, Cambridge (2015)
61. Sampat, A.M., Martin, E., Martin, M., Zavala, V.M., Optimization formulations for multi-product supply chain networks. Comput. Chem. Eng. **104**, 296–310 (2017)
62. Sharma, S., Shah, B.: A proposed hybrid storage assignment framework: a case study. Int. J. Prod. Perform. Manag. **64**, 870–892 (2015)
63. Sörensen, K.: Metaheuristics–the metaphor exposed. Int. Trans. Oper. Res. **22**(1), 3–18 (2015)
64. Sormaz, D., Sarkar, A.: Process sequencing for features with multiple processing requirements. Procedia Manufact. **38**, 726–734 (2019). https://doi.org/10.1016/j.promfg.2020.01.098
65. Todd, P.M.: Heuristics for decision and choice. In: Smelser, N.J., Baltes, P.B. (eds.) International Encyclopedia of the Social & Behavioral Sciences, pp. 6676–6679. Pergamon, Oxford (2001). ISBN 978-0-08-043076-8. https://doi.org/10.1016/B0-08-043076-7/00629-X. https://www.sciencedirect.com/science/article/pii/B008043076700629X
66. Tompkins, J.A., White, J.A., Bozer, Y.A., Frazelle, E.H., Tanchoco, J.M.A., Trevino, J.: Facilities Planning, pages 36–47, 2nd edn. Willey, Hoboken (1996)
67. Tu, M., Yang, M.-F., Kao, S.-L., Lin, F.-C., Ming-Hung, W., Lin, C.-K.: Using a heuristic multi-objective genetic algorithm to solve the storage assignment problem for CPS-based pick-and-pass system. Enterpr. Informat. Syst. **15**(9), 1238–1259 (2021)
68. Verhoeven, M.G.A., Aarts, E.H.L., Swinkels, P.C.J.: A parallel 2-opt algorithm for the traveling salesman problem. Future Gener. Comput. Syst. **11**(2), 175–182 (1995)
69. Viveros, P., González, K., Mena, R., Kristjanpoller, F., Robledo, J.: Slotting optimization model for a warehouse with divisible first-level accommodation locations. Appl. Sci. **11**(3), 936 (2021)
70. Wutthisirisart, P., Noble, J.S., Alec Chang, C.: A two-phased heuristic for relation-based item location. Comput. Ind. Eng. **82**(C), 94–102 (2015). ISSN 0360-8352. https://doi.org/10.1016/j.cie.2015.01.020
71. Xiao, J., Zheng, L.: A correlated storage location assignment problem in a single-block-multi-aisles warehouse considering BOM information. Int. J. Prod. Res. **48**(5), 1321–1338 (2010)

72. Yang, M.-F.: Using simulation to object-oriented order picking system. Informat. Technol. J. **7**(1), 224–227 (2008)
73. Yang, T., Peters, B.A.: Flexible machine layout design for dynamic and uncertain production environments. Eur. J. Oper. Res. **108**(1), 49–64 (1998)
74. Yang, P., Miao, L., Xue, Z., Qin, L.: An integrated optimization of location assignment and storage/retrieval scheduling in multi-shuttle automated storage/retrieval systems. J. Intell. Manufact. **26**, 11 (2013). https://doi.org/10.1007/s10845-013-0846-7
75. Yang, P., Peng, Y., Ye, B., Miao, L.: Integrated optimization of location assignment and sequencing in multi-shuttle automated storage and retrieval systems under modified 2n-command cycle pattern. Eng. Optim. **49**(9), 1604–1620 (2017). https://doi.org/10.1080/0305215X.2016.1261128
76. Yilmaz Balaman, S.: Modeling and optimization approaches in design and management of biomass-based production chains. Basic Concepts Methodol. 185–236 (2019). ISBN 9780128142783. https://doi.org/10.1016/B978-0-12-814278-3.00007-8
77. Yu, Y., De Koster, R.B.: On the suboptimality of full turnover-based storage. Int. J. Prod. Res. **51**(6), 1635–1647 (2013)
78. Yuan, R., Li, J., Wang, W., Dou, J., Pan, L.: Storage assignment optimization in robotic mobile fulfillment systems. Complexity **2021**, 1–11 (2021)
79. Zaerpour, N., De Koster, R.B.M., Yu, Y.: Storage policies and optimal shape of a storage system. Int. J. Prod. Res. **51**(23–24), 6891–6899 (2013)
80. Zhang, G., Nishi, T., Turner, S., Oga, K., Li, X.: An integrated strategy for a production planning and warehouse layout problem: modeling and solution approaches. Omega **68**, 85–94 (2017). https://doi.org/10.1016/j.omega.2016.06.005
81. Zhang, R.-Q., Wang, M., Pan, X.: New model of the storage location assignment problem considering demand correlation pattern. Comput. Ind. Eng. **129**, 210–219 (2019)
82. Zhang, G., Shang, X., Alawneh, F., Yang, Y., Nishi, T.: Integrated production planning and warehouse storage assignment problem: an IoT assisted case. Int. J. Prod. Econ. **234**, 108058 (2021)
83. Zhou, K., Wan, W., Chen, X., Shao, Z., Biegler, L.T.: A parallel method with hybrid algorithms for mixed integer nonlinear programming. In: Computer Aided Chemical Engineering, vol. 32, pp. 271–276. Elsevier, Amsterdam (2013)

A Greenfield Analysis for Supply Chains Enhanced with Agent-Based Simulation

Eduardo García-Roa, José Antonio Marmolejo-Saucedo, Ana Paula Martínez, and Fernando Elizarrarás

Abstract This chapter presents a strategy to improve the greenfield analysis of a supply chain. The proposed strategy is based on a multiparadigm simulation model that uses discrete event simulation and agent-based simulation for vehicles used in freight transport. The modeling was carried out using Anylogic software, which is the only software that allows the integration of different simulation methods in the same project. A hypothetical case study was developed to implement the strategy.

1 Introduction

Factory location is a critical factor in the supply chain networks. Operations research (OR) is fundamental to decide a strategic location, which has been the subject of investigation of various books [1] and papers [2].

Due to the importance of facility location through time, the following characteristics should be taken into consideration: (1) understand the properties that are necessary for a facility to be acceptable in an SC context model, (2) know whether the facility fits into an SC context model, and (3) understand whether the supply chain management fits into facility location models.

Facility location faces some problems. Serving demand [2, 3] correctly is a challenge we face when locating a facility, that is why it is important to take into consideration the time, the distance, and the cost and, therefore, it is necessary to identify the facilities that are to be opened and the target customers to be served in each facility so as to minimize the costs.

E. García-Roa · A. P. Martínez · F. Elizarrarás
Universidad Panamericana, Facultad de Ingeniería, Mexico City, Mexico
e-mail: 0237168@up.edu.mx

J. A. Marmolejo-Saucedo (✉)
Facultad de Ingenieria, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico
e-mail: jose.marmolejo@fi.unam.edu

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_3

In supply chain management (SCM), the planning process focuses on finding the supply chain configuration that fits the most. Besides the strategical location of the facility, other areas such as procurement, inventories, production, and distribution should also be considered (see Cordeau et al. [4]).

In order to solve the facility location, the final site location must be selected from a group of possible alternatives. The problem can be solved by selecting *p* number of facilities to decrease the total distances or costs. The proposal of multi-period location is necessary due to the fact that parameters change depending on the time to be predectible. The objective is to align these parameters to the facilites according to the planned horizons. By that means, the consideration of planning horizon divided into time slots.

The first step to start the project is to identify potential sites, to where facilities can be studied and there is a required capacity. The location of the facility is a project that, due to the complexity and importance it represents for the companies, implies an important investment in resources, financial an time, which typically becomes a long-term project. Therefore, it can be supposed that the projects which require a new facility will operate an extended period of time. It is also important to state that changes in the environment can change the status of a location from a good one to a bad one in time. For facilities that require a small investment, such as warehouses, it makes more sense to do large changes in the configuration of facilities.

There are four basic features that can be included in the resolution of a similar model to make use of it in strategic supply chain planning: multiple commodities, multi-layer facilities, single/multiple period(s), and deterministic/stochastic parameters (see Bortolini et al. [5]).

The center of gravity analysis (or GFA, which stands for greenfield analysis) is used to select the ideal number and location of distribution facilities [6]. This type of analysis is widely used in the area of supply chain planning because of the benefits it provides in terms of possible outcomes when making the best strategic decision. The greenfield analysis gives the decision-makers the ability to:

- Select the ideal location for distribution centers as close to places of demand as possible to reduce costs
- Determine the best facility to supply to each demand point
- Maintain the number of locations at their optimum level so that they can operate at high service standards.
- Select the best location for facilities in order to minimize costs
- Have access to several scenarios simultaneously, so that they can be consulted or used in case some factors such as demand or supply changes in the future.

To obtain a result in the green field analysis it is necessary to start by structuring a model. You will need to provide information about current locations, supply and demand, products, customers, and the objective to be solved by the model. Once the structure is designed, you can upload the information in a solving tool.

Traditional methods are the ones commonly used, such is the case of the "solver" tool in microsoft excel, however, it can the counterproductive because of the time the computer will take running the simulation compared to other tools that were built for solving greenfield analysis models. The more complicated the supply chain structure is, the more sense it makes to use a GFA specialized solving tool, as the tools contains all the variables and drivers.

This type of tool can be used for the entire supply chain and can be tailored to the customer's need. The investment made here can be considerable, however, the benefits obtained will pay off by giving great value to the company in aspects such as finding the models that fits, improving the demand/supply accuracy, increased response times and flexibility in the future if needed.

2 Mathematical Model

The mathematical model used for the simulations has the ability to optimize the facilities needed to run the operation as required. These facilities can be distribution centers and factories. The model bases the decision on the locations of customers and the quantity ordered by location. This mathematical model is called a gravity model. This analysis, called the greenfield analysis, is calculated by solving the following equations for the center of gravity.

$$C_x = \frac{\sum_i d_{ix} w_i}{\sum_i w_i} \tag{1}$$

$$C_y = \frac{\sum_i d_{iy} w_i}{\sum_i w_i} \tag{2}$$

Where:

$d_{ix} = x$ coordinate of location i
$d_{iy} = y$ coordinate of location i
$w_i =$ volume for location i

The location of the facility is the first "data" that we must consider and register on the model and tool to calculate the distances between the locations of different facilities, which is advantageous compared to classic mathematical programming solutions. The center of gravity mathematical model is implicit in the solutions obtained by the Anylogistix software, which was used to validate the results obtained by the agent simulation (see Fig. 1).

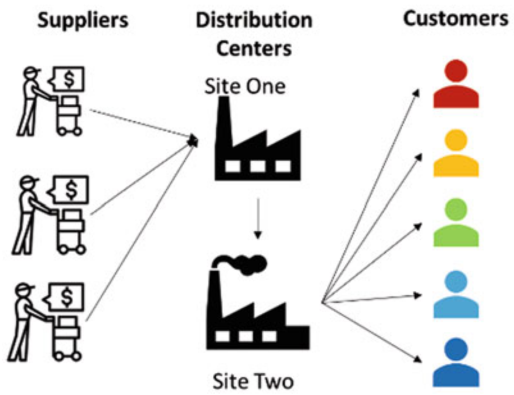


Fig. 1 Facility location in a supply chain design

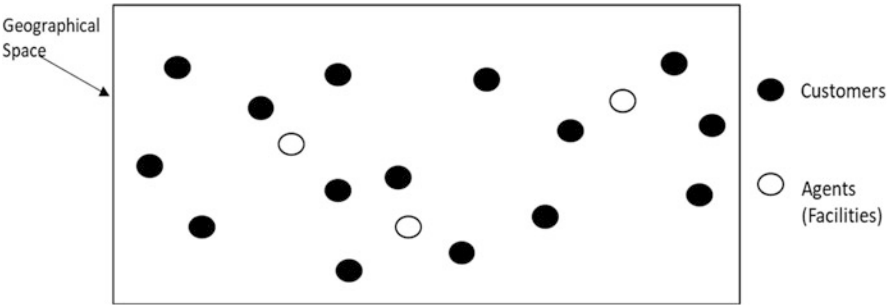


Fig. 2 Greenfield analysis in a geographical space

3 General Framework for an Agent-Based Model with Location Problems

The problems regarding facilities' location may be solved by an agent-based simulation model (see Earnest [7], Pu [8] and Pan [9]). This can be solved as a planar region, where the demand can be set as discrete points. The problem can be solved by supposing p number of facilities, with an n number of demand points. In this study, the vehicles that move merchandise between distribution centers were modeled as agents. The objective of the simulation is to define rules that the agents must follow, satisfy the demand, and solve the problem with all the specifications given. The environment is given by a space, where agents are positioned and linked by roads. The agents represent the vehicles that are in the distribution centers. These vehicles not only meet customer demand but also meet their own decision rules.

The active agents react to signals (demand orders) and DC's locations. Also, they change position depending on the demand orders. See Figs. 2, 3, 4, and 5.

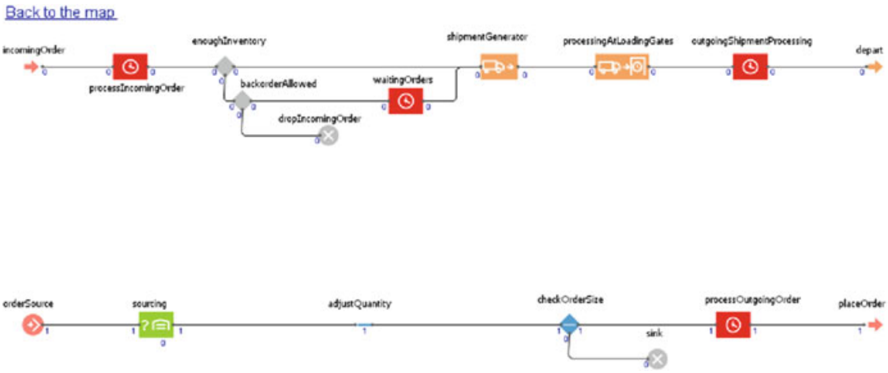


Fig. 3 Distribution center event diagram using discrete simulation

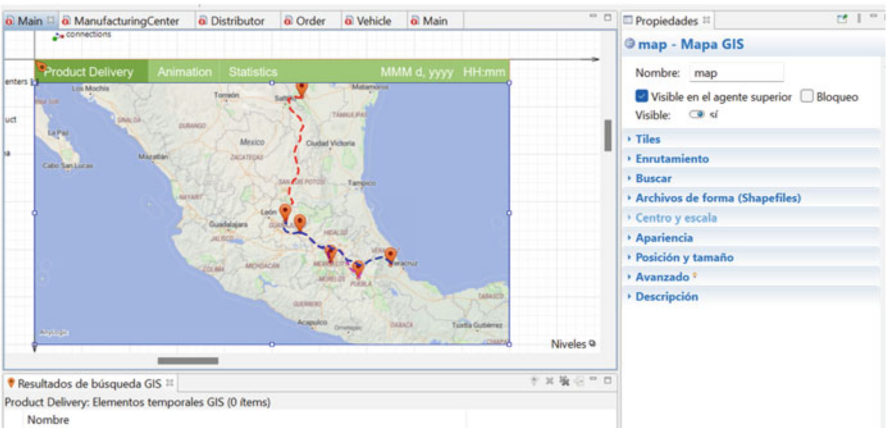


Fig. 4 Agent simulation using Anylogic-integrated georeferencing system

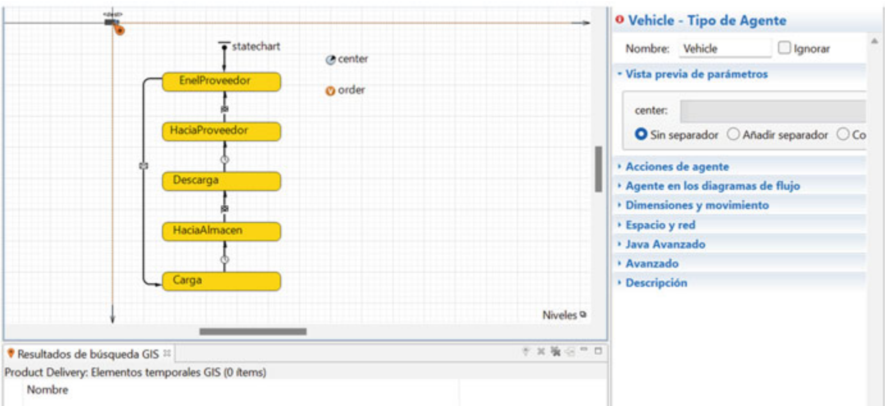


Fig. 5 State chart of vehicle agents

4 Results and Discussion

This study was designed as follows:

1. Modeling of product demand orders through discrete event simulation
2. Modeling of vehicles that transport merchandise through agent-based simulation
3. Utilization of Anylogistix to determine the best geographic locations that minimize the average distances to all customers (Figs. 6, 7, and 8)
4. Utilization of geographic information system (GIS) from Anylogic to simulate the behavior of the agents (vehicles) and validate their decision rules for choosing their trip to the distribution centers (see Figs. 3, 4, and 5). These distribution centers were previously located at Anylogistix.

The simulations show that the best alternative is to select two distribution centers. These alternative decreases both the number of vehicle trips and the distance traveled, thus minimizing transportation costs in general (see Figs. 7 and 8). The

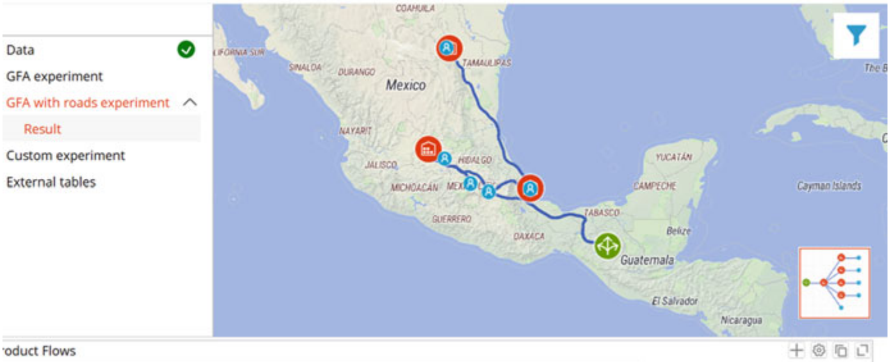


Fig. 6 3 DC – greenfield analysis using Anylogistix

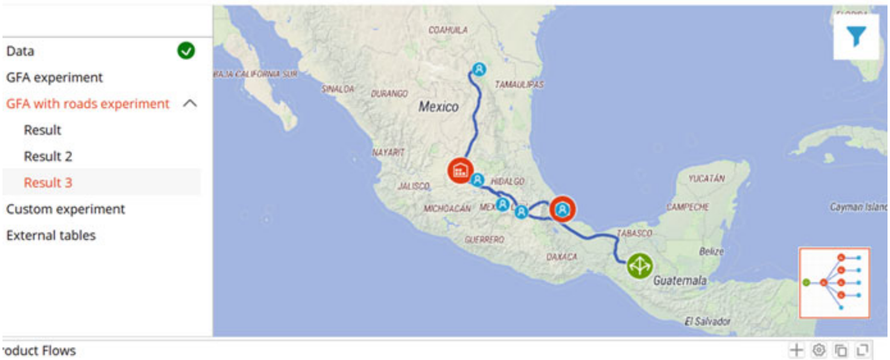


Fig. 7 2 DC – greenfield analysis using Anylogistix

Product Flows

	From	To	Product	Period	Flow, pcs	Distance, km	Flow Cost Esti...
1	Supplier	DC1	Product1	diario: 2022-0...	20	1,369.057	27,381.131
2	Supplier	DC1	Product2	diario: 2022-0...	16	1,369.057	21,904.905
3	DC1	Customer	Product1	diario: 2022-0...	10	354.41	3,544.103
4	DC1	Customer 4	Product1	diario: 2022-0...	10	150.16	1,501.604
5	DC1	Customer 4	Product2	diario: 2022-0...	6	150.16	900.962
6	DC1	Customer	Product2	diario: 2022-0...	10	354.41	3,544.103
7	Supplier	GFA Jardines d...	Product2	diario: 2022-0...	5	1,705.681	4,264.203
8	Supplier	GFA Jardines d...	Product1	diario: 2022-0...	10	1,705.681	8,528.406
9	Supplier	GFA Veracruz	Product2	diario: 2022-0...	30	738.684	11,080.267
10	Supplier	GFA Veracruz	Product1	diario: 2022-0...	40	738.684	14,773.689
11	GFA Veracruz	Customer 5	Product2	diario: 2022-0...	20	1.187	23.744
12	GFA Veracruz	Customer 2	Product2	diario: 2022-0...	10	272.669	2,726.69
13	GFA Veracruz	Customer 5	Product1	diario: 2022-0...	20	1.187	23.744
14	GFA Veracruz	Customer 2	Product1	diario: 2022-0...	20	272.669	5,453.38
15	GFA Jardines d...	Customer 3	Product1	diario: 2022-0...	10	16.844	168.441
16	GFA Jardines d...	Customer 3	Product2	diario: 2022-0...	5	16.844	84.221

Fig. 8 Distribution center location data

frequency of trips is lower than that obtained in the simulation of three distribution centers (Fig. 6).

5 Conclusions

This study presents a strategy to improve the greenfield analysis of a supply chain. The proposed strategy is based on a multiparadigm simulation model that uses discrete event simulation and agent-based simulation for vehicles used in freight transport.

The vehicles that transport merchandise are modeled as agents with decision rules based on distances and demand orders. Realistic paths were considered to minimize delivery distances and frequencies. To achieve this, the built-in GIS in Anylogic software was used. All the analyses were tested in a hypothetical supply chain located in the center of Mexico that represents the high complexity of commercial transactions typical of the region.

References

1. Chopra, S., Meindl, P., Pino, R.: Supply Chain Management. Pearson, Peru (2017)
2. Tsao, Y.-C., Chyi Lu, J.: A supply chain network design considering transportation cost discounts. Transp. Res. E Logist. Transp. Rev. **48**, 401–414 (2011)
3. ReVelle, C.S., Eiselt, H.A.: Location analysis: a synthesis and survey. Eur. J. Oper. Res. **165**(1), 1–19 (2005)

4. Cordeau, J.-F., Pasin, F., Solomon, M.M.: An integrated model for logistics network design. Ann. Oper. Res. **144**, 59–82 (2006)
5. Bortolini, M., Calabrese, F., Gabriele Galizia, F., Mora, C.: A three-objective optimization model for mid-term sustainable supply chain network design. Comput. Ind. Eng. **168**, 108131 (2022)
6. Nickel, S., Puerto, J.: Location Theory: A Unified Approach. Springer Science & Business Media (2006)
7. Earnest, D.C.: An agent-based model of the evolution of supplier networks. Comput. Math. Organ. Theory. **24**, 112–144 (2017)
8. Pu, S.: Agent-based supply chain allocation model and its application in smart manufacturing enterprises. J. Supercomput. **76**, 3188–3198 (2018)
9. Pan, A.: An agent-based negotiation model on price and delivery date in a fashion supply chain. Ann. Oper. Res. **242**, 529–557 (2013)

Resilience in Supply Chains: A Strategy Based on Inventory Policies

Rodrigo Eyssautier-Alvarado, José Antonio Marmolejo-Saucedo, and Eduardo García-Roa

Abstract The resilience of supply chains is an increasingly important factor today. The effects of the COVID-19 pandemic have highlighted the importance of generating new strategies to reduce the risk of supply chain disruption. This paper proposes an inventory policy modeled in a simulation context. The case study is a hypothetical two-stage supply chain. The solution is given by the simulation of different scenarios where different data and policies are being taking into account to see different results and discard those policies that were not useful for this case study; all the simulations are made in anyLogistix software, which is a supply chain analytics software for designing and optimizing a supply chain. The results show the performance of the proposal in several disruption scenarios similar to those experienced in the COVID-19 pandemic. The solution proposal can serve as a reference in the functioning of supply chains in the post-COVID-19 era.

Keywords Resilience in supply chains · Inventory policies

1 Introduction

A supply chain (SC) is the direct or indirect network that an organization needs in order to meet a customer's need. These are dynamic structures where the various players share information both ways. In other words, depending on the nature of the product and/or service available on the market, it may be the producer who makes the good available, or otherwise the customer makes the direct request, and from there the production process begins.

R. Eyssautier-Alvarado (✉) · E. García-Roa
Universidad Panamericana, Facultad de Ingeniería, Mexico City, Mexico
e-mail: 0199466@up.edu.mx

J. A. Marmolejo-Saucedo
Facultad de Ingenieria, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico
e-mail: jose.marmolejo@fi.unam.edu

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_4

Fig. 1 Supply chain elements

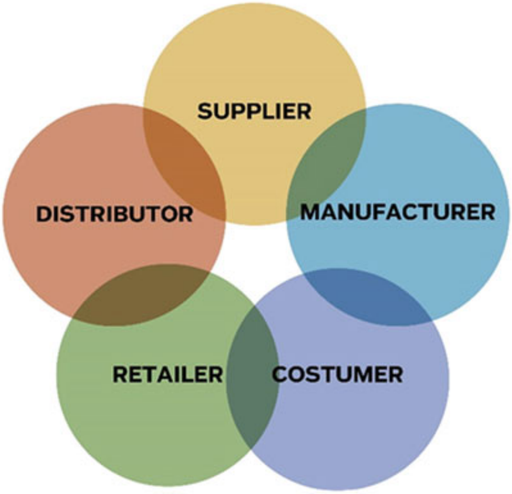


Figure 1 is a graphic example of the different ways in which a SC can be displayed:

It is important to consider that in this network not only manufacturers, suppliers, and/or carriers are the only protagonists, but the various functional areas of the organization such as sales, marketing, finance, etc. play a key role. The goal of the SC lies in maximizing the value generated by the customer, and at the same time and not least, to find a break-even point in which all the players of this SC find the desired profitability based on the revenues and costs repealed. Finally, it is important to mention that for this to be successful, it is key to adapt to changing times, because if it does not remain in force, the future of this will not be promising [1].

1.1 Role of the Coronavirus (SARS CoV-2) in the Business Environment

On December 31, 2019, in Wuhan, a city belonging to the People's Republic of China, there was a group of cases of viral pneumonia, a disease caused by the new coronavirus known as SARS-CoV-2, which affected worldwide from different perspectives [2]. Two years after its discovery and still with sequelae of this disease, this article seeks to focus on the impact it continues to have on SC in various industries.

The impact of SARS-CoV-2 on health remains critical; however, for the purposes of this publication, we feel it is important to mention the negative effects it has had on industries in general; although we have seen improvements, the risks still exist. Taking as an example China, the cradle of this pandemic and not so far from other nations, studies show that, at a macroeconomic level, levels declined from 0.4%

to 0.8%, unemployment rose by about 0.7%, and prices increased by about 0.9% compared to normal levels reported in 2020 [3].

This pandemic can be compared to major earthquakes and tsunamis, so it can be categorized as a global natural disaster. From a supply chain standpoint, it has had major repercussions on manufacturing industries. Firsthand, we saw border closures, more thorough security protocols, and labor shortages, which led to facility closures, capacity reductions, increased costs, and general uncertainty, among others, so that, to meet these challenges, these industries must seek resilient strategies focused on new consumer habits without forgetting that technology plays a very important role [4].

2 Literature Review

This research is based on literature by various authors who previously delved into the topic, whose keywords with which there is correlation are "Resilience in supply chains" and "Simulation based on agents in supply chains" taken from the scientific database Springer (https://link.springer.com) and ScienceDirect (https://www.sciencedirect.com) in publications made between 2011 and 2022 to measure the impact before, during, and after the pandemic.

All the following information and research will be helpful to design various scenarios and simulate them on the software anyLogistix (ALX), and all the results will be presented later as proposals for industries in this new era.

Supply chain design is a highly valued asset that allows companies to actively compete in today's dynamic markets, the focus continues to be on optimizing costs and achieving good environmental sustainability, the supply chain network design is an issue of a strategic nature for the management of the supply chain, it is necessary to know how to correctly establish the network of suppliers, the production and distribution centers, and all those channels that are between the distribution centers and the final customers [5].

The objective of designing optimization models for supply chain network design is to define the location, the capabilities of the facilities in terms of production, and material flow to achieve a good performance in the sustainable and economic role [5].

With the Fourth Industrial Revolution, led by Industry 4.0, companies are facing several challenges in terms of innovation and changing customer demand, forcing companies to improve their supply chain design to meet customer needs. Many of the current strategies, such as Make To Stock or Make To Order, have limitations that prevent them from meeting delivery times and cost reduction goals [5].

The analysis of the literature on optimization models for the supply chain network design affirms that despite the existence of several models and formulations referring to the design of the supply chain and its optimization, the focus is still directed to the minimization of the total costs of the network, improvement of the environmental impact, and maximizing the response capacity of the system, since there are models of single, bi-, and multi-objectives that have a traditional economic and environmental orientation [5].

The current market is very competitive and dynamic, and the interests of customers are constantly changing, which has led to SC becoming global environments, which give rise to innovation; however, the interruption of these also increases the uncertainty.

Maureen S. Golan explains the lack of literature related to resilience and the need to deepen the analysis and impact at a global level that the failure between connections and nodes has. Today studies focus on SC, excluding critical issues such as transport, command, and control networks [6].

Derived from this situation, several authors have studied, written, and proposed on the subject; such is the case of Chandra Shekhar, who firsthand proposes to analyze the resilience of these SC, against the aforementioned interruptions, and identify the key indicators that mark the course of the organization in favor of making it more solid [7].

Already with an established course, María de Arquer proposes a hybrid system in a balanced way and, through a system, takes into account both the efficiency of the system and the product or service in question. With this, she knows if the strategy implemented is adequate [8].

Times have changed, and the traditional strategies where costs were the priority have been left behind, giving way to strategies where the important thing is to be prepared for the rapid changes of trend that the market presents and scenarios in which technology will play a very important role.

Several authors have made contributions; such is the case of David C. Earnest, who pointed out that the operation of SC is in the form of networks, where various players make alliances seeking to maximize their own performance, which can increase the costs and complexity of these. The author proposes a simulation of agents using algorithms to improve the performance of these [9].

On the other hand, Zhensi Pu in his proposal adds a little more structure to this idea, identifying that the factors to consider in this simulation are a dynamic planning, the network that manages it, the flow of information between various parties, and logistics, among others [10].

Sabrina Backs, with a more technological vision, proposes this simulation in a dynamic way, multiple factors to consider in real time, generating various scenarios that help take the strategy that makes the most sense [11].

Research at a more specific level continues; such is the case of An Pan, who proposes a simulation of agents in which better communication is sought, a cooperation between the manufacturer and the supplier where in a granular way all costs and delivery times are considered in order to seek mutual benefit [12].

Ashkan Negahban proposes a simulation of agents considering several factors, where it is sought to demonstrate that long connections and the accumulation of inventories are not always the most appropriate strategy. On the other hand, analyzing these variables, a way forward can be identified to suit them [13].

R. Rajesh focuses on a resilient plant localization model, in which the following factors are taken into account: the resilience of the network and the density, complexity, and criticality of the nodes.

The author proposes a system of novel algorithms that seeks a resistant design that guarantees cost efficiency and minimizes the probability of possible interruptions due to bottlenecks [14].

Finally, with the technology that is in Vogue, Amine Belhadi proposes a model of structural equations applied to artificial intelligence, which can cope with the constant demand and provide a sustainable system to the SC [15].

2.1 Some Important KPIs to Consider

Cost savings have been one of the great objectives of companies when we talk about designing and managing their supply chain. Over the years, it has been seen that the distribution networks of distribution centers are very complex and need correct decisions for facility placement and inventory flow in the supply chain [16].

The management of the supply chain is one of the most challenging problems faced by the directors of each organization since there are substantial costs such as the transport of goods from the factories to the distribution centers, which is why this cost is an important component of the supply chain [16].

Transportation costs are essential when defining how inventory replenishment will be, as it can help reduce the total cost of the supply chain. A great example is discounts for large amounts of cargo. On the other hand, the cost of transportation is directly related to the distance traveled to deliver the goods. In the same way, the amount of fuel used depends on the distance. For this reason, the shorter the distance, the lower the transportation cost is [16].

Another key performance indicator is the lead time, which is known as that space of time that exists between when an order is placed and when it is received, to have a good relationship between customers and retailers. The lead time must be reduced as much as possible. Hence, it is said that a supply chain can become more reliable when the lead time is reduced [17].

In any supply chain system, the objective is to reduce this delivery time and maximize profits by satisfying customers. In addition, by having controlled lead times, the costs involved in the operation can be more easily estimated [17].

A standardized lead time promotes correct management of the supply chain, hence the importance of finding methods that can reduce it and establish measures to standardize it in order to satisfy the customers [17].

Additionally, if you seek to achieve and maintain a competitive advantage over the years in an existing market that is customer focused, what you have to do is to focus on improving the service level. This key performance indicator tells us the probability of satisfying the demand that comes from customers [18].

Moreover, many companies promise and claim to have a 100% service level, which is the number that every company aims to compete and win the trust of customers. However, it has been questioned if a 100% service level is really the best option for all companies. The answer can be quite complicated since when seeking to increase the level of service, several costs come into play, such as inventory costs [18].

A good level of service is key to improving the income of companies, but they are usually accompanied by a greater risk in the inventory. Even Schalit and Vermorel [19] showed that if we seek to increase the service level from 95% to 97% which is much more expensive than from 85% to 87%, this means that the increase in the level of service and the increase in costs are related. Hence, it is essential to establish an appropriate service-level objective. So the company can find the balance between customer satisfaction and final income. It is not an easy task, but the necessary strategies must be sought to achieve this goal in an environment where market demand is constantly changing [18].

3 Problem Statement and Contribution

Based on the evidence provided above, which is called and considered the state of the art of this research, we have been able to identify a common problem in all industries that want to have a competitive advantage in a dynamic market, how to have a resilient supply chain, and the capacity for resistance and recovery in the face of disruptive events such as COVID-19.

We have already analyzed how diverse and important authors expose their points of view through scientific journals, and we can assure that supply chains are vital for all types of industry, regardless of their type. For this it is essential that the design of the supply chain takes into account all the needs of the organization and finds mutual benefit between all the parts and elements of it. Every organization requires suppliers that supply the raw materials, facilities that transform the product or service, and finally a network that puts that product or service in the hands of end customers.

The market and consumer needs have changed dramatically. So traditional schemes are no longer enough. If we add to this footprint caused by COVID-19, the impact that SC has globally is important, a problem that we should not ignore in this new era.

For this reason, it is vital to move to innovative strategies rich in technology, such as agent-based simulation, which allow us to identify the needs and benefits of each one of them, through concrete and fluid communication. Today both the manufacturer, suppliers, and other entities have to cooperate for the best possible result in terms of costs and time.

This paper seeks to give as a contribution to the scientific community, a strategy based on inventory policies that will allow us to optimize key KPIs both for the customer, such as the level of service, and for the supply chain, such as production costs, transportation costs, and lead time, among others.

4 Methodology

To achieve the desired results from a technological and simulation approach, we have used the anyLogistix software, which is a tool that, through a digital twin of our organization, allows us to consider various factors in real time and create multiple scenarios in favor of make the best decisions.

5 Results and Discussion Section of Different Scenarios

With the help of the aforementioned software and for simulation purposes, the most relevant data of a fictitious company were entered which, as its name implies, is dedicated to the sale and marketing of these. It has headquarters and production plant in Germany, which sends product to four distribution centers strategically located to serve 11 customers in the main cities of the European Union (Fig. 2.)

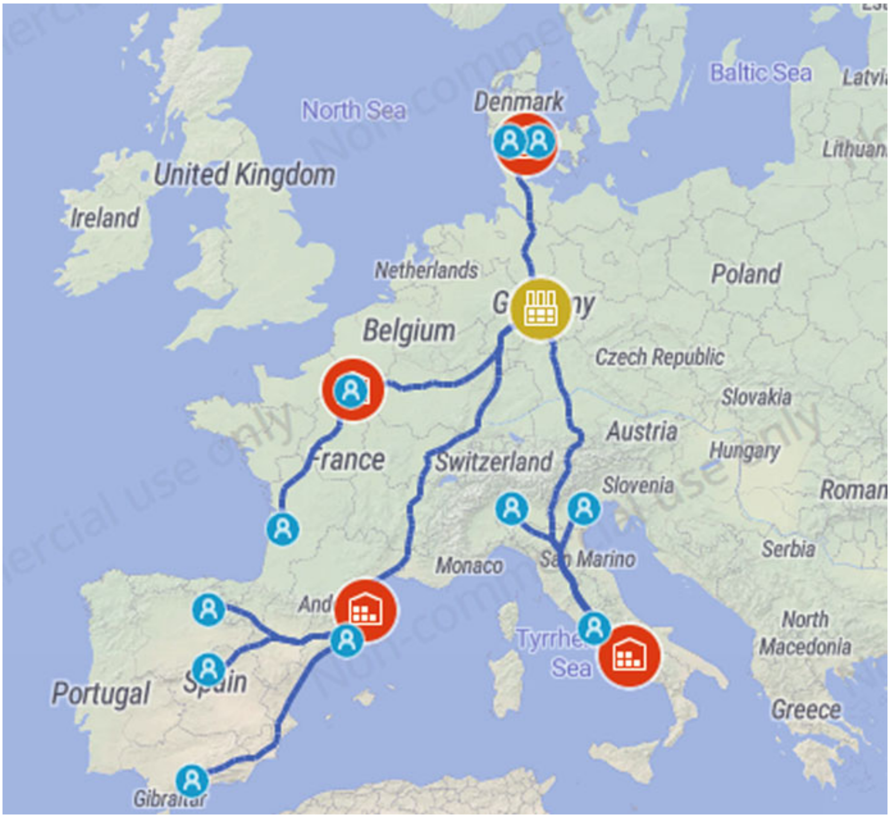


Fig. 2 Supply chain design in anyLogistix software

The first step was to establish the metrics. In this way, we would be clear about where we were going. These were delivery times, cost of existence, production costs, revenue, profit, and service level.

The second step was to identify the policies that would allow us to achieve these goals. These were sourcing and inventory.

Finally, considering the factors previously mentioned, the tool gave us the possibility of running simultaneously 19 different scenarios with which we could compare numerically and graphically the pros and cons of each one. This in order to choose the most appropriate and take it as a starting point for a new strategy.

Also, taking into account the initial scenario, the decision that has been made is to modify some important data and policies for the supply chain. This is in order to demonstrate the different results that we can obtain at the moment of simulating several of these scenarios taking into account the key performance indicators such as lead time, service level, production indicators, and all those related to the costs contained in the profit and loss statement.

It is important to mention that another objective that we seek to demonstrate is that the policies and values of the initial scenario were raised based on the design of the supply chain that we have, where we seek an optimization that gives us the best possible result with the simulation in the software called anyLogistix.

In all scenarios, a brief description of what data and policies are being changed or modified will be given, in addition to including only the KPIs that suffered a significant variation compared to the initial scenario, in order to give a general overview of the situation and focus on the relevant results and variations.

In the first case, we have a policy of minimums and maximums with security inventory in some cases (Fig. 3). Finding the perfect mix of these factors allows you to be the one that obtains the most revenue and profits. The level of service is 100% to meet the needs of end customers.

Once we found that minimums and maximums with security stock inventory policy were giving a 100% service level, several inventory policies were modified in order to see what other result we could have. It began with a simulation where

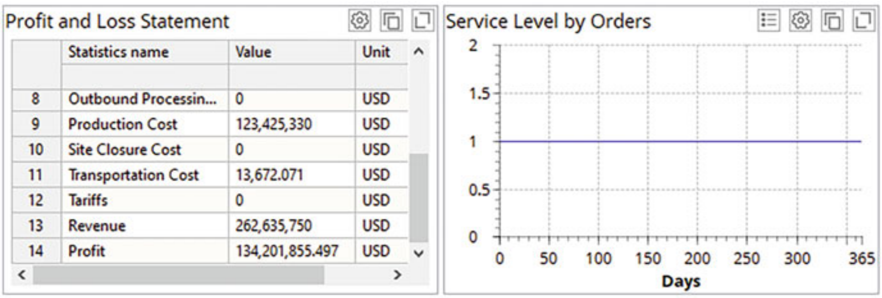


Fig. 3 Profit and loss statement report and service level by orders KPI with minimums and maximums with security stock inventory policy

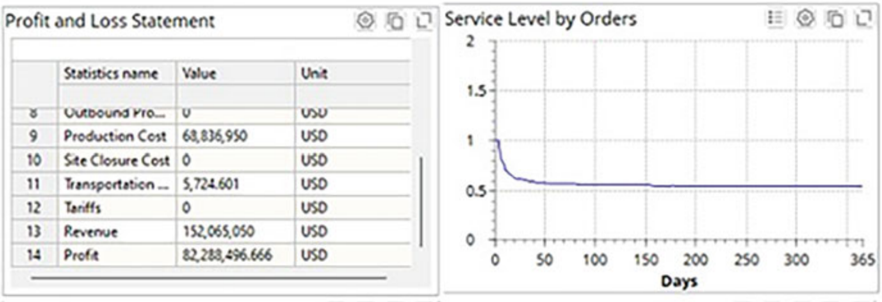


Fig. 4 Profit and loss statement report and service level by orders KPI with reorder point inventory policy

unlimited inventory was established, where we assume that the products are always in stock in any required quantity.

Clearly, when the simulation was made, several of the KPIs did not even appear in the results, since units are not produced and they are already in stock as mentioned in the previous paragraph, which greatly helps the profit and loss statement, since all the costs related to inventory production are non-existent. With this exercise the final profit went from 134,201,855.497 USD to 262,622,077.929 USD, which represents an increase of 270%.

Although the scenario is unreal, it serves and exists to see how the supply chain would work and perform without the limitations of inventories, which, at the end of the day, is what governs the pace at which the supply chain will work and the profits to be had.

In the third case, we have a reorder point policy (Fig. 4). Production costs go down by not having surplus inventory. However, we see a drop in the level of service that can be mostly attributable to not complying on time with the orders of end customers.

In the fourth case, we have a cross-dock policy (Fig. 5). Production costs are 0 when outsourced, and transport costs remain minimal. However, not being responsible for the entire operation, the service level goes 0, mostly attributable to the fact that the orders of the final customers were not fulfilled.

When using other inventory policies such as order on demand, regular policy, no replenishment, and cross-dock policy, among others, we could see that many of the important indicators for this study were severely affected, especially profit, which was negative or very low at the end of the simulation cycle in most of the policies of this type and the service level which fell between 0% and 1% levels as shown in Fig. 5, and this is only because we have an initial stock that allows us to satisfy the demand in the first days of the cycle.

With this, it can be deduced and demonstrated that for the purposes of the current exercise and with the design of the supply chain in the initial scenario, the best inventory policy is the min-max policy and min-max policy with safety stock.

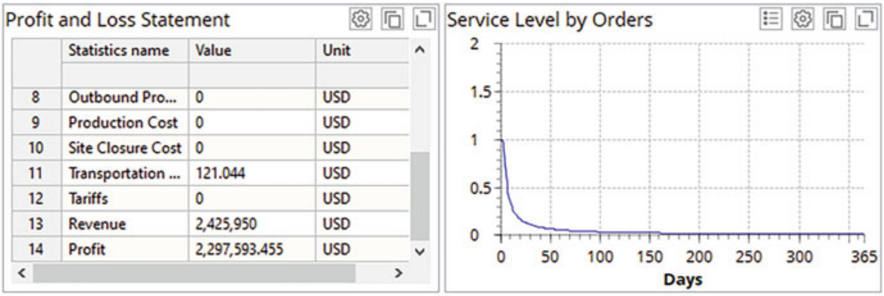


Fig. 5 Profit and loss statement report and service level by orders KPI with cross-dock inventory policy

It is important to mention that the goodness of ALX is that it allows you to run these in the future. You can predict the following exercises without having to wait for them to happen.

The fifth case consists of varying the sources policy. Originally the configuration is defined so that the factory delivers the product to the Paris DC, Barcelona DC, Italy DC, and Denmark DC. Later the distribution of the DCs to the final customers is defined to do it to customers belonging to the country of DC. For example, Paris DC will take care of delivering the product to Bordeaux and Paris, Italy DC to Milan, Rome and Venice, and so on.

What was modified is the distribution configuration of the DCs to the final clients, where any DC can provide to any final client of any country. The results when simulating this scenario were as expected. Since the lead time toward the end customer went from 1889 to 11,656 days, an increase of practically 617%, significantly affecting transportation costs that were originally at 13,672,071 USD and became 25,096.73 USD, which represented an increase of 183%, which is enough to define that this parameter of distribution of the DCs to the final clients is not optimal.

The sixth case is quite interesting. Instead of considering an initial stock in each facility of the supply chain that would allow us to have a service level of 100%, we assume that the initial stock of all the elements of the supply chain is 0 units.

The KPIs that were most affected were those of production cost, which went from 123,425,330 USD to 125,272,490 USD (an increase of 101%), and the service level (Fig. 6), where we can see that in this simulation it takes approximately the first quarter of the year to recover and obtain a level of 100%.

The time that takes the factory to produce enough stock to satisfy the demand and the DCs accumulates what is necessary to meet the requirements of the final customer, affects the service level, and will suffer a recovery quite slow depending on the production capacities that are established for the factory.

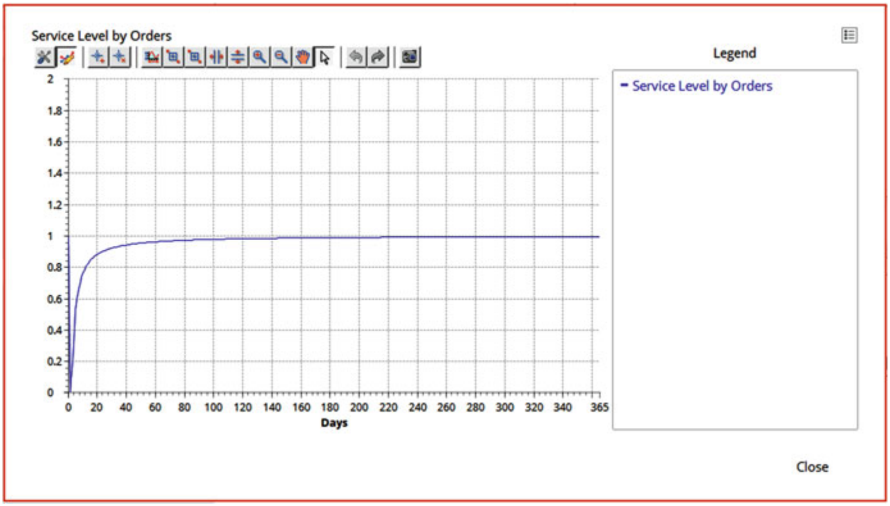


Fig. 6 Service level by orders with initial stock of 0 units

In the seventh case, it was decided to give a greater focus to the productive capacity of the plant and the inventory levels handled by each DC. Originally the values of s (minimum inventory) and S (maximum inventory) were configured to exceed what was required by each final customer as a protection measure. However, what was done was to modify these values so that $s = S$ and the quantity was exactly the daily demand that each DC had.

The result at the time of the simulation was very interesting, since the service level remained at a level of 100% as shown in Fig. 7, while in the profit and loss statement, we can see that in terms of final profit we went from an initial result of 134,201,855.497 to 141,600,129.89, which represented an increase of approximately 105%.

Where a change was perceived was in the KPI of production utilization where in the original scenario we came from a value of 1 and we stayed around ~.3. In this case it was similar only that in the first 80 days of the cycle (365 days), we are below the .3 of the initial scenario (Fig. 8), which means that there was more idle time and made the production utilization factor low in the first days of the year.

The scenario is good, and the profit is even higher than the initial one. However, we are talking about a scenario where the demand forecast is 100% accurate and there are no variations in the daily product requirements by final customers, for protection, it is always good to have that amount of inventory that allows you to face the peaks in demand that may arise along the way.

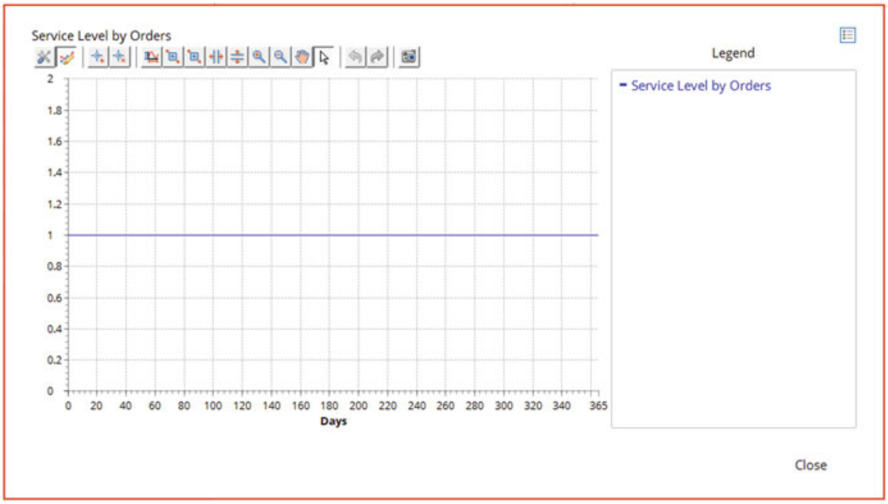


Fig. 7 Service level with exact inventory and production in factory and DCs

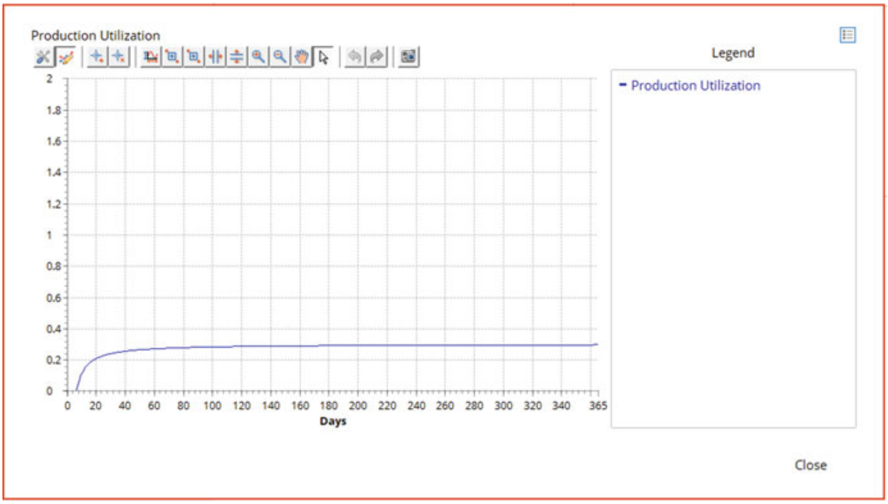


Fig. 8 Production utilization KPI with exact inventory and production in factory and DCs

6 Conclusions

We can conclude that the scientific contributions made by these and other authors are key to the transcendence of the supply chain and avoid disruptions in these that prevent continuous improvement and promote risks in different operational areas. As far as possible, organizations should invest in acquiring these tools and properly train their collaborators in order to have clear, well-established processes that are as flexible and adaptable as possible. It is not guaranteed that the use of these tools will bring us 100% safe and immediate benefits, but it will undoubtedly put us in a better position and will be a preventive measure, from which we will be more competitive, both internally and externally.

Of all the scenarios that we simulated, we were able to show that for the purposes of this paper and the perfumery supply chain, the best inventory policy is the min-max policy and min-max policy with safety stock, since in this policy where the products are ordered when the inventory level falls below a fixed replenishment point, we find the best results in terms of transportation costs, production costs, service level, lead time, and revenue, among others, that leave us being very profitable at the end of the year.

The fear that these tools will replace man is present. However, the real benefit will be that complex and repetitive tasks are left aside, giving way to more strategic tasks focused on global results.

This paper shows how the software called anyLogistix is a great tool to see and visualize different simple and complex scenarios that give expected results within a supply chain. On this occasion, we find, within our established parameters, an optimal solution of inventory policy in a supply chain context that will help the organization have resilience in his supply chain, based on an organization that produces and distributes perfume products. However, each organization must adapt its policies to the supply chain design they have and focus on which KPIs want to improve or optimize.

References

1. Chopra, S., Meindl, P., Pino, R.: Supply Chain Management. Pearson, Peru (2017)
2. World Health Organization: Obtained from https://www.who.int/es/news-room/questions-and-answers/item/coronavirus-disease-covid-19 (2021, May 13)
3. Wu, F., et al.: The impact of COVID-19 on China's regional economies and industries. J. Geogr. Sci. **31**, 565–583 (2021)
4. Kapoor, K., et al.: How is COVID-19 altering the manufacturing landscape? A literature review of imminent challenges and management interventions. Ann. Oper. Res. (2021). https://doi.org/10.1007/s10479-021-04397-2
5. Bortolini, M., Calabrese, F., Gabriele Galizia, F., Mora, C.: A three-objective optimization model for mid-term sustainable supply chain network design. Comput. Ind. Eng. **168**(4), 108131 (2022)
6. Golan, M.S., Jernegan, L.H., Linkov, I.: Trends and applications of resilience analytics in supply chain modeling: systematic literature review in the context of the COVID-19 pandemic. Environ. Syst. Decis. **40**(2), 222–243 (2020)
7. Sing, C.S.: Performance indicators for supply chain resilience: review and conceptual framework. J. Ind. Eng. Int. **15**(Suppl 1), 105–117 (2019)
8. de Arquer, M.: Examining the balance between efficiency and resilience in closed-loop supply chains. Cent. Eur. J. Oper. Res. **30**, 1307–1336 (2022)
9. Earnest, D.C., Wilkinson, I.F.: An agent-based model of the evolution of supplier networks. Comput. Math. Organ. Theory. **24**(1) (2018). https://doi.org/10.1007/s10588-017-9249-1

10. Pu, S., et al.: Agent-based supply chain allocation model and its application in smart manufacturing enterprises. J. Supercomput. **76**(11) (2018). https://doi.org/10.1007/s11227-018-2536-x
11. Backs, S., Jahnke, H., Lüpke, L., et al.: Traditional versus fast fashion supply chains in the apparel industry: an agent-based simulation approach. Ann. Oper. Res. **305**, 487–512 (2021)
12. Pan, A., Choi, T.-M.: An agent-based negotiation model on price and delivery date in a fashion supply chain. Ann. Oper. Res. **242**(2) (2013). https://doi.org/10.1007/s10479-013-1327-2
13. Negahban, A., Smith, J.S.: A joint analysis of production and seeding strategies for new products: an agent-based simulation approach. Ann. Oper. Res. **268**(3) (2018). https://doi.org/10.1007/s10479-016-2389-8
14. Rajedh, R.: Network design for resilience in supply chains using novel crazy elitist TLBO. Neural Comput. & Applic. **32**(2) (2020). https://doi.org/10.1007/s00521-019-04260-3
15. Belhadi, A., et al.: Artificial intelligence-driven innovation for enhancing supply chain resilience and performance under the effect of supply chain dynamism: an empirical investigation. Ann. Oper. Res. (2021). https://doi.org/10.1007/s10479-021-03956-x
16. Tsao, Y.-C., Chyi Lu, J.: A supply chain network design considering transportation cost discounts. Transp. Res. E: Logist. Transp. Rev. **48**(2), 401–414 (2011)
17. Dey, B.K., Bhuniya, S., Sarkar, B.: Involvement of controllable lead time and variable demand for a smart manufacturing system under a supply chain management. Expert Syst. Appl. **184**(2), 115464 (2021)
18. Chen, X., Luo, J., Wang, X., Yang, D.: Supply chain risk management considering put options and service level constraints. Comput. Ind. Eng. **140**(3), 106228 (2020)
19. Schalit, S., Vermorel, J.: Service level definition: optimizing the service levels (2014).

Part II
Computational Intelligence and Computer Sciences

Intelligent Technology in Geometric Design

Igor Litvinchev, Andrii Chuhai, Sergey Shekhovtsov, Tatiana Romanova, and Georgiy Yaskov

Abstract An integrated intelligent approach for solving geometric design problems is studied. A general optimization placement problem of objects with arbitrary shapes in a bounded container is constructed as a mathematical programming problem in terms of the phi-function technique. Various technological requirements (geometric and mechanical) are considered, including continuous translations and rotations of the objects, allowable distances between objects, prohibited zones in the container, balancing conditions, and mechanical strength constraints. Grouping the optimization placement problems based on the typology of the geometric design problems is provided. Solution strategies and various approaches to solve different variants of the model are discussed. A methodology of solving optimization placement problems is developed and illustrated with different examples.

Keywords Intelligent technology · Geometric design · Optimization placement problem · Mathematical models · Phi-functions · Optimization

1 Introduction

One of the tools for studying and optimizing complex technical systems to achieve the state of their optimal functioning is the geometric design technique aimed

---

I. Litvinchev
Graduate Program in Systems Engineering, Nuevo Leon State University (UANL), Monterrey, San Nicolas de los Garza, Nuevo Leon, Mexico

A. Chuhai · S. Shekhovtsov · G. Yaskov
Department of Mathematical Modeling and Optimal Design, Institute for Mechanical Engineering Problems, National Academy of Sciences of Ukraine, Kharkiv, Ukraine

T. Romanova (✉)
Department of Mathematical Modeling and Optimal Design, Institute for Mechanical Engineering Problems, National Academy of Sciences of Ukraine, Kharkiv, Ukraine

University of Leeds, Maurice Keyworth Building, University of Leeds, Leeds, UK

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_5

to modeling and solving optimization placement problems. These problems are referred to the creation of energy- and resource-saving technologies in a wide variety of industries (energy, machinery, ship and aircraft manufacturing, construction, chemical industry, as well as scientific research in the field of nanotechnology), biomedicine, mineralogy, material sciences, information encoding problems, image recognition systems, and spacecraft management systems when automating and modeling the processes of placement of 3D solids.

As is known, geometric design problems (including cutting, packing, layout, loading, covering) are NP-hard [1, 2]. A lot of publications address this class of problems (see, e.g., [3–17] and reference therein). A variety of methods for searching for feasible solutions are used: heuristics based on approximation approaches [18], genetic algorithms [7, 19], simulated annealing method [6], method of the artificial colony of bees [9], monkey algorithm [16], traditional optimization methods [20], and hybrid methods with heuristics and nonlinear optimization [8].

Models and approaches for the following classes of geometric design problems are developed: placement of spherical and ellipsoidal objects [21, 22]; placement polyhedral objects [23]; placement of objects bounded by spherical, conical, and cylindrical surfaces [24]; balance layout problems [25, 26]; and sparse layout problems [27, 28].

The complexity of an analytical description of the solution space and the multi-dimensionality and multi-extremality of optimization problems require the development of modern intelligent technologies to solve them effectively.

The modern method of achieving progress in solving challenging scientific problems is "artificial intelligence." Intelligent information technologies [29] allow creating intelligent systems that are able to solve challenging problems arising in scientific and industrial applications.

We propose a unique approach for mathematical and computer modeling as a powerful tool for creating an information system for automatic decision-making support in solving the geometric design problems.

2 Problem Formulation

Despite the different formulations, the geometric design problems can be described by a general statement.

General Optimization Placement Problem (GOPP). Arrange a given set of geometric objects O_i, $i \in I_n$, into a container considering the technological requirements so that an optimization criterion reaches its optimal value.

The input data for GOPP contain the following information:

- Information about a set of objects (spatial shapes: 2D or 3D, simple or composed, regular or irregular, homogeneous or heterogeneous; metric characteristics: fixed

or variable, upper and low bounds; orientation of objects: fixed or orthogonal orientation, continuous or discrete rotation; transformation type: homothetic or stretching)
- Information about a container (spatial shapes: 2D or 3D, simply connected or disconnected; prohibited zones; division into subcontainers; metric characteristics: fixed or variable, upper and low bounds)
- Technological restrictions [geometric restrictions: containment of objects in container (packing and layout problems), non-overlapping conditions for objects (packing and layout problems), container covering conditions (covering problems), minimum allowable distances between objects (sparse packing and layout problems), upper and low bounds of rotation parameters of objects (packing and layout problems); mechanical conditions: restrictions on behavior system (balance layout problems), mechanical strength conditions (topology optimization of the object), influence of physical fields (thermodynamic problems)]
- Type of an optimization criterion (maximization of the number of objects, maximization of metrical characteristics of objects, maximization of the minimum of distances between objects, minimization of metric characteristics of a container, multi-criteria optimization)

According to the objective, GOPP can be classified as a knapsack problem with output maximization or a problem with minimization an open dimension of the container [30]. In the first type of problem, some geometric objects should be packed into a container, maximizing their value. Problems of the second type pack a set number of objects into an optimized container.

We use a special class of point sets, phi-objects, as a mathematical models of real parts. By definition [31] a phi-object is a non-empty object $O \subset R^d$, $d = 2$, 3 which satisfies the following conditions: O is a canonically closed point set; the homotopic type of int(O) coincides with the homotopic type of cl(O)). Here int(O) is the interior, and cl(O) is the closure of O.

For example, the 3D objects under consideration can belong to three groups. Convex objects bounded by cylindrical, conical, and spherical surfaces belong to the first group. Arbitrary-shaped objects approximated by polyhedral. Ellipsoids together with spheroids form the third group (Fig. 1).

Objects can be subjected to congruent, homothetic, and stretching transformations. The vector of variables $u_o = (v_o, \theta_o, \lambda_o)$ is associated with the object $O \subset R^d$, $d = 2$, 3, where v_o is a translation vector, θ_o is a vector of rotation parameters, and λ_o is a scaling parameter. We denote the object O translated by the vector v_o, rotated by angles θ_o, subject to homothetic transformation with parameter λ_o, as $O(u) = \left\{p : p = v_o + \lambda_o \cdot M(\theta_o) \cdot \tilde{p},\ \ \forall \tilde{p} \in O\right\}$.

We consider the following restrictions on placement of objects: orientation, allowable distances, prohibited zones, and balancing conditions.

The set of objects are placed into the container Ω having different shapes, such as a sphere, a cuboid, and prisms with prohibited zones (see Fig. 2).

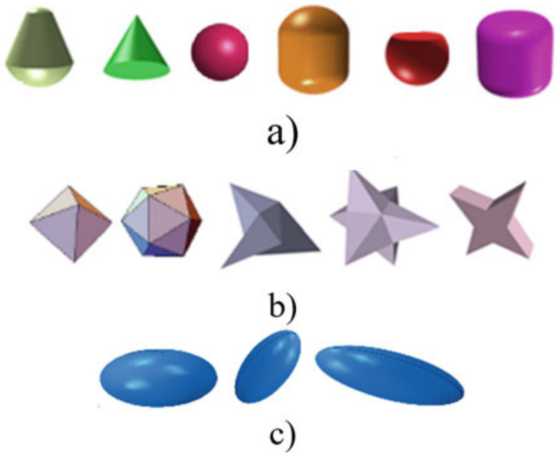


Fig. 1 Objects: (**a**) group 1, (**b**) group 2, and (**c**) group 3

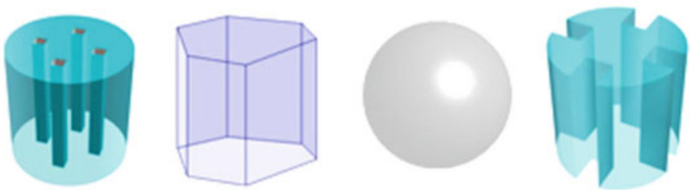


Fig. 2 Types of containers

Problem GOPP can be aimed at optimizing the container or object metric characteristics, minimizing the container volume, maximizing the minimal distance between objects, and maximizing the number of objects to be placed into a given container.

3 The Phi-Function Technique

We describe non-overlapping and containment constraints by means of the phi-function approach that is used for mathematical modeling of interactions between geometric objects. Phi-functions allow us to present the placement problem as a mathematical programming model.

Let O_1 and O_2 be the two arbitrary objects. The location of the object O_1 is specified by $u_1 = (v_1, \theta_1)$, where $v_1 = (x_1, y_1)$, θ_1 contains rotation parameters.

Three cases are monitored: O_1 and O_2 do not overlap; O_1 and O_2 are in contact; and O_1 and O_2 have common interior points.

We refer a function $\Phi(u_1, u_2)$ to a phi-function for O_1 and O_2 [32] if

$\Phi(u_1, u_2)$ is a continuous and everywhere-defined function in each variable and
$\Phi(u_1, u_2) < 0$, if int $O_1 \cap$ int $O_2 \neq \varnothing$;
$\Phi(u_1, u_2) = 0$, if int $O_1 \cap$ int $O_2 = \varnothing$ and fr $O_1 \cap$ fr $O_2 \neq \varnothing$;
$\Phi(u_1, u_2) > 0$, if cl$O_1 \cap$ cl$O_2 = \varnothing$.

Note that non-negative values of a phi-function provides the non-overlapping of O_1 and O_2. For describing the containment condition, $O_1 \subset O_2 \iff$ int $O_1 \cap O_2^* = \varnothing$, a phi-function for objects is used, where $O_2^* = R^d \backslash$ int O_2.

For some complex shapes, we use a quasi-phi-function. We say that a function $\Phi^{'}(u_1, u_2, u^{'})$ is a quasi-phi-function for O_1 and O_2 [33] if a maximum of the function in auxiliary variables $u^{'}$ ($\max_{u'} \Phi'(u_1, u_2, u')$) is a phi-function for O_1 and O_2.

Therefore $\Phi^{'}(u_1, u_2, u^{'}) \geq 0$ implies int $O_1 \cap$ int $O_2 = \varnothing$ for some $u^{'}$.

4 General Model of GOPP

We introduce the following variables of the optimization placement problem:

- u_Ω is a vector of variable dimensions of Ω.
- n_Ω is the number of variable dimensions of Ω.
- $u = (u_1, u_2, \ldots, u_n)$ contains variable motion vectors of all objects O_i, $i \in I_n$.
- $u_i = (v_i, \theta_i, g_i)$ is a variable motion vector of O_i.
- v_i is a translation vector of O_i.
- θ_i contains rotation angles of O_i.
- $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ contains variable scaling parameter of objects O_i, $i \in I_n$.
- $u^{'} = (u'_{ij}, i \in I_n, j \in I_n, i > j)$ is a vector of additional variables used for deriving quasi-phi-functions for objects $O_i(u_i)$ and $O_j(u_j)$.

In terms of phi-functions, a general model of GOPP can be presented as follows:

$$\kappa\left(X^*\right) = \min_{X \in W} \kappa(X) \tag{1}$$

subject to

$$W = \left\{X = \left(u_\Omega, u, u', \lambda\right) \in \mathbf{R}^\sigma : \Psi_s(X) \geq 0, s = 1, \ldots, 5\right\}. \tag{2}$$

In the model (1) and (2):

- $\kappa(X)$ is at least a twice-differentiable function.
- $\Psi_1(X) \geq 0$ describes containment condition $O_i \subset \Omega$ for all objects

$$\Psi_1(X) = \min\left\{\Phi_i(X), i \in I_n\right\}.$$

$\Phi_i(X)$ is a phi-function/a quasi-phi-function for O_i and the object $\Omega^* = R^d \backslash$ int Ω.

- $\Psi_2(X) \geq 0$ describes distance (non-overlapping) constraints for all pairs of objects $O_i(u_i)$ and $O_j(u_j)$

$$\Psi_2(X) = \min\left\{\Phi_{ij}(X), i \in I_n, j \in I_n, i < j\right\}.$$

$\Phi_{ij}(X)$ is the phi-function (or quasi-phi-function) for the objects $O_i(u_i)$ and $O_j(u_j)$.

- $\Psi_3(X) \geq 0$ describes distance conditions for objects $O_i(u_i)$, $i \in I_n$, and prohibition zones O_k, $k \in I_z$,

$$\Psi_3(X) = \min\left\{\Phi_{ik}(X), i \in I_n, k \in I_z\right\}.$$

$\Phi_{ik}(X)$ is a phi-function/a quasi-phi-function for $O_i(u_i)$ and O_k.

- $\Psi_4(X) \geq 0$ describes additional constraints (restrictions on the metric characteristics of the container and/or placement objects).
- $\Psi_5(X) \geq 0$ describes some mechanical conditions.

Realizations of the general model (1) and (2) can be constructed depending on the type of objective function and the system of constraints in (2). It covers a wide spectrum of applications.

Below we provide some basic characteristics of the model (1) and (2):

- The problem (1) and (2) is a nonlinear programming model that involves all global solutions of GOPP.
- Each inequality $\Psi_2(X) \geq 0,\ \Psi_3(X) \geq 0$ can be defined by an inequality system with continuously differentiable functions using phi-functions or an inequality system using quasi-phi-functions.
- A set of feasible solutions W is defined by a system of inequalities with nonsmooth functions that involve finite operators "max" and "min." Therefore $W = \bigcup_{l=1}^{m} W_l$, where W_l is defined by an inequality system with smooth functions. The model of GOPP can be rewritten as

$$\kappa\left(X^*\right) = \text{extr}\left\{\kappa\left(X^{*l}\right), l = 1, \ldots, m\right\}$$

where

$$\kappa\left(X^{*l}\right) = \underset{X \in W_l}{\text{extr}}\, \kappa(X).$$

On the ground of the model characteristics above, a methodology for solving GOPP is proposed.

5 Methodology of Solving GOPP

The methodology of solving GOPP is based on the analysis of information on objects, the container, the technological restrictions, the optimization criteria discussed in Sect. 2, and the characteristics of the mathematical model (1) and (2) provided in Sect. 4.

The fundamental components of the methodology structure are shown in Fig. 3.

The structure can be considered as a foundation of creating an intelligent system for solving different variants of GOPP.

Figure 4 illustrates input and output information in the intelligent system.

It easily sees that the input data for the system contain the information on objects being placed, a container, technological constraints, and an objective function. As a result, the system outputs a result of solving the GOPP of geometric design.

The input information on the problem differs by option of changing the orientation of objects. Allowing orientation of the objects makes this process much more complex and requires other approaches, including construction of feasible starting

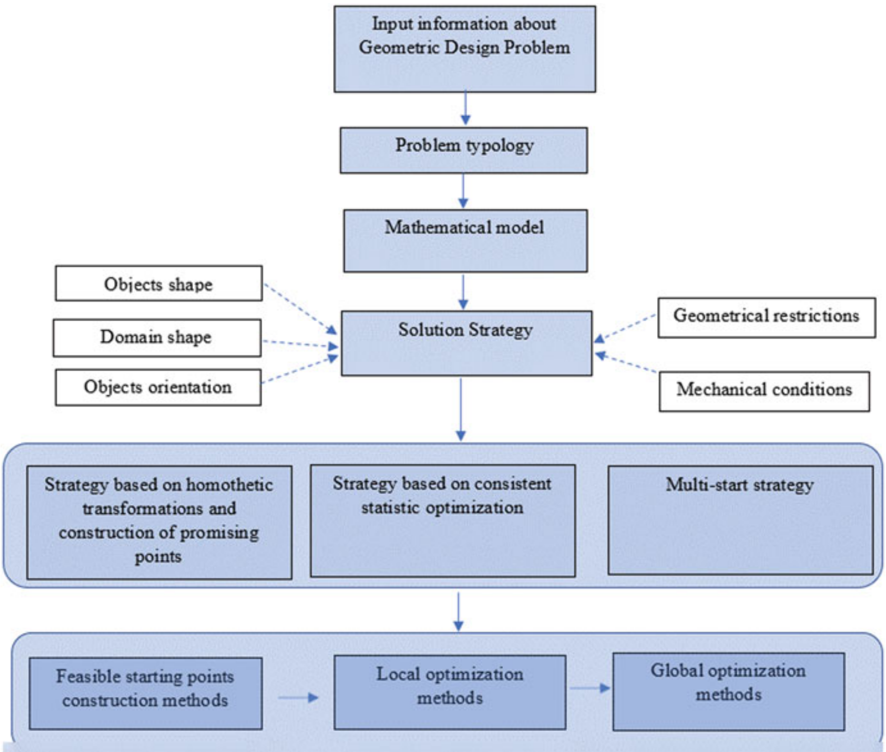


Fig. 3 Fundamental components of the methodology structure

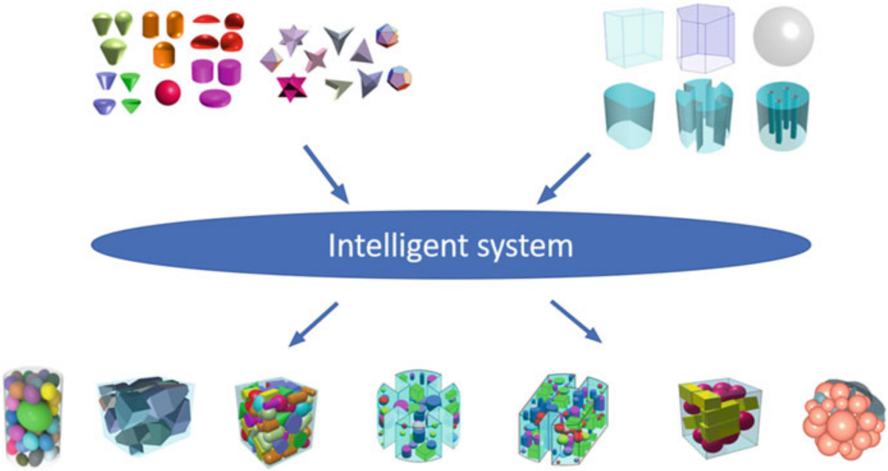


Fig. 4 Problem information

placements. The methodology involves two basic solution approaches for convex objects and rotating objects.

In the case of rotating objects, the approach suggests two strategies to search for an approximation to a global solution depending on the shape of objects. For convex objects, a homothetic transformation strategy and search of special points is used. When packing non-convex objects, the complexity of the problem increases dramatically. Therefore, a multi-start strategy is applied. Firstly, the strategy packs rotating convex objects.

Each of the proposed strategies includes the following stages:

- Construction of feasible points (the block-coordinate descent method, the homothetic transformation strategy, the regular placement method based on grids, clusterization of objects)
- Local optimization (the feasible direction method combined with the ε-active restriction strategy and the interior point solver (IPOPT) [34] together with the special decomposition strategy)
- Global optimization (the neighborhood search method (NSM), the "object-swapping" method, and the multi-start method)

For each of the strategies, different solution methods are developed.

5.1 *Neighborhood Search Strategy*

A strategy based on NSM [35] is applied to find an approximation to the global extremum of GOPP.

NSM is based on a directed random search on a set of permutations using the properties of a probabilistic distribution of local extrema of the objective function of GOPP.

We associate a tuple of the metric characteristics with each object. Then a set of the tuple permutations are formed for all objects. The Euclidean metric on the permutation set is defined. We introduce a neighborhood on a set of permutations. Then searching for the best value of the objective function of GOPP is carried out for the set of neighborhoods. A center and a radius of a new neighborhood are selected at each step of NSM. The radius of the neighborhood is reduced if there is no improvement of the objective function value.

Selection of promising permutations (centers of new neighborhoods) is based on the probability of obtaining smaller values of the objective function. A hypothesis about the distribution law of the objective function values at local minima is defined for calculating the probability. We assume that each local extremum is a random value that depends on the permutation of arranged objects. Computational experiments have shown that the distribution of the objective function values for GOPP is close to the normal.

Each promise permutation depends on the sample mean and the sample variance of the objective function values for the neighborhood under consideration.

The described approach to managing neighborhoods on a set of permutations ensures the convergence of the algorithm.

We use the sequential addition technique based on the block-coordinate descent method [36] for generating starting feasible points.

A modification of the feasible direction method [37] combined with the active inequalities approach is used for local optimization.

5.2 Homothetic Transformation Strategy

The strategy includes the following steps: constructing feasible starting points using a homothetic transformation method, searching for local extrema using IPOPT combined with the decomposition strategy, and finding an approximation to the global extremum applying the “object-swapping” technique.

The approach for generating feasible starting points uses variable metric characteristics of objects. Then, a random choice of placement parameters of scaling objects inside a container is performed. A nonlinear programming problem maximizing the sizes of scaling objects is solved using IPOPT. If all obtained sizes of objects become original and the objects fit to the container, then the found placement parameters of the objects can be considered as a starting point to search for a local extremum of GOPP.

For searching for a local extremum of GOPP, we apply the following technique. Since the set of feasible solutions is determined by a large number of variables and constraints, a special decomposition procedure is developed. The iterative procedure allows reducing the runtime by decreasing the number of the nonlinear inequalities.

Thanks phi-functions, the set of feasible solutions for GOPP is presented as a union of subregions. At each step of the decomposition procedure, a subregion is selected, and then subsets of the subregion are generated using feasible starting points. Additional linear constraints on the arrangement parameters of each object are added to the selected inequality system. The placement objects are allowed to move within the "individual" box. If the boxes do not overlap each other, we remove the appropriate phi- inequalities from the inequality system. This technique allows us to reduce the number of nonlinear constraints and additional variables in quasi-phi-functions. To construct a starting feasible point for the next iteration, we use the local solution of the previous subproblem.

In order to define an approximation to the global extremum, we use an approach for constructing a set of "promising placements" (promising starting points). We classified all objects in two groups.

The first group involves objects that can be potentially extended due to "surrounded free space."

The second group of objects comprises those that cannot be extended due to a lack of available surrounding space.

We choose the object of the first group and find a larger object from the second group.

We then swap the objects, ensuring that the larger one is resized to match the dimensions of the smaller object.

In the object-swapping method, an auxiliary NLP problem maximizing variable homothetic coefficients of objects is solved. As a result, sizes of some objects are reduced and some enlarged. A series of promising starting points are constructed to solve the auxiliary problem. According to the given sequence, we change over the objects of two groups. This procedure terminates when all objects of original sizes without overlapping fit to the container. This allows to find a feasible starting point from a new feasible subregion of GOPP. If recovering the original objects is successful, then the corresponding local solution is taken as a starting point for searching for a new local extremum of GOPP.

5.3 Irregular Packing Strategy

The strategy is used for optimized packing irregular objects. To reduce runtime, we divide the solution process into two stages.

At the first stage, clusterization of a pair of objects is applied for constructing feasible starting points of GOPP. The problem of minimizing the volume of the enclosing spherical/cuboidal volume for two irregular objects is solved. Local optimization is performed using IPOPT.

At the second stage, a set of starting points for GOPP are generated, and then corresponding local minima are calculated. Three approaches are proposed depending on the obtained cluster shapes: packing cuboids with orthogonal rotations, packing spheres, and packing spheres and rotated cuboids. To solve these

Fig. 5 Clustering objects

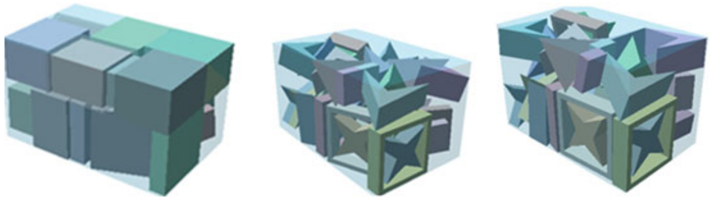


Fig. 6 Stages of generating starting point using the clusterization method

problems, the homothetic transformation method combined with construction of "object swapping" is used.

The best local solution obtained is taken as a solution of GOPP.

5.4 Methods for Constructing Feasible Starting Points

We consider the following approaches: the lattice placement method for packing congruent objects, the homothetic transformation method for packing convex objects, and the clusterization method for non-convex objects (see Figs. 5 and 6).

5.5 Local Optimization Methods

We decompose the large-scale problem into a sequence of smaller problems. The next steps are performed: generation of subsets of the set of feasible solutions, choice of a subsystem of ε-active constraints, search of local extrema in the selected subsets, and purposeful transition to another feasible subset.

Figure 7 depicts packings corresponding to starting points and appropriate local minima for two problems: packing convex objects bounded by cylindrical, conical, and spherical surfaces into minimum-height cuboid (Fig. 7a); packing irregular polyhedral objects into a minimum-volume cuboid (Fig. 7b); and packing ellipsoids into a minimum-volume cylinder (Fig. 7c).

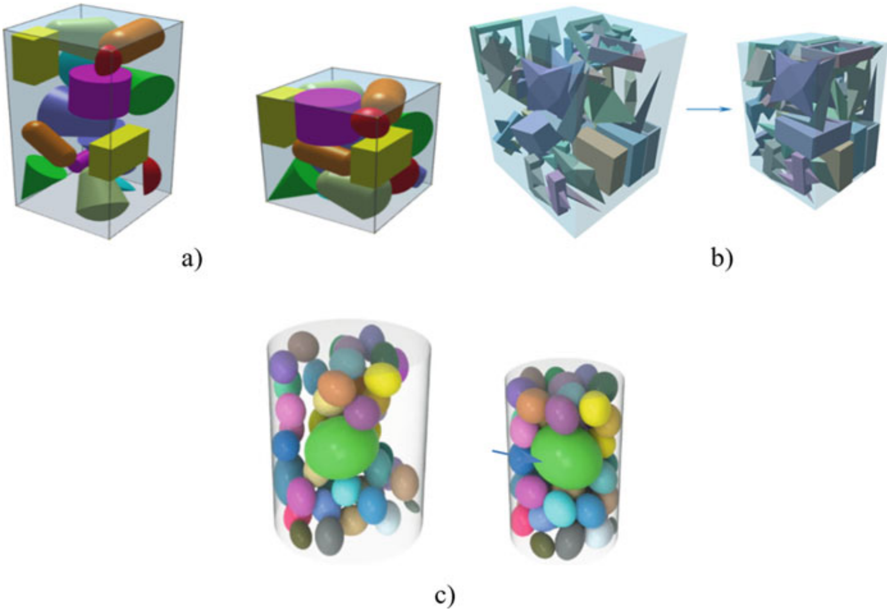


Fig. 7 Local optimization: (**a**) for the first group of objects, (**b**) for the second group of objects, and (**c**) for the third group of objects

6 Conclusions

This paper proposes an intelligent optimization technology based on a unique approach to mathematical and computer modeling for GOPP.

The methodology uses advanced tools of mathematical modeling and optimization methods of geometric design and state-of-the-art software packages to solve NP-hard placement problems. Efficiency of our algorithms is supported by computational results compared with those published.

The methodology is centered around applying cutting-edge mathematical and computational techniques involving a local NLP solver for real-world problems.

Optimized packing models considered in this work are referred to large-scale NLP. Construction of convex hulls for clustering composed objects [38, 39] as well as aggregation/decomposition techniques are aimed to use a special structure of the problem and provide a reasonable alternative to direct solution approaches [40, 41].

Future research in this area could focus on further developing and refining these techniques, exploring new applications and domains, and investigating the potential benefits and limitations of using these tools in different contexts. Additionally, there may be opportunities to integrate these tools with other emerging technologies, such as artificial intelligence and machine learning, to create even more powerful problem-solving systems.

Acknowledgments This work is supported by the National Research Foundation of Ukraine (grant #02.2020/167), Volkswagen Foundation (grant #97775), and British Academy (grant #100072).

References

1. Chazelle, B., Edelsbrunner, H., Guibas, L.J.: The complexity of cutting complexes. Discrete Comput. Geom. **4**(2), 139–181 (1989)
2. Lenstra, J.K., Rinnooy, A.H.G.: Complexity of packing, covering, and partitioning problems. In: Schrijver, A. (ed.) Packing and Covering in Combinatorics, pp. 275–291. Mathematisch Centrum, Amsterdam (1979)
3. Stillinger, F.H., Lubachevsky, B.D.: Crystalline-amorphous interface packings for disks and spheres. J. Stat. Phys. **73**, 497–514 (1993)
4. Yaskov, G.N.: Methodology to solve multi-dimensional sphere packing problems. J. Mech. Eng. Problemy Mashynobuduvannia. **22**(1), 67–75 (2019)
5. Pankratov, A.V., Romanova, T.E., Chugay, A.M.: Optimal packing of convex polytopes using quasi-phi-functions. J. Mech. Eng. Problemy Mashynobuduvannia. **18**(2), 55–65 (2015)
6. Cao, P., Fan, Z., Gao, R., Tang, J.: Complex housing: modeling and optimization using an improved multi-objective simulated annealing algorithm. In: Proceedings the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. 2B: 42nd Design Automation Conference, Charlotte, North Carolina, USA. August 21–24, 2016. V02BT03A034. ASME (2016). https://doi.org/10.1115/DETC2016-60563
7. Egeblad, J., Nielsen, B.K., Odgaard, A.: Fast neighborhood search for two- and three-dimensional nesting problems. Eur. J. Oper. Res. **183**(3), 1249–1266 (2007)
8. Fasano, G.: A global optimization point of view to handle non-standard object packing problems. J. Glob. Optim. **55**(2), 279–299 (2013)
9. Li, G., Zhao, F., Zhang, R., Du, J., Guo, C., Zhou, Y.: Parallel particle Bee Colony Algorithm approach to layout optimization. J. Comput. Theor. Nanosci. **13**(7), 4151–4157 (2016)
10. Kallrath, J.: Packing ellipsoids into volume-minimizing rectangular boxes. J. Glob. Optim. **67**(1–2), 151–185 (2017)
11. Kampas, F.J., Castillo, I., Pintér, J.D.: General ellipse packings in optimized regular polygons. http://www.optimization-online.org/DB_HTML/2016/03/5348.html (2016)
12. Duriagina, Z., Lemishka, I., Litvinchev, I., et al.: Optimized filling of a given cuboid with spherical powders for additive manufacturing. J. Oper. Res. Soc. China. **9**, 853–868 (2021)
13. Litvinchev, I., Ozuna, L.: Integer programming formulations for approximate packing circles in a rectangular container. Math. Probl. Eng. **2014**, 317697, 1–6 (2014)
14. Pankratov, A., Romanova, T., Litvinchev, I.: Packing ellipses in an optimized rectangular container. Wireless Netw. **26**, 4869–4879 (2020)
15. Stoian, Y.E., Chugay, A.M., Pankratov, A.V., Romanova, T.E.: Two approaches to modeling and solving the packing problem for convex polytopes. Cybern. Syst. Anal. **54**(4), 585–593 (2018)
16. Torres-Escobar, R., Marmolejo-Saucedo, J.A., Litvinchev, I.: Binary monkey algorithm for approximate packing non-congruent circles in a rectangular container. Wirel. Netw. **26**(7), 4743–4752 (2020)
17. Wang, Y., Lin, C.L., Miller, J.D.: Image segmentation for analysis of multisize particles in a packed particle bed. Powder Technol. **301**, 160–168 (2016)
18. Litvinchev, I., Ozuna, L.: Approximate packing circles in a rectangular container: valid inequalities and nesting. J. Appl. Res. Technol. **12**(4), 716–723 (2014)

19. Xiao, W.S., Wu, L., Tian, X., Wang, J.L.: Applying a new adaptive genetic algorithm to study the layout of drilling equipment in semisubmersible drilling platforms. Math. Probl. Eng. **2015**, 146902, 1–9 (2015). https://doi.org/10.1155/2015/146902
20. Romanova, T., Stetsyuk, P., Chugay, A., Shekhovtsov, S.: Parallel computing technologies for solving optimization problems of geometric design. Cybern. Syst. Anal. **55**(6), 894–904 (2019)
21. Stoyan, Y., Yaskov, G., Romanova, T., Litvinchev, I., Yakovlev, S., Cantú, J.M.V.: Optimized packing multidimensional hyperspheres: a unified approach. Math. Biosci. Eng. **17**(6), 6601–6630 (2020)
22. Romanova, T., Litvinchev, I., Pankratov, A.: Packing ellipsoids in an optimized cylinder. Eur. J. Oper. Res. **285**, 429–443 (2020). https://doi.org/10.1016/j.ejor.2020.01.051
23. Romanova, T., Bennell, J., Stoyan, Y., Pankratov, A.: Packing of concave polyhedra with continuous rotations using nonlinear optimisation. Eur. J. Oper. Res. **268**(1), 37–53 (2018)
24. Stoyan, Y.G., Semkin, V.V., Chugay, A.M.: Modeling close packing of objects. Cybern. Syst. Anal. **52**(2), 296–304 (2016)
25. Stoyan, Y., Romanova, T., Pankratov, A., Kovalenko, A., Stetsyuk, P.: Balance layout problems: mathematical modeling and nonlinear optimization. In: Fasano, G., Pintér, J.D. (eds.) Space Engineering. Springer Optimization and Its Applications, vol. 114, pp. 369–400. Springer, Cham (2016)
26. Romanova, T., Stoyan, Y., Pankratov, A., Litvinchev, I., Tsegelnyk, Y., Shypul, O.: Sparsest balanced packing of irregular objects in a cylindrical container. Eur. J. Oper. Res. **291**(1), 84–100 (2021)
27. Romanova, T., Stoyan, Y., Pankratov, A., Litvinchev, I., Avramov, K., Chernobryvko, M., Yanchevskyi, I., Mozgova, I., Bennell, J.: Optimal layout of ellipses and its application for additive manufacturing. Int. J. Prod. Res. **59**(2), 560–575 (2021)
28. Romanova, T., Pankratov, A., Litvinchev, I., Plankovskyy, S., Tsegelnyk, Y., Shypul, O.: Sparsest packing of two-dimensional objects. Int. J. Prod. Res. **59**(13), 3900–3915 (2021)
29. Pashchenko, F.F., An, B.T., Hieu, T.D., Pashchenko, A.F., Van Trong, N.: Intelligent technologies in decision-making support systems. In: Proceedings of 2020 International Conference Engineering and Telecommunication (En&T), 2020, pp. 1–4 (2020). https://doi.org/10.1109/EnT50437.2020.9431248
30. Wäscher, G., Haußner, H., Schumann, H.: An improved typology of cutting and packing problems. Eur. J. Oper. Res. **183**(3), 1109–1130 (2007)
31. Stoyan, Y.G.: Mathematical methods for geometry design. In: Ellis, T.M.R., Semenkoc, O.J. (eds.) Advances in CAD/CAM, vol. 1982, pp. 67–86, Amsterdam (1983)
32. Stoyan, Y., Pankratov, A., Romanova, T.: Placement Problems for Irregular Objects: Mathematical Modeling, Optimization and Applications, vol. 130, pp. 521–559. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68640-0_25
33. Stoyan, Y., Pankratov, A., Romanova, T.: Quasi-phi-functions and optimal packing of ellipses. J. Glob. Optim. **65**(2), 283–307 (2016)
34. Wächter, A., Biegler, L.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. **106**, 25–57 (2006)
35. Stoyan, Y.G., Socolovsky, V.Z.: The minimization method for some permutation functionals. Inf. Process. Lett. **8**(2), 110–111 (1979)
36. Nishijima, M., Nakata, K.: A block coordinate descent method for sensor network localization. Optim. Lett. **16**, 1051–1071 (2022)
37. Chen, X., Kostreva, M.M.: Methods of feasible directions: a review. In: Yang, X., Mees, A.I., Fisher, M., Jennings, L. (eds.) Progress in Optimization. Applied Optimization, vol. 39. Springer, Boston (2000)
38. Kallrath, J., Romanova, T., Pankratov, A., Litvinchev, I., Infante, L.: Packing convex polygons in minimum-perimeter convex hulls. J. Glob. Optim. **85**(1), 39–59 (2023). https://doi.org/10.1007/s10898-022-01194-4

39. Romanova, T., Pankratov, A., Litvinchev, I., Dubinskyi, V., Infante, L.: Sparse layout of irregular 3D clusters. J. Oper. Res. Soc. **74**(1), 351–361 (2023). https://doi.org/10.1080/01605682.2022.2039568
40. Litvinchev, I.: Decomposition-aggregation method for convex programming problems. Optimization. **22**(1), 47–56 (1991)
41. Litvinchev, I., Rangel, S.: Localization of the optimal solution and a posteriori bounds for aggregation. Comput. Oper. Res. **26**(10–11), 967–988 (1999)

An Information Architecture for the Engineering and Design of Industrial Electrical Systems

Francisco de Asis López-Fuentes

Abstract Electrical systems are one of the most important and necessary facilities that an industrial plant has. An industrial electrical system consists of various components such as conductors or cables, connectors, switches, and electrical circuits. During the design of an industrial electrical system, engineers and designers require a great deal of technical information. It is a tedious process for an electrical engineer because of the need to collect information from different digital or printed catalogues. In this situation, an information system with information of all components of an electrical system can help to designers and engineers in these tasks. This paper presents an information architecture to build this information system. Our proposed architecture will allow to acquire, analyze, and monitor data in the way that the user requires.

Keywords Power systems · Information systems · Data modeling · Electrical design

1 Introduction

Today the industrial companies face many challenges to diversify their products, increase the resource efficiency, and improve time to market [1]. The Fourth Industrial Revolution has defined a new level of organization and control in the life cycle of products [2, 3] which affects different areas in a company. The Industry 4.0 concept presents different challenges in the industry, and the companies should review their designing and drafting methods in all their disciplines to base the development of their products in a modern and interdisciplinary approach [4]. The growth of the digitization, IT penetration, and networking has increased

F. de Asis López-Fuentes (✉)
Department of Information Technology, Universidad Autónoma Metropolitana – Cuajimalpa, Cuajimalpa, Mexico City, Mexico
e-mail: flopez@cua.uam.mx

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_6

these challenges in the industry. Digitization of the designing and engineering processes for industrial plants has an important role in today's industry. Industrial or residential infrastructure projects generally involve a variety of facilities. Within these infrastructures, electrical installations have an important role. The design of an electrical system involves modeling and design calculations of lighting, power and cable routing systems, panel board design and balancing, lighting protection, cost estimation, installation detailing, and layout planning. During the design processes, engineers require technical information on the various components as well as their characteristics. During the design of an electrical distribution system for a building, industrial plant or residential area is required to observe norms and standards of the electrical industry such as NEC, ANSI, and NEMA [2, 4]. This activity is useful to define the concepts, components, electrical symbology, calculation recommendations, and electrical design. Currently, there are methodologies as BIM (Building Information Modeling) for the digitization of the design and engineering processes of the industrial plants [5]. However, government agencies and researchers need to do significant efforts for the standardization of BIM methods related to projects of electrical engineering. Also, object modeling in BIM must be compatible with manufacturer-based analysis and design studies [6]. In this paper, we present an information architecture for planning and design electrical industrial systems through the selection of the most important electrical components. Our proposed architecture tries to provide a solution to develop specialized CAD/CAE systems and more focused on small groups of design engineers in the planning of electrical installations.

The remainder of this paper has the following organization. Section 2 gives an overview about the different stages for the planning and design of an industrial electrical system. Conceptual architecture design with the interaction between the different programs and the database is presented in Sect. 3. Interaction between the different types of programs and the files is presented in Sect. 4. The data modeling and the relationships between entities based on the entity-relationship model are discussed in Sect. 5. Our article concludes in Sect. 6.

2 Planning and Design of an Industrial Systems

To define the information architecture and the data modeling of an industrial electrical system, we need to define the electrical design process for an industrial plant. The most common planning and development for an electrical project has the following stages [7, 8]:

1. Development of a preliminary project with the distribution of equipment, machinery, and the production areas in the industrial plant.
2. Definition of the electrical load zones within the industrial plant based on the distribution of the equipment to be connected.

3. Planning of the design of the electrical distribution system based on the following procedure: examination of the load, analysis of the demand in its different divisions (demand, peak load, demand factor, maximum demand, diversity factor, demand factor, load factor, coincident demand), investigation and selection of the appropriate distribution system for the plant, and its representation in a single-line diagram with the project loads.
4. Definition of the type of distribution system to be used (simple radial, expanded radial, primary system, secondary system, ring bus, etc.).
5. Definition of the following aspects:

   (a) Distance from load centers to the equipment
   (b) Characteristics of the power lines
   (c) Loads of the main feeders and branch circuits
   (d) Measurement and protection equipment according to the characteristics of the load
   (e) Capacities of the electrical equipment and elements involved in the system

6. Once the capacity of the electrical elements involved in the project has been defined, their physical dimensions are defined, and the physical location is carried out within the distribution plan of the industrial plant equipment.
7. Once the distribution of electrical equipment and feeders is completed, a catalog of specifications is made with the construction features for each element involved.

In the planning and design of an electrical system, it is necessary to define the elements that constitute it.

3 Conceptual Architecture Design

To develop our conceptual architecture, the following steps are been considered:

As the first step, we propose to create a database with relations (tables) based on entity-relationship diagrams. These tables are used by the queries required from the different application in the information system. The entity and relationship tables between entities contain the attributes of the electrical elements for an industrial electrical system and will allow us to make direct queries, such as the following:

- Characteristics of a conductor (e.g., weight, price and resistance)
- Characteristics of a transformer (e.g., manufacturer, capacity, and price)

In the database, we can make direct queries between entities, relating to the different entities, and these queries could be as follows:

- Select an electrical switch for a transformer based on its voltage.
- Consult different types of electrical switches for a transformer.
- Consult the different electrical protections for a conductor.

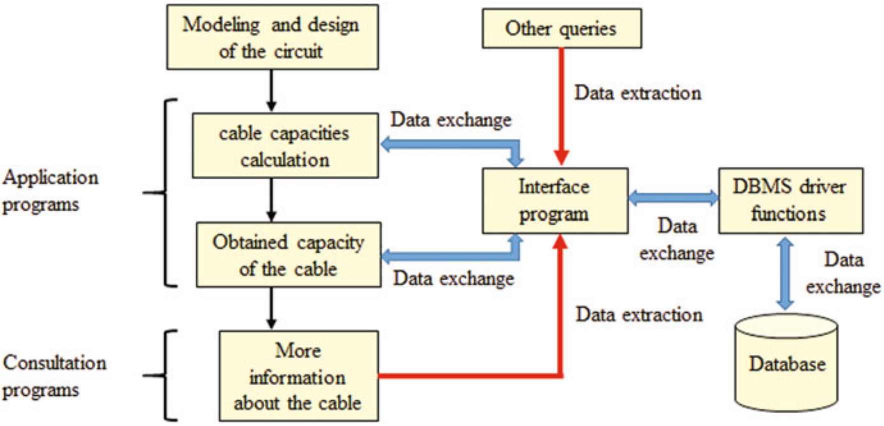


Fig. 1 Interaction between application programs and databases

In the second step, we can develop different applications to access the database [9]. These applications are programs which perform calculations to determine the capacities of the conductors for an industrial electrical system. The different data required by the application programs such as resistance, area, and capacity of the conductor are extracted from the database. Once the calculation process is finished in the application programs, we obtain optimal values with which we can independently access the database and select the appropriate conductor. This process is illustrated in Fig. 1.

For the handling and design of the graphic parts of an equipment or a component such as side views, front views, dimensions, and photographs, two options will be considered (see Fig. 2) [9]:

1. Use the own CAD, which allows to draw different schemes or equipment components. Subsequently, these drawings can be saved in a graphic format such as GIF (Graphics Interchange Format) or another.
2. Import external and independent files from an external CAD to the system. This will allow that drawings done by an external CAD system can be managed by the information systems (e.g., image file formats or graphic formats). Therefore, these drawings or images can be invoked by application and interface programs and can be displayed where they are required. This facilitates the construction of graphic parts from different CAD systems based on industry standards.

When a query is done to know the graphical characteristics of an equipment or device, the interface program invokes a function that allows visualizing the graphic characteristics of the equipment with data extracted from the database. In this way, we can obtain a technical sheet with the specifications of the device or equipment. Also, the system will be able to query the different units, without necessarily having to invoke their graphic features. By using the CAD system will be possible to build drawings for electrical systems using symbols from the CAD system itself and save

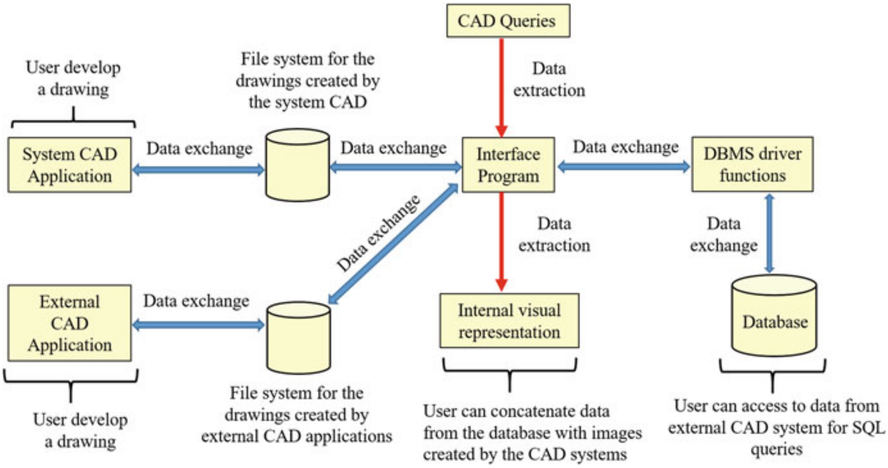


Fig. 2 Interaction between CAD applications, database, and the graphical interface

these drawings in some predefined format (e.g., GIF). This allows the manipulation of these drawings in other information systems that support this type of format. Database tables can be accessed from an external CAD system to perform SQL queries or applications with different drawing formats.

4 Interface with the Programs

For our information architecture, we consider three different types of programs, which have been defined based on their functionality:

- *Application programs:* These programs allow to obtain a specific result for a particular use. In this case, the application programs will be for the electrical calculation algorithms of conductors. The results obtained from these programs allow an adequate selection of the electrical conductors using three different methods.
- *Query programs:* These are programs to access the database. Unlike application programs, query programs will not require returning any data to be processed by any program. The requested data only is displayed on the screen for the user. These programs are also responsible for invoking the graphical characteristics of the entities (electrical equipment and components) from the database.
- *Interface programs:* These programs are used to open and access the database. In our case, these programs will be used, and each data is required, either by the application programs or by the query programs.

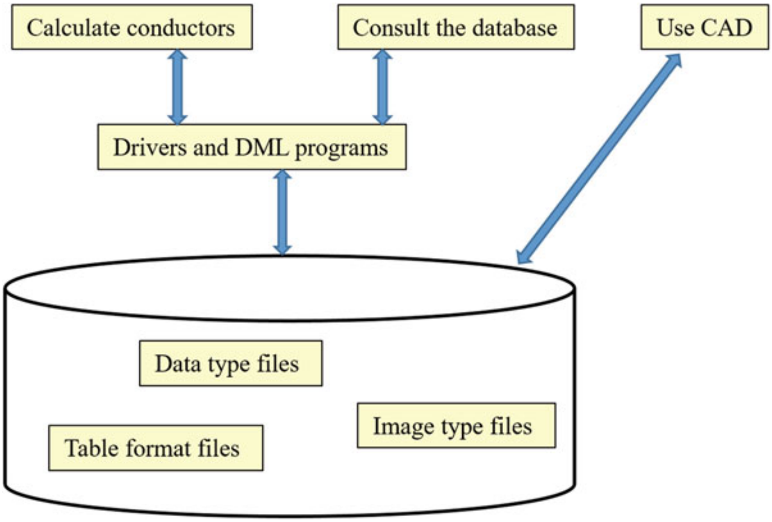


Fig. 3 Interaction between the different programs and the different types of files

The interaction of these programs with the database is divided into different steps which allow controlling access to the database and its tables through the DBMS (Database Management Systems). These steps are as follows:

1. *Connection to database.* This step allows to access the database and must be done before any queries can be processed.
2. *Opening the communication.* This action is performed by the DBMS driver. This step opens the communication and the table to be consulted in the database.
3. *Data extraction or reading.* It is the next step after the request is executed from an application program. Data is extracted or read from the corresponding tuples.
4. *Closing the communication.* This action closes the queried database table and the communication with the database.

Figure 3 shows the interaction between these different programs and the different types of files in our information architecture [9].

In this architecture, there are three application programs. The first performs the selection of conductors. The second program allows data to be extracted from table-type files and associated with the corresponding image files, while the third program performs the CAD operations. The selection of electrical conductors is made based on the procedures of current capacity, voltage drop, and short circuit and according to the NEC regulations. The calculation of conductors in an electrical system is important because its correct selection allows the adequate flow of electrical energy from a primary source to the final load. In this way, the conductors distribute energy to the different electrical loads in an industrial plant. If a user requires a conductor analysis, then the “Calculate conductors” option is chosen. However, we can see in Fig. 3 that the database is used by the “Calculate conductors” program, as well as the query programs (*Consult the database*). Depending on the needs of the user, the corresponding option will be chosen, and the files will be required in a different way. For example, the "Calculate conductors" option requires the use of DML (Data Manipulation Language) programs and drivers because this program will interact with the database tables, extracting data from the table format files in order to calculate different operations. If the user decides to make diagrams or drawings of the electrical system, then the option "Use CAD" must be selected. In this architecture, using the CAD option, the user can make diagrams and equipment drawings, as well as save the images made in a defined image format. The "Consult the database" program queries the different elements of an electrical system and allows knowing the technical data, images, diagrams, and dimensions of the corresponding equipment. For example, if a user searches for the motor element of a specific capacity, the program accesses the database, extracts the motor information, and associates it with the corresponding image.

5 Data Modeling

For the data modeling, the following technical points have been considered [9]:

- Design the conceptual part of the database, such as the following: define the entities, the attributes of each entity, the relationships, the ER diagram (entity-relationship) [10, 11], the different affinities based on the relational model, the normalization required for these affinities, and the data dictionary.
- Also for the design criteria of the database, it is necessary to consider the relationships that the devices and components of an industrial electrical system have among themselves, the recommendations of manuals in the electrical industry, the recommendations of the equipment manufacturers, and the diagrams. In this data, modeling was considered the following standards: NEC, NEMA, IEEE, and ANSI [12].

In the data modeling, the relationships are the associations that can be established between different entities [10, 13]. In this work, the main components of an industrial electrical system are represented as entities. An E-R diagram used to build the information architecture is shown in Fig. 4. Here, all entities that are involved in the design of the electrical system are considered. In this diagram, the attributes of the entities and the relationships are not indicated.

6 Conclusions

An information architecture plays an important role for the digitization of the design and engineering processes of industrial electrical systems. In this paper, we present an information architecture for building an information system which

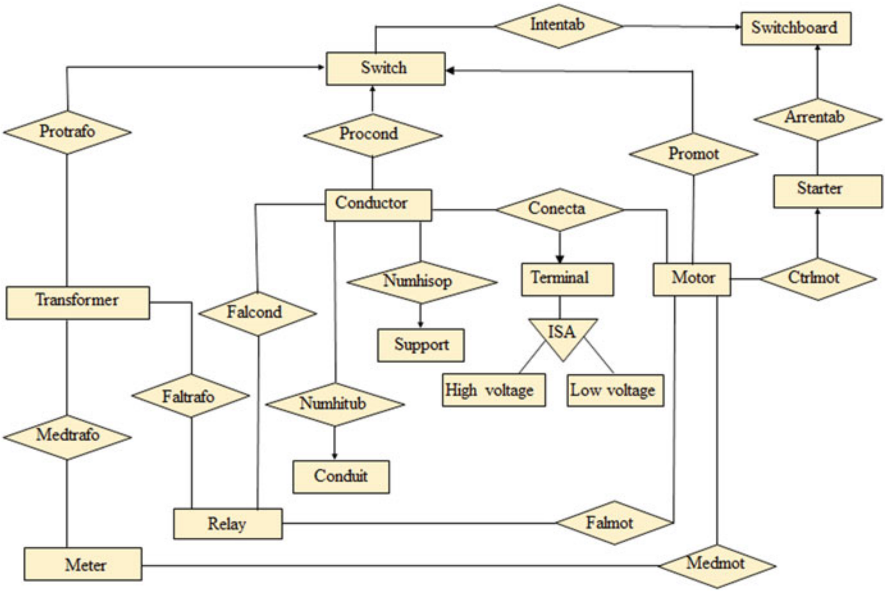


Fig. 4 Diagram E-R used as reference to build the information architecture [9]

helps the engineers select the equipment during the electrical design processes of an industrial plant. Our proposed architecture integrates information generated from three different applications program with a database. These application programs are used for conductor calculation, computer-aided design, and database consultation. Our conceptual design shows the interaction between the application programs and databases. Also, there are interaction between different programs and different types of files. In our architecture, the data are modeled using the E-R model, where each principal electrical component of an industrial electrical system is modeled as an entity. We believe that our current work can be extended to other fields related to Industry 4.0. We plan to do this activity as a future work.

References

1. Rennung, R., Luminosu, C.T., Draghici, A.: Service provision in the framework of industry 4.0. In: SIM 2015/13th International Symposium in Management, Procedia – Social and Behavioral Sciences, vol. 221, pp. 372–377. Elsevier, Timisoara (2016)
2. Vaidya, S., Ambad, P., Bhosle, S.: Industry 4.0 – glimpse. In: Procedia Manufacturing, vol. 20, pp. 233–238. Elsevier, Aurangabad, Maharashtra (2018)
3. Oztemel, E., Gursev, S.: Literature review of industry 4.0 and related technologies. J. Intell. Manuf. **31**(1), 127–182 (2020)
4. Eigner, M.: Modellbasierte Virtuelle Produktentwicklung auf einer Plattform für System Lifecycle Management, in Industrie 4.0. In: Beherrschung der industriellen Komplexität mit SysLM. Springer, Heidelberg (2013)

5. Eastman, C., Tiecholz, P., Sacks, R., Liston, K.: BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors, 2nd edn. John Wiley, Hoboken (2011)
6. Farooq, J., Sharma, P., Kumar, S.: Applications of building information modeling in electrical systems design. J. Eng. Sci. Technol. Rev. **10**(6), 119–128 (2017)
7. IEEE: IEEE Recommended Practice for Electric Power Distribution for Industrial Plants, ANSI/IEEE Std 141-1993. IEEE Inc., New York (1993)
8. Short, T.A.: Electric Power Distribution Handbook. CRC Press, Boca Raton (2004)
9. Lopez-Fuentes, F.A.: Implementación de una base de datos gráfica para el desarrollo de proyectos eléctricos, Master thesis, ITESM (1998)
10. Silberschatz, A., Korth, H., Sudarshan, S.: Database Systems Concepts, 7th edn. McGraw-Hill, New York (2019)
11. Codd, E.F.: The Relational Model for Database Management: Version 2, 1st edn. Addison-Wesley Pub. Co, Reading (1990)
12. Jay, F.: IEEE Standard Dictionary of Electrical and Electronics Term, 3rd edn. IEEE Press, New York (1984)
13. Kroenke, D.: Database Processing, Fundamentals, Design and Implementation, 5th edn. Prentice-Hall Inc., New York (1995)

Optimized Packing Soft Convex Polygons

Igor Litvinchev, Luis Infante, Tatiana Romanova, Alberto Martinez-Noa, and Luis Gutierrez

Abstract Packing soft convex polygons in an optimized convex container is considered. It is assumed that the soft polygonal object can change its shape in certain limits, while its area remains constant. Non-overlapping, containment, and area conservation constraints are formulated for soft polygonal objects, and a corresponding nonlinear programming model is presented. Numerical experiments for packing soft triangles and pentagons in optimized circular and quadratic containers are presented to demonstrate efficiency of the proposed approach.

Keywords Soft polygons · Optimized packing · Layout conditions

1 Introduction

In optimized packing problems, several objects must be allocated completely inside a container, without mutual overlapping, and optimizing a certain objective. The typical objective is minimizing the container's dimensions to allocate a given number of objects. Although the shape of the container can vary and is defined by optimization, the objects are typically fixed (rigid objects). Packing soft objects changing their shapes and dimensions to optimize a certain objective is much less investigated. To the best of our knowledge, so far only rectangular and ellipsoidal soft objects were considered in optimized packing (see [1–6] and the references therein). The dimensions of the rectangles or ellipses (circles) were allowed to change the subject to their area or perimeter conservation.

I. Litvinchev (✉) · L. Infante · A. Martinez-Noa · L. Gutierrez
Graduate Program in Systems Engineering, Nuevo Leon State University (UANL), Monterrey, San Nicolas de los Garza, Nuevo Leon, Mexico

T. Romanova
Department of Mathematical Modeling and Optimal Design, Institute for Mechanical Engineering Problems of the National Academy of Sciences of Ukraine, Kharkiv, Ukraine

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_7

Packing soft object has various applications in biology [7], mechanics [8], and material science [9], to mention a few. Packing soft triangles in 2D containers is closely related to mesh generation in finite element analysis and, more generally, with the ability to handle complex shapes easily [10].

In this paper, we focus on packing soft convex 2D (two-dimensional) polygonal objects in a convex optimized container. It is assumed that the soft object can change its shape in certain limits, but its area remains constant. Using coordinates of the vertices of the objects as decision variables [11], layout and area conservation constraints are formulated. The resulting nonlinear optimization problem is solved by the global solver BARON (Branch and Reduce Optimization Navigator) [12, 13]. Numerical examples for packing soft triangles and pentagons in an optimized circle or square are presented.

2 Problem Formulation

Let $\Omega(R)$ be a convex bounded 2D domain (container) defined by an m-vector of its metric parameters R. For example, for a circular domain, R stands for radius, while for a rectangular domain, R is defined as the dimensions of the rectangle.

Let $T_k(X_k),\ k = 1, 2 \ldots K$ be bounded 2D convex polygons with their vertices $X_k = \left\{X_k^j, j \in J_k\right\}$. Denote by $S_k(X_k)$ the area of the corresponding polygon.

The problem of *Optimized Packing Soft Polygons* (OPSP) is stated as follows. Find $X = \{X_k,\ \ k = 1, 2 \ldots K\}$ and R minimizing the objective $F(X, R)$ subject to:

(a) $\text{int}T_k(X_k) \cap \text{ int } T_p(X_p) = \varnothing,\ k, p = 1, 2 \ldots K,\ p > k$ (non-overlapping)
(b) $T_k(X_k) \subseteq \Omega(R),\ k = 1, 2 \ldots K$ (containment)
(c) $S_k(X_k) = S_{k0},\ k = 1, 2 \ldots K$ (area conservation)
(d) $G(X, R) \leq 0$ (metric constraints)

Here the objective $F(X, R)$ may represent the area/perimeter/size of the container, constraints (a) and (b) state the standard packing layout conditions, and constraint (c) assures that the area of the polygon remains constant all feasible vertices positions. Metric constraints (d) characterize "elasticity" of the soft polygons. For example, metric constraints can be used to limit the variance of the distances between adjacent vertices. Similarly, constraint (c) can be relaxed to permit a limited variance of the area. Below constraints (a)–(d) are stated explicitly.

2.1 *Non-overlapping and Containment*

Convex polygons T_k, T_p are non-overlapping (though they can be tangent) if there exists a nonzero 2D vector v_{kp} and a scalar b_{kp}, such that a hyperplane $v_{kp}^t x = b_{kp}$ separates T_k and T_p, i.e.,

$$v_{kp}^t x \leq b_{kp} \ \text{ for any } \ x \in T_k,$$

$$v_{kp}^t x \geq b_{kp} \ \text{ for any } \ x \in T_p.$$

This can be re-written in the form

$$\max\left\{v_{kp}^t x : x \in T_k\right\} \leq b_{kp},$$
$$\min\left\{v_{kp}^t x : x \in T_p\right\} \geq b_{kp}$$

or, equivalently,

$$\max\left\{v_{kp}^t x : x \in T_k\right\} \leq \min\left\{v_{kp}^t x : x \in T_p\right\}.$$

Since an extremum of a linear function is attained at a vertex of a bounded convex polygon, then we have

$$\max\left\{v_{kp}^t X_k^j : j \in J_k\right\} \leq \min\left\{v_{kp}^t X_p^j : j \in J_p\right\}. \tag{1}$$

Denote

$$\alpha_{kp} = \max\left\{v_{kp}^t X_k^j : j \in J_k\right\}, \ \ -\beta_{kp} = \min\left\{v_{kp}^t X_p^j : j \in J_p\right\}.$$

Then by the definition of max and min, the inequality (1) is equivalent to the following system of inequalities

$$\begin{aligned} &\alpha_{kp} + \beta_{kp} \leq 0, \\ &\alpha_{kp} \geq v_{kp}^t X_k^j, \ j \in J_k, \\ &-\beta_{kp} \leq v_{kp}^t X_p^j, \ j \in J_p, \\ &v_{kp} \neq 0. \end{aligned} \tag{2}$$

If v_{kp} is feasible to (1) or (2), then ρv_{kp} is also feasible for any scalar $\rho > 0$. Thus, to assure $v_{kp} \neq 0$, we may normalize v_{kp} by setting, e.g., $\|v_{kp}\| \geq 1$, where $\|\cdot\|$ is a certain norm, not necessary Euclidean. We can also use $e^t v_{kp} \geq 1$, where $e \in \mathbb{R}^2$ is a unit vector. This inequality eliminates 0 and vectors with all negative components. However, if v_{kp} corresponds to a separating plane, then $-v_{kp}$ also separates T_k and T_p.

By convexity of $\Omega(R)$ containment conditions $T_k(X_k) \subseteq \Omega(R), \ k = 1, 2 \ldots K$ are equivalent to

$$X_k^j \in \Omega(R), \ j \in J_k, \ \ k = 1, 2 \ldots K \tag{3}$$

Note that non-overlapping and containment conditions can be formulated in the form (2) and (3) in n-dimensional space for $n \geq 2$.

2.2 Area Conservation and Metric Constraints

The area of the polygon $X_k = \left\{X_k^j,\, j \in J_k\right\}$ having counterclockwise oriented two-dimensional vertices $X_k^j = \left(x_k^j, y_k^j\right)$, $j = 1, 2 \ldots J_k$ can be calculated using the shoelace formula, also known as Gauss's area formula or the surveyor's formula (see, e.g., [14, 15] and the references therein). Correspondingly, area conservation constraints have the form

$$S_{k0} = 0.5\left(x_k^1 y_k^2 - x_k^2 y_k^1 + x_k^2 y_k^3 - x_k^3 y_k^2 + \cdots + x_k^{J_k} y_k^1 - x_k^1 y_k^{J_k}\right), \quad k = 1, 2 \ldots K. \tag{4}$$

In this paper metric constraint (d) in OPSP is used in the form

$$(1 - \rho_k)\, d_k \leq \left\| X_k^j - X_k^{j+1} \right\| \leq (1 + \rho_k)\, d_k,\ j = 1, 2 \ldots J_k,\quad k = 1, 2 \ldots K. \tag{5}$$

and to simplify the notation, it is assumed that $X_k^{J_k+1} = X_k^1$. Here the scalar parameter $0 \leq \rho_k \leq 1$ represents the level of "elasticity" of the polygon. This parameter is used to establish bounds for the allowed deviation of the polygon's side-length from the side-length d_k. For simplicity it is assumed that for $\rho_k = 0$ all polygons are regular and S_{k0} is the area of the corresponding regular polygon with the side-length d_k.

Two containers are considered, a circle $\Omega_c(R) = \{(x, y) : x^2 + y^2 \leq R^2\}$ and a square $\Omega_s(R) = \{(x, y) : 0 \leq x \leq R,\ 0 \leq y \leq R\}$. In both cases, the objective function to minimize is $F(X, R) \equiv R$.

3 Computational Results

In this section, numerical examples of packing soft objects are presented. The objective was to see how the packing density, the size of the container, and the shapes of the objects are changed for different values of the elasticity parameter. The same value of the elasticity parameter was used for all objects. All experiments were running on NEOS server using AMPL modeling language. Nonlinear programming problems were solved by the global solver BARON (Branch and Reduce Optimization Navigator) for AMPL version 19.12.7 (see, e.g., [12, 13]).

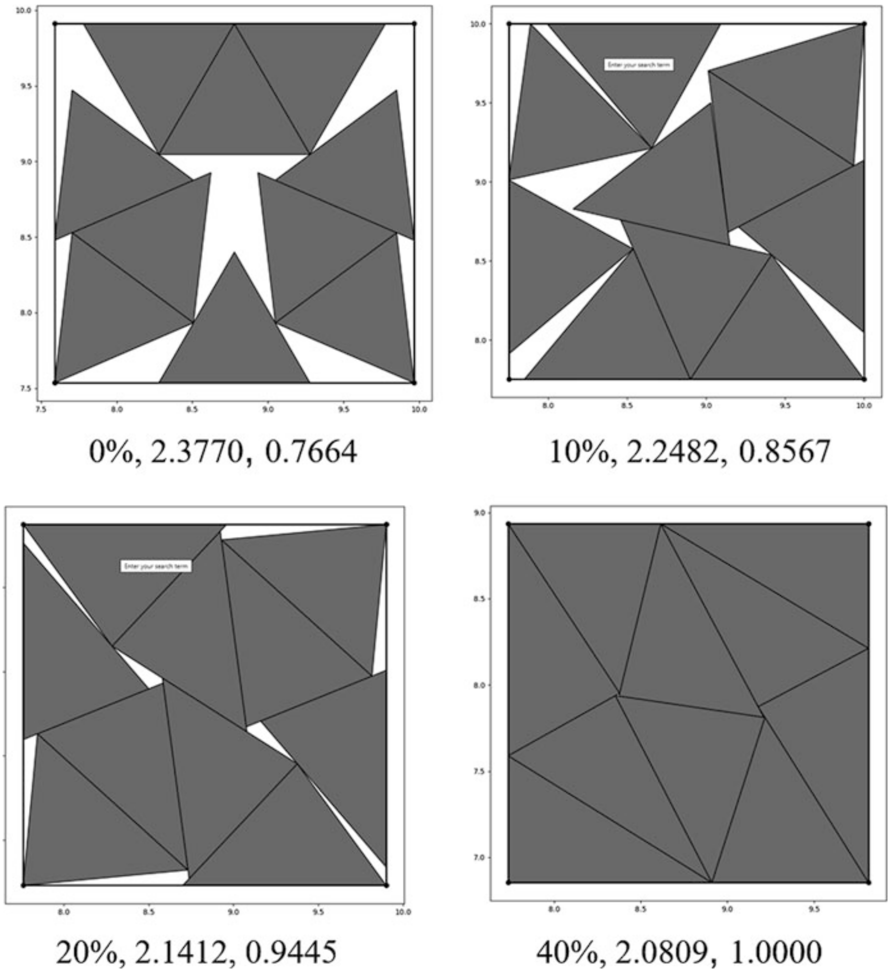


Fig. 1 Packing ten soft triangles in a minimal square

Packing ten soft triangles in a minimal square is presented in Fig. 1. Figure captions provide elasticity parameter $\rho \times 100\%$, corresponding optimal size of the square and packing density (total area of the objects over the area of the container). For $\rho = 0$ all triangles are equilateral with side-lengths $d_k = 1$ and $S_{k0} = 0.25\sqrt{3} \approx 0.43301$.

Figure 2 shows packing 45 soft triangles in a minimal circle. For $\rho = 0$ all triangles are regular with side-lengths $d_k = 1$. Figure captions provide elasticity parameter in %, corresponding optimal radio of the container, and packing density.

Packing ten soft pentagons is presented in Fig. 3. Figure captions provide elasticity parameter in %, corresponding optimal radio of the circle, and packing density. For $\rho = 0$ all pentagons are regular with the side-lengths $d_k = 1$.

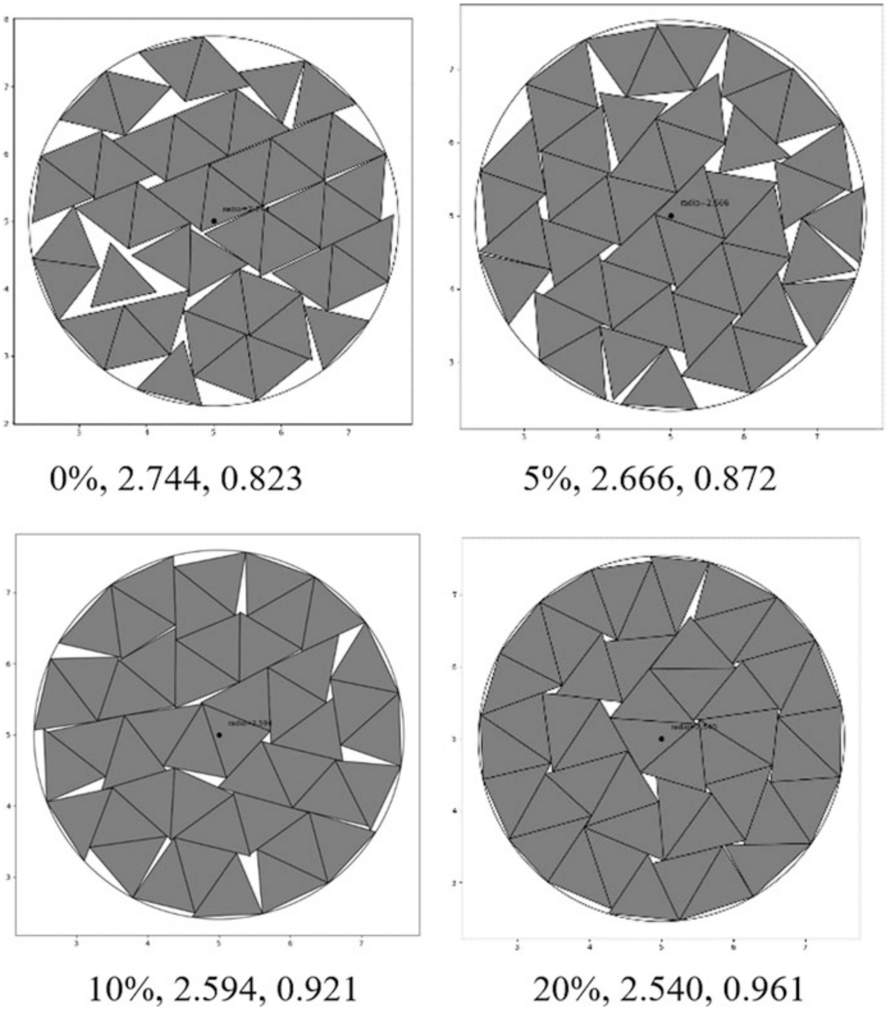


Fig. 2 Packing 45 soft triangles in a minimal circle

Figure 4 demonstrates packing ten soft pentagons for unrestricted side-length (constraint (5) is omitted). In the first case, all pentagons can freely change their dimensions. In the second case, four pentagons are rigid and stay regular with their side-lengths 1, while six other pentagons can freely change their dimensions. Corresponding figure captures indicate the radio of the circular container and density.

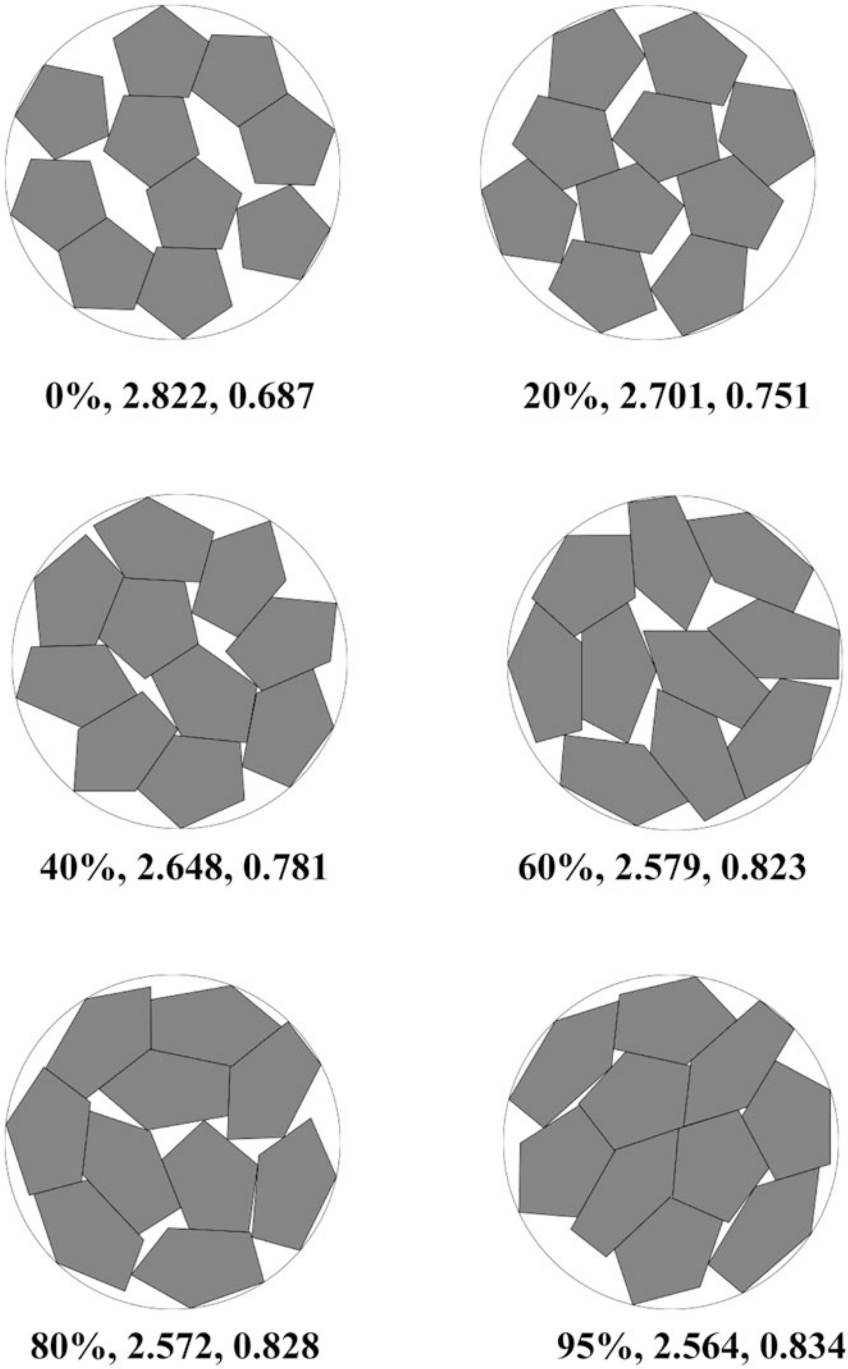


Fig. 3 Packing ten soft pentagons in a minimal circle

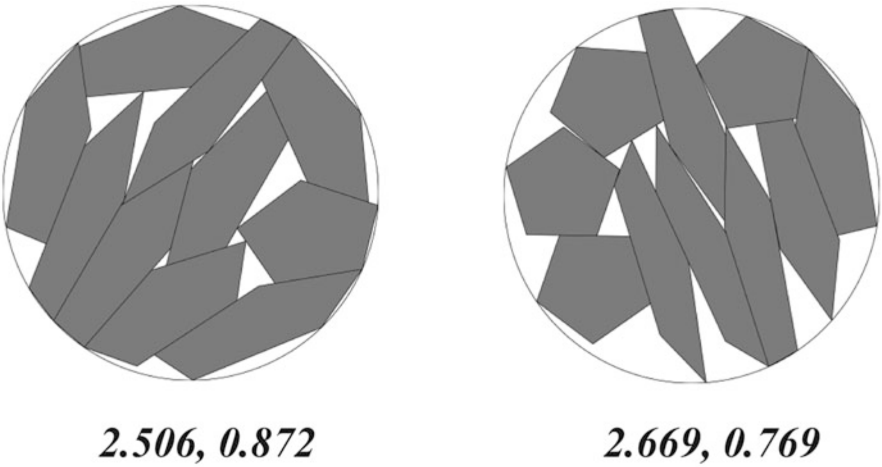


Fig. 4 Packing ten soft pentagons with unrestricted side-length

4 Conclusions

Optimized packing for soft 2D polygons was considered subject to their area conservation. Using coordinates of the vertices as decision variables, non-overlapping, containment, and area conservation constraints were formulated. Numerical results were presented to demonstrate the correspondence between elasticity and packing density of polygons for optimized circular and square containers. To linearize approximately arising nonlinear models, grid approximations of the container can be used [16].

An interesting direction for the future research is considering 3D soft objects. In this case volume and/or surface area conservation (or quasi-conservation) conditions may arise. In natural sciences applications, the volume of the object corresponds to its weight, while the surface area is related to the surface energy. Additional balancing or sparsity conditions may be imposed [17, 18]. Packing soft objects in a soft container is also an interesting area for the future research. In this case, the shape of the container is not fixed and must be defined by optimization (see [19] for the 2D case). Some results in these directions are on the way.

References

1. Fügenschuh, A., Junosza-Szaniawski, K., Lonc, Z.: Exact and approximation algorithms for a soft rectangle packing problem. Optimization. **63**(11), 1637–1663 (2014)
2. Ji, P., He, K., Jin, Y., Lan, H., Li, C.: An iterative merging algorithm for soft rectangle packing and its extension for application of fixed-outline floorplanning of soft modules. Comput. Oper. Res. **86**, 110–123 (2017)
3. Brenner, U.: γ -Soft packings of rectangles. Comput. Geom. **70**, 49–64 (2018)

4. Bui, Q.T., Vida, T., Hà, M.H.: On three soft rectangle packing problems with guillotine constraints. J. Glob. Optim. **74**, 45–62 (2019)
5. Winkelmann, J.: Structures of Columnar Packings with Soft and Hard Spheres, PhD Thesis, School of Physics, Trinity College, Dublin (2020)
6. Romanova, T., Stoyan, Y., Pankratov, A., Litvinchev, I., Kravchenko, O., Duryagina, Z., Chugai, A.: Optimized packing soft ellipses. In: Manshahia, M.S., Litvinchev, I., Thomas, J.J., Vasant, P., Weber, W.G. (eds.) Human Assisted Intelligent Computing: Modelling, Simulations and Applications. IOP, United Kingdom (2023)
7. Jiang, J., Garikipati, K., Rudraraju, S.: A diffuse interface framework for modeling the evolution of multi-cell aggregates as a soft packing problem driven by the growth and division of cells. Bull. Math. Biol. **81**, 3282–3300 (2019)
8. Yuan, Q., Li, Z., Gao, Y., Wang, Y.H., Li, X.: Local responses in 2D assemblies of elliptical rods when subjected to biaxial shearing. Acta Geotech. **14**, 1685–1697 (2019)
9. Chen, Y., Yuan, M., Wang, Z., Zhao, Y., Li, J., Hu, B., Xia, C.: Structural characterization and statistical properties of jammed soft ellipsoid packing. Soft Matter. **17**, 2963 (2021). https://doi.org/10.1039/d0sm01699c
10. Lewis, B.A., Robinson, J.S.: Triangulation of planar regions with applications. Comput. J. **21**(4), 324–332 (1978). https://doi.org/10.1093/comjnl/21.4.324
11. Litvinchev, I., Romanova, T., Corrales-Diaz, R., Esquerra-Arguelles, A., Martinez-Noa, A.: Lagrangian approach to modeling placement conditions in optimized packing problems. Mob. Netw. Appl. **25**, 2126–2133 (2020)
12. Sahinidis, N.V.: BARON 19.12.7: Global Optimization of Mixed-Integer Nonlinear Programs, User's Manual (2019)
13. Tawarmalani, M., Sahinidis, N.V.: A polyhedral branch-and-cut approach to global optimization. Math. Program. **103**(2), 225–249 (2005)
14. Allgower, E.L., Schmidt, P.H.: Computing volumes of polyhedra. Math. Comput. **46**(173), 171–174 (1986)
15. Braden, B.: The surveyor's area formula. Coll. Math. J. **17**(4), 326–337 (1986)
16. Litvinchev, I., Infante, L., Ozuna Espinosa, E.L.: Approximate circle packing in a rectangular container: integer programming formulations and valid inequalities. In: González-Ramírez, R.G., Schulte, F., Voß, S., Ceroni Díaz, J.A. (eds.) Computational Logistics. ICCL 2014. Lecture Notes in Computer Science, vol. 8760. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11421-7_4
17. Romanova, T., Litvinchev, I., Pankratov, A.: Packing ellipsoids in an optimized cylinder. Eur. J. Oper. Res. **285**(2), 429–443 (2020)
18. Romanova, T., Stoyan, Y., Pankratov, A., Litvinchev, I., Plankovskyy, S., Tsegelnyk, Y., Shypul, O.: Sparsest balanced packing of irregular 3D objects in a cylindrical container. Eur. J. Oper. Res. **291**(1), 84–100 (2021)
19. Kallrath, J., Romanova, T., Pankratov, A., Litvinchev, I., Infante, L.: Packing convex polygons in minimum-perimeter convex hulls. J. Glob. Optim. **85**(1), 39–59 (2023). https://doi.org/10.1007/s10898-022-01194-4

Part III
Industry 4.0 Applications

E-Commerce on Startup: A Systematic Literature Review

Agung Purnomo, Mega Firdaus, Yogi Tri Prasetyo, Elsa Rosyidah, Satria Fadil Persada, Thalea Christy Nathaniela, and Fairuz Iqbal Maulana

Abstract This study's objectives include reviewing the existing literature and studying e-commerce on startups. The PRISMA protocols for performing and reporting systematic reviews were implemented when conducting a systematic literature review. The Scopus database was thoroughly searched, and 22 peer-reviewed papers were found. The research program was provided, such as publication sources, disciplines, and geographic location. E-commerce on startups was grouped to highlight the focus on particular elements and the direction of future research in this field.

Keywords E-commerce · Entrepreneurship · PRISMA protocols · Startup · Systematic review

1 Introduction

Today's business and entrepreneurial landscape is becoming more diversified and competitive, and its rate of development is very unpredictable. As a result, having the appropriate strategy will be essential to a company's ability to compete and survive. At this time, network-based or online transactions for buying and selling have rapidly increased. Electronic commerce, or e-commerce for short, is a relatively new phrase that can either exchange goods, services, or information via an information network like the Internet or the process of purchasing and selling

A. Purnomo · S. F. Persada · T. C. Nathaniela · F. I. Maulana
Bina Nusantara University, Jakarta, Indonesia

M. Firdaus · E. Rosyidah
Universitas Nahdlatul Ulama Sidoarjo, Sidoarjo, Indonesia

Y. T. Prasetyo (✉)
Mapua University, Manila, Philippines
e-mail: ytprasetyo@mapua.edu.ph

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_8

things or services online. Implementing e-commerce is one of the essentials for a company's success [1–3].

Startup entrepreneurship is the underpinning of economic development in every nation and a vital part of discovery, industry innovation, and entrepreneurialism, particularly in contexts with global markets [4]. A startup is a learning company that continuously works to improve its operations and outcomes, much like entrepreneurship itself is a learning process [5, 6].

An accurate, scholarly assessment of the corpus of work on a particular subject or discipline is a thorough systematic literature review (SLR) [7]. A systematic, repeatable process must be used to find, assess, and synthesize all relevant studies [8]. Specific protocols provide a written record of the procedures followed for document search, document exclusion and inclusion, and document analysis [9]. An SLR attempts to gather as much current, evidence-based research that is relevant to the topic, depending on where it was performed [10]. Although they promote the implementation of specific tactics that can reduce accuracy and bias, SLRs are recognized for producing rigorous evidence reviews [8]. What is the status of the e-commerce literature and research related to startups? This is one of the research questions this review suggests. This study's objectives include reviewing the existing literature and studying e-commerce on startups from an SLR perspective.

2 Methods

To conduct a systematic literature review, this work adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) criteria [9]. This study used a comprehensive literature database to conduct a systematic review of the literature.

To identify and link relevant publications in the global Scopus database, this study has paired the right keywords related to e-commerce on startup research. Although academics consider the Scopus database a reliable source of scholarly papers, it was used as the primary information source. As seen in Fig. 1, this study used the terms "e-commerce" and "startup" from the title, abstract, and keywords to find pertinent information in the Scopus database. To collect all published data for a year, the data mining was limited to yearly data. The following was the search option utilized in data mining (TITLE-ABS-KEY (e-commerce) AND TITLE-ABS-KEY (startup)) AND PUBYEAR >1998 AND PUBYEAR <2022 AND (LIMIT-TO (SUBJAREA, "BUSI")) AND (LIMIT-TO (SRCTYPE, "j")) as of July 2022. In this stage, 22 items were found. The quantitative analysis used in this SLR includes examining annual publications, publishing sources, regional contexts, and discipline-specific studies.

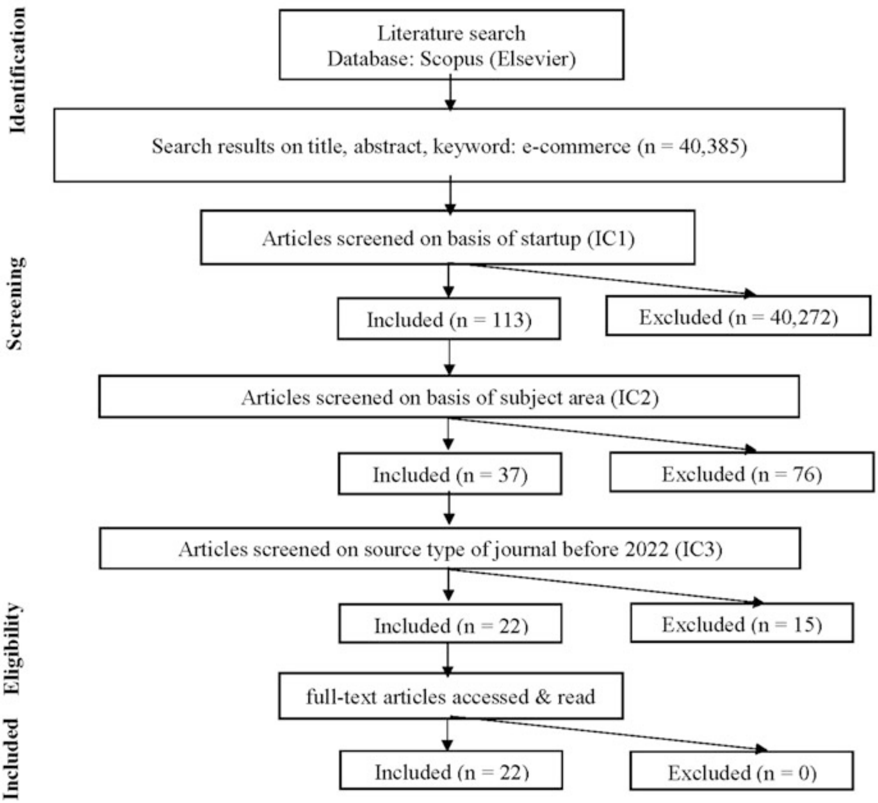


Fig. 1 PRISMA protocols

3 Results and Discussion

In the field of e-commerce on startup, this section discusses the status of existing literature and research based on quantitative analysis and disciplines analysis.

3.1 *Annual Publications*

Figure 2 displays 22 documents that were released each year. According to these statistics, it can be seen that the number of publications about e-commerce startups has fluctuated, starting to stabilize in 2020 with a significant growth rate. This publication growth rate was caused by the COVID-19 pandemic in 2020, which increased e-commerce [11].

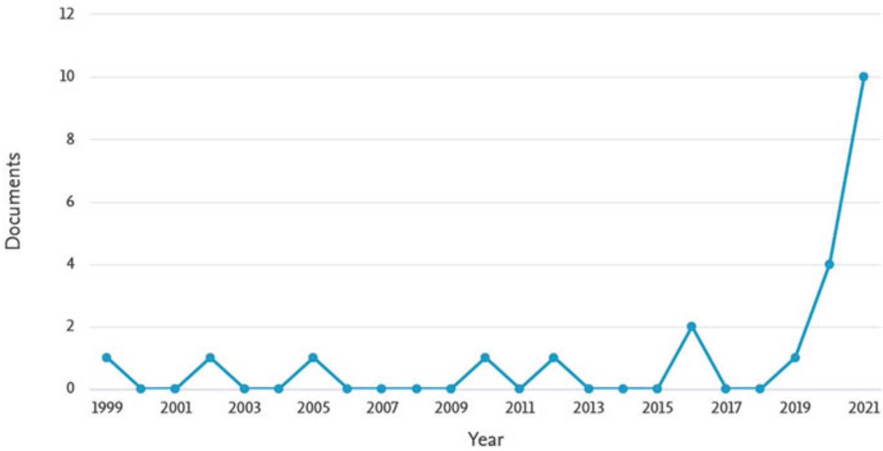


Fig. 2 The e-commerce in the startup sector's annual publications

3.2 Publications Sources

Table 1 lists the 17 publication sources where 22 documents with categories of diverse issues have been published. *Emerald Emerging Markets Case Studies* was a journal that regularly publishes articles with the theme of e-commerce on startups (published by Emerald Insight), *Journal of Electronic Commerce in Organizations* (posted by IGI Global), and *Asian Journal of Management Cases* (published by Sage Publications). All of these publications were listed on Scopus and had good scientific solidity.

3.3 Geographical Contexts

Asian and European nations have conducted most of the e-commerce startup research (seen in Table 2). We examined the location of the research data (if the researcher identifies the research site) and the author to discover which countries contributed the most to this study (if the research uses observation, case studies, or literature review). Otherwise, whether the piece publishes secondary data or employs empirical analysis has been taken into account based on the author's location.

We discovered that there were 12 distinct countries represented in the publication out of a total of 22 articles. There were 12 publications from the research done in Asia, 7 of which were from India, 3 from Indonesia, and 1 from the Philippines and the United Arab Emirates. Europe was the second continent to contribute to this subject; the United Kingdom, Croatia, and Germany published the most. The variety

Table 1 Papers by publication sources

Publication sources	Publication number
Emerald Emerging Markets Case Studies	3
Journal of Electronic Commerce in Organizations	3
Asian Journal of Management Cases	2
Data Base for Advances in Information Systems	1
European Business Review	1
Fashion and Textiles	1
Industry and Innovation	1
International Journal of Applied Business and Economic Research	1
International Journal of Physical Distribution and Logistics Management	1
International Journal of Scientific and Technology Research	1
International Journal of Supply Chain Management	1
International Journal on Food System Dynamics	1
Journal of Business Venturing Insights	1
Journal of Cases on Information Technology	1
Journal of International Business Education	1
Revista De Gestao	1
Transnational Marketing Journal	1

Table 2 Geographical contexts of e-commerce on startups studies

Geographical	Publication number	Percentage (%)
Asia	**12**	**50%**
India	7	29%
Indonesia	3	13%
Philippines	1	4%
United Arab Emirates	1	4%
Europe	**5**	**25%**
United Kingdom	4	17%
Croatia	1	4%
Germany	1	4%
American continent	**4**	**17%**
Brazil	2	9%
Canada	1	4%
United States	1	4%
Australian continent	**1**	**4%**
Australia	1	4%
Not specified	**1**	**4%**
Undefined	1	4%
Total	**22**	**100%**

of places shows the expanding significance of e-commerce in startup research as a topic and its ongoing appeal to professional academics around the globe.

Table 3 The most frequently researched disciplines

Disciplines	Paper	Articles
Marketing	8	[4, 5, 12–17]
Entrepreneurship	5	[11, 18–21]
Strategic	4	[22–25]
Operation and supply chain	2	[26, 27]
Finance	1	[28]
Organization and human resource	1	[29]
Management information system	1	[30]

3.4 Disciplines

E-commerce represents a variety of management disciplines in startups (can be seen in Table 3). There were 22 studies specifically about e-commerce on startups that validated their findings across a range of scientific disciplines. The area of marketing that has received the most research on startups is e-commerce ($n = 8$). Entrepreneurship was mentioned in as many as five documents after that. This study also utilized other academic fields, including strategic ($n = 4$), operation and supply chain ($n = 2$), finance ($n = 1$), organization and human resource ($n = 1$), and management information system ($n = 1$). The 22 articles that were chosen did not just concentrate on one or two areas. But also in other sectors, marketing was the most considerable discipline. So it can be concluded that marketing in this era was essential in e-commerce in the startup world.

4 Conclusions

E-commerce for startups has a vital role in encouraging economic independence and making it easier for businesses to expand their reach without limits. Through offering numerous quantitative analyses of e-commerce in startup literature, including publication sources, an annual publication, and geographic location, this study looks into the distribution of e-commerce research on startups. The findings of the discipline's study demonstrate that several scientific fields and diverse industries have investigated e-commerce on startups. The greatest of this field's advancements have come from Asia. The yearly publication reveals that the field's research has stabilized since 2020, and *Emerald Emerging Markets Case Studies* has emerged as the research source with the most significant number of articles. It is believed that this evaluation will open the door for a new investigation into fields with a shortage of data and skilled analysis.

Acknowledgment This research was funded by Mapúa University Directed Research for Innovation and Value Enhancement (DRIVE).

References

1. Sofiyah, F.R., Yulinda, Muda, I.: The influence of e-commerce toward internal factors of startup business success (case in Indonesia). Int. J. Sci. Technol. Res. **8**(12) (2019)
2. Maulana, F.I., Zamahsari, G.K., Purnomo, A.: Web design for distance learning Indonesian language BIPA. In: 2020 International Conference on Information Management and Technology (ICIMTech), pp. 988–991 (2020). https://doi.org/10.1109/ICIMTech50083.2020.9211175
3. Purnomo, A., Sari, A.K., Mufidah, E., Asitah, N., Aziz, A.: Digital business: a scientific publication positioning using scientometric analysis. In: 2020 International Conference on Information Management and Technology (ICIMTech), pp. 588–593 (2020). https://doi.org/10.1109/ICIMTech50083.2020.9211174
4. Bodolica, V., Kasih, B.: Reimagining the post-pandemic industry of arts in the Arab world: fast-tracked digital uplift or back to business as usual? Emerald Emerg. Mark. Case Stud. **11**(3), 1–25 (2021). https://doi.org/10.1108/EEMCS-06-2020-0181
5. Sood, S.: The death of social media in start-up companies and the rise of s-commerce. J. Electron. Commer. Organ. **10**(2), 1–15 (2012). https://doi.org/10.4018/jeco.2012040101
6. Purnomo, A., Susanti, T., Rosyidah, E., Firdausi, N., Idhom, M.: Digital economy research: thirty-five years insights of retrospective review. Proc. Comput. Sci. **197**, 68–75 (2022). https://doi.org/10.1016/j.procs.2021.12.119
7. Petticrew, M., Roberts, H.: Systematic reviews in the social sciences: a practical guide. Blackwell Publishing Ltd., Malden (2006)
8. Tranfield, D., Denyer, D., Smart, P.: Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manag. **14**(3), 207–222 (2003). https://doi.org/10.1111/1467-8551.00375
9. PRISMA: PRISMA transparent reporting of systematic review and meta-analysis. Prisma. [Online]. Available: https://prisma-statement.org/ (2022)
10. Jones, M.V., Coviello, N., Tang, Y.K.: International entrepreneurship research (1989–2009): a domain ontology and thematic analysis. J. Bus. Ventur. **26**(6), 632–659 (2011). https://doi.org/10.1016/j.jbusvent.2011.04.001
11. Frare, A.B., Beuren, I.M.: Job autonomy, unscripted agility and ambidextrous innovation: analysis of Brazilian startups in times of the COVID-19 pandemic. Rev. Gestão. **28**(3), 263–278 (2021). https://doi.org/10.1108/REGE-01-2021-0005
12. Kim, N.L., Shin, D.C., Kim, G.: Determinants of consumer attitudes and re-purchase intentions toward direct-to-consumer (DTC) brands. Fash. Text. **8**(1), 8 (2021). https://doi.org/10.1186/s40691-020-00224-7
13. Garg, N., Gera, S., Punia, B.K.: Gratitude and turnover intention in e-commerce startups: investigating the mediating role of organizational citizenship behaviour. J. Electron. Commer. Organ. **19**(2) (2021)
14. Taneja, B.M., Puri, S.: Indian smart TV market: a case of Realme's entry. J. Int. Bus. Educ. **16**, 237–254 (2021)
15. Pathak, P., Mehta, N., Sunil, A., Pandya, K.: On the way of making Alagrand.com a grand business! Emerald Emerg. Mark. Case Stud. **10**(4), 1–27 (2020). https://doi.org/10.1108/EEMCS-08-2019-0208
16. Hartono, B.D., Yusuf, B.P., Arifiati, A.: Implementation of digital marketing strategies through social media marketing, supply chain management and online sales of Bill Chilly product. Int. J. Supply Chain Manag. **9**(4) (2020)
17. Wismiarsi, T., Shihab, M.S., Hatta, H., Tjandra, N.: Startup branding: a case study of Pegipegi. Transnatl. Mark. J. **8**(1), 27–50 (2020). https://doi.org/10.33182/tmj.v8i1.721
18. Dushnitsky, G., Stroube, B.K.: Low-code entrepreneurship: Shopify and the alternative path to growth. J. Bus. Ventur. Insights. **16**, e00251 (2021). https://doi.org/10.1016/j.jbvi.2021.e00251
19. Rajan, T.: The Flipkart story in India: from the start to Walmart. Asian J. Manag. Cases. **18**(2), 126–143 (2021). https://doi.org/10.1177/0972820120914526

20. Sarva, M., Dube, P., Kashyap, S.K.: Startups can be the game changer for Indian economy. Int. J. Appl. Bus. Econ. Res. **14**(7) (2016)
21. Jiwa, S., Lavelle, D., Rose, A.: E-entrepreneurship: learning in a simulated environment. J. Electron. Commer. Organ. **3**(3), 42–56 (2005). https://doi.org/10.4018/jeco.2005070103
22. Frare, A.B., Beuren, I.M.: Fostering individual creativity in startups: comprehensive performance measurement systems, role clarity and strategic flexibility. Eur. Bus. Rev. **33**(6), 869–891 (2021). https://doi.org/10.1108/EBR-11-2020-0262
23. Ghura, A.S., Kumar, I.: Shradhanjali.com: creating Blue Ocean. Emerald Emerg. Mark. Case Stud. **10**(2), 1–16 (2020). https://doi.org/10.1108/EEMCS-05-2019-0133
24. Pejic-Bach, M., Knezevic, B., Pejic-Bach, M.: Development of a web based business oriented towards a market niche in an emerging economy. J. Cases Inf. Technol. **12**(2), 31–48 (2010). https://doi.org/10.4018/jcit.2010040103
25. Kenney, M., Curry, J.: E-commerce: implications for firm strategy and industry configuration. Ind. Innov. **6**(2), 131–151 (1999). https://doi.org/10.1080/13662719900000008
26. Wong, C.Y.: Celebrating IJPDLM's 50th anniversary: a reflection on its contributions and future directions. Int. J. Phys. Distrib. Logist. Manag. **51**(10), 1049–1064 (2021). https://doi.org/10.1108/IJPDLM-10-2021-0427
27. Sundmaeker, H.: Accelerating system development for the food chain: a portfolio of over 30 projects, aiming at impact and growth. J. Food Syst. Dyn. **7**(4) (2016). https://doi.org/10.18461/ijfsd.v7i4.747
28. Moirangthem, N.S., Nag, B.: Value-added activities of venture capitals in entrepreneurial finance: evidence from the growth of Flipkart. Asian J. Manag. Cases, 097282012199812 (2021). https://doi.org/10.1177/0972820121998120
29. Dessyana, A., Dwi Riyanti, B.P.: The influence of innovation and entrepreneurial self-efficacy to digital startup success. Int. Res. J. Bus. Stud. **10**(1), 57–68 (2017). https://doi.org/10.21632/irjbs.10.1.57-68
30. Chiasson, M.W.: The tail & the dog: agency and structure influences on the development of an e-commerce information system in a new venture startup. ACM SIGMIS Database DATABASE Adv. Inf. Syst. **33**(4), 24–37 (2002). https://doi.org/10.1145/590806.590811

A Strategy to Analyze the Metal Packaging Market in the Food Cans Industry Using Agent-Based Simulation

Luis F. Chaverra Vargas and José Antonio Marmolejo-Saucedo

Abstract Companies that manufacture metal packaging for canned jalapeño peppers require forecasts to determine the amounts of inventory needed to meet customer demand and anticipate sales opportunities. The nature of jalapeño pepper production depends on factors such as planting, harvesting, seasonality, pests, and climate change, among others, which generate a lot of uncertainty when estimating the needs of the client.

The purpose of this research is to solve the problem of having a simple and more accurate forecast of jalapeño pepper availability in the market for the metal packaging business. Forecasts are currently based on historical data; the adjustment of these predictions does not always follow sudden changes in the data, and for this reason, there are other more complex methods, which use mathematics, where it seems that a mathematical expert is needed more than an expert in the business where the forecasting tool is used.

Agent-based simulation allows the development of interaction rules between the different market agents, which, in a dynamic and flexible way, allows the development of simple models with complex behaviors where there are no formulas that govern them and allow interaction between them.

The ABS model provides more details about the behavior of the jalapeño pepper crop, compared to the models evaluated in this chapter, which are based on equations.

Keywords SEMARNAT (secretary of environment and natural resources) · INEGI (national institute of geography and statistics) · ABS (agent-based simulation)

L. F. Chaverra Vargas (✉)
Universidad Panamericana, Facultad de Ingeniería, Ciudad de Mexico, Mexico

J. A. Marmolejo-Saucedo
Facultad de Ingenieria, Universidad Nacional Autonoma de Mexico, Ciudad de Mexico, Mexico
e-mail: jose.marmolejo@fi.unam.edu

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_9

1 Introduction

The evolution of current markets, the need to optimize both human and physical resources, the growing demand for products that go beyond just covering a primary need, the insistent pressure on manufacturing costs, and the demand for raw materials friendly to the environment are some of the market factors that put pressure on competition between companies with similar products.

The first Industrial Revolution marked the transition from manual manufacturing to mechanized manufacturing between the years of 1760 and 1830, the second revolution brought with it electricity and therefore mass production between the 1850s, the third revolution brought electronics, information technology, and telecommunications only from the middle of the twentieth century, and now the fourth revolution brings everything related to the automation of production processes and the creation of intelligent networks that can be controlled by themselves [29].

The fourth revolution has origins around the automotive industry and has been expanding to other sectors; the basis of this revolution is data, and their analysis will guarantee timely decision-making and, therefore, the financial performance of companies.

The food and beverage industry in Mexico continues the path of growth that has marked the last decade. The rise of Mexican cuisine at a global level and the internationalization of many Mexican food companies, together with strong domestic consumption, have placed the industry in the crosshairs of foreign investment. The NAFTA continues to be an unknown in terms of how it can affect the sector in Mexico, but the latest news coming from both Mexico and Canada suggests some optimism. According to statistics published by the National Institute of Statistics and Geography (INEGI), the food, beverage, and tobacco sector accounted for 3.5% of Mexico's gross domestic product in 2017. The outlook is positive for the industry for the year 2020. Although the statistics vary, the ANTAD puts the growth rate of the industry at 4.3% for the period from 2014 to 2020. On the other hand, other sources determine that the growth in value of the industry for the year 2020 will be 7.6%.

In the last four years, there has been an increase in agricultural production of 11% according to SAGARPA data. This increase in production corresponds to the 52 main crops and is mainly due to fruit crops that increased by 22%. But there were also significant increases in vegetables (21%), grains (13.4%), agro-industrial crops (9.6%), oilseeds (8.8%), and fodder (5.7). SAGARPA also reported that there was also an increase in the percentage production of many other crops such as soybeans (105.7%), apples (91.2%), asparagus (81%), blackberries (77.8%), broccoli (51.7%), avocado (43.6%), palay rice (42.1%), cucumber (38.4%), green chili (37.8%), and fodder corn (34%).

This increase in production has also resulted in an increase in exports in 2017. The agri-food sector obtained 32.583 million dollars in sales of exported products, a figure that represents an increase of 12.47% compared to 2016. According to data from the Bank of Mexico, the products that increased the most in export volume were citrus fruits with 37.2%, tomato (or tomato as it is known in Mexico) with 27.4%, and fresh legumes and vegetables with 7.6% [92].

For the development of this work, the focus as a case study is the jalapeño pepper market for canning, which is one of the main peppers that is produced and consumed in Mexico.

Chili is one of the crops native to Mexico and one of the most important worldwide. Its different varieties adapt to different climates and types of soil, which has contributed to its successful and wide geographical distribution. The multiple uses of chili and its derivatives date back to pre-Hispanic times and go beyond forming an extraordinary condiment. Today, its production and cultivation in Mexico are still relevant, although due to biological and technical factors, the country has been at a disadvantage, mainly compared to Asian producers.

Among horticultural crops, chili is the most important at the national level. Mexico is the country with the largest genetic variety of capsicum, which is produced in the 32 states of the Republic; however, the main producing states are Chihuahua, Sinaloa, Guanajuato, Zacatecas, and Sonora. The most cultivated varieties are jalapeño, serrano, poblano, bell pepper, and habanero. In the country, there are around 150,000 hectares planted with more than two million tons of dry and green chili per year [95]. This represents a commercial value of approximately 13,224 million pesos. Mexico ranks as the main exporter of green chili on an international scale and is the second largest producer in the world. The country is divided into six green chili producing zones, which differ from one another by the types of chilies that are planted. In the Gulf area (Veracruz and Tamaulipas), jalapeños and serranos are mostly obtained; in the south (Yucatán and Tabasco), there are jalapeños, costeños, and habaneros; in the Bajío area (Guanajuato, Jalisco and Michoacán), anchos, mulatos, and pasilla are grown; in the central table (Puebla and Hidalgo), they specialize in poblanos, miahuatecos, and carricillos; in the north (Chihuahua and Zacatecas), they produce jalapeños, mirasol, and anchos; and in the North Pacific zone (Baja California, Sinaloa, and Sonora), they have bell, anaheim, jalapeños, and caribes [45].

When we talk about agricultural-based products, there are a number of variables that in one way or another affect availability, size, quality, and harvest time and that in one way or another cause an oversupply due to good harvesting and harvesting conditions or on the contrary product shortages due to poor conditions. All these factors, in addition to affecting product availability in the market, directly affect the sale price both for the wholesaler, who reprocesses it to make a processed product, and for the final consumer, such as housewives for homemade preparation.

Being able to be clear about how the market behaves is one of the needs of companies to be able to optimize their processes and be able to achieve good customer satisfaction; which allows you to improve your sales and therefore improve your profits.

The evolution in forecasts has been influenced by the use of technologies. Today, companies that use big data from a predictive level are still a minority; however, they are mature in their use of data, guided by more accurate forecasts and leaders with statistical skills [97].

In Mexico, a low number of about 20% of mature companies use big data from a level of predictive and adaptive analysis to design agile supply chains, driven by the flow of information. The role that forecasts currently play in data analysis is more precise because in big data it adds other variables depending on the type of organization, such as temperature, social networks, changes in monetary policy, etc.; with those, which are more unstructured, the idea is that you have more accurate forecasts. Traditionally, it has been the area where it is used, but there is going to be a new wave where that data is used, not necessarily unstructured, but managed to have better precision [97].

As Bauza argues "Do not use the data to see what happened, but to determine what will happen." Currently, many companies base their forecasts on past events, but this places them in a complicated situation due to the number of variables that can cause significant fluctuations in the behavior of demand, and they can run the risk of not having what the client really needs, causing unwanted inventories and therefore affecting capital.

2 Background and State of the Art

Most companies can forecast the total demand for all products, as a group, with errors of less than 5%. However, forecasting the demand for a product can lead to considerably larger errors. A forecast is simply a prediction of what will happen in the future. Managers must learn to accept the fact that no matter what technique is used, they will not be able to produce perfect forecasts [5]. There are different demand forecasting methods, which differ in both the form and the types of data associated with the forecast, as well as the characteristics of the markets that are evaluated. Figure 1 shows the different types of methods that are currently used by companies to make their forecasts.

Forecasting methods are classified as quantitative or qualitative.

Quantitative methods are used when:

(i) Past information is available on the variable to be forecast.
(ii) Information can be quantified.
(iii) It is reasonable to assume that the pattern of the past will continue to occur in the future. In these cases, a forecast can be made with a time series method or a causal method.

If the historical data are restricted to past values of the variable we are trying to forecast, the forecasting procedure is called a time series method. The goal of time series methods is to discover a pattern in historical data and then extrapolate it into the future; the forecast is based only on past values of the variable we are trying to forecast or on past errors [5].

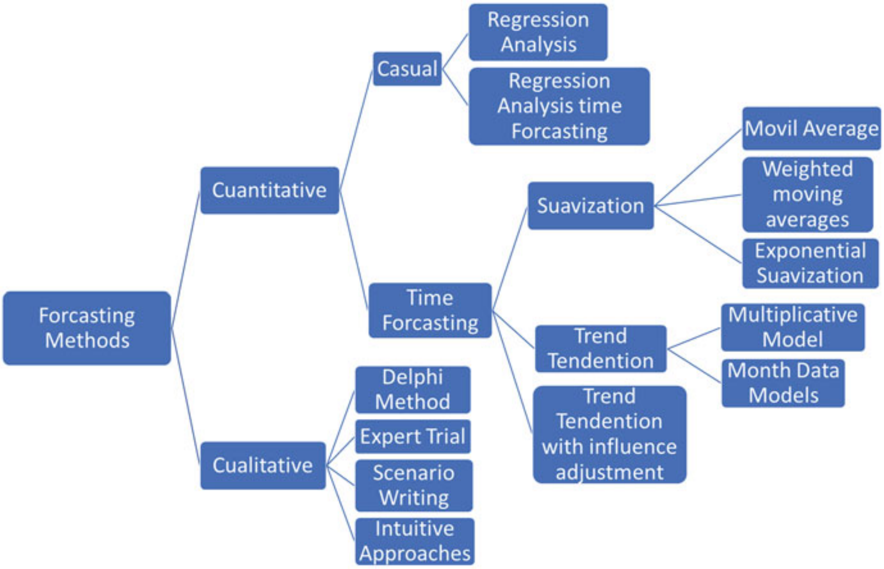


Fig. 1 Forecast estimation methods, own creation based on [5]

Causal forecasting methods are based on the assumption that the variable we are trying to forecast exhibits a cause-and-effect relationship with other variables.

Qualitative methods generally involve the use of expert judgment to make forecasts. An advantage of qualitative procedures is that they can be applied when information about the variable being forecast cannot be quantified or data about it are scarce.

In the case of this research, qualitative methods will not be treated since the type of market that is being treated has sufficient information to use a quantitative method. Table 1 shows the most significant generalities of the most used methods, where it can be seen that starting from the moving average method, which is the most basic, the following add some component that allows adjustments, to which the previous models did not contain, or for the applications that were established it was not required.

The first 5 methods only estimate one variable at a time, but they are very limited since the analyzed variable is estimated to only change due to its own behavior; the linear regression methods allow estimations based on several variables such as the simple regression model and with additional parameters such as time series such as time series regression methods; but with some limitations since the variables must have linear behavior.

There are some works that seek to improve the weaknesses that the aforementioned models present, but they have a mathematical complexity that implies very specific knowledge such as differential equations, genetic algorithms and programming, below they are presented in a general way:

Table 1 Differences between forecast estimation methods, own creation based on [5]

Quantitative method	Advantages	Disadvantages
Moving averages	– It is a direct calculation	– It is a direct calculation
	– Provides a smoother line, less prone to whipping up and down in response to slight, temporary oscillations back and forth	– Provides a smoother line, less prone to whipping up and down in response to slight, temporary oscillations back and forth
Weighted moving averages	– Being weighted with the most recent data changes, it responds more quickly to changes	– It is very sensitive to false data
Suavización exponencial	– Simple and requires a minimum of data	– The series that present trends alter their results
	– Gives more weight to the closest values than to the oldest data	
Multiplicative model	– Adds an element of temporality which allows to assume sudden changes in the variables	– The calculation of the irregular component I as well as that of the seasonality component and depends on the experience of the variable
Models based on periodic data	– Models can be created that have more than one independent variable	– The relationship between the variables must be linear
	– You can make a prediction of the behavior of some variable at a certain point or moment	– The errors in the measurement of the explanatory variables are independent of each other
		– The total error is the sum of all errors
Regression analysis	– The temporary component allows to make adjustments of sudden changes of variable	– The relationship between the variables must be linear
	– You can work with ranges of periods	– The errors in the measurement of the explanatory variables are independent of each other
		– The total error is the sum of all errors

(i) Bayesian Methods: It is an approximate reasoning method in which there is uncertainty and admits a certain degree of variation. They are based on Bayes´theorem [93]. Nadini and Tokimatsu made estimates of fossil fuel consumption to assess the impact of climate change [80] Domit, Monti and Sokol made projections of the growth of the economy to estimate the behavior of UK inflation [32].

(ii) Fuzzy Logic: It is a heuristic method and is based on the relative of what is observed as a differential position. It is used when the complexity of the process in question is very high and there are no precise mathematical models, for highly nonlinear processes and when definitions and knowledge that are not strictly defined are involved. On the other hand, it is not a good idea to use it when some mathematical model already efficiently solves the problem, when the problems are linear, or when they have no solution [7]. Killian, Mayer, and Kozek use this method to estimate climate behavior and thereby optimize energy consumption in an air conditioning system [60]. Lia and Cheng used this method to work with projections where insufficient data existed [67].

(iii) Support Vector Machines (SVMs): They are supervised learning algorithms, normally used for classification and regression problems. Given a set of training examples (data), we can label the classes and train an SVM to build a model that predicts the class of a new sample [4]. Guoa, Lia, and Zhang used this method to estimate oil prices [42]. Chena, Tana, and Song used the method to estimate electricity consumption as a tool for the construction of power generators [24]. Zhanga, Wanga, and Zhang used the method to estimate electrical load to plan power generation [124].

(iv) Autoregressive Integrated Moving Average Methods (ARIMAs): It is a statistical model that uses variations and regressions of statistical data in order to find patterns for a prediction toward the future. It is a dynamic time series model, that is, future estimates are explained by past data and not by independent variables [86]. Widowatia, Putrob, Koshioc, and Oktaferdian used this model to predict weather conditions and assess the impact of water pollution [119]. Carvalho, Monteiro, and Sá-Soares used this method to estimate patient demand in an emergency center [102].

(v) Gray Models: They are based on the gray theory developed by Deng [103]. They are statistical models used to work with scarce data. It is considered a short-term method since its results can be well adjusted to very close data [54]. Wang used the method to forecast electricity demand by combining it with a Gaussian quadrature formula [115], and in another paper, he used the method to forecast power demand using a Gauss–Chebyshev formula [116]. Songa, Zhangb, Zhanga, Xiac, and Miao used the method to forecast power consumption in Hebei province based on an IOWA operator [106]. Chang, Lin, and Jin used the model considering few data to make the manufacturing demand projection statistic [21]. Hu and Jiang used this method in combination with neural networks to estimate energy demand [48]. Tsai, Lee, and Guo used the method to estimate the production of green raw materials and to estimate the reduction of polluting emissions by comparing 3 gray models: GM(1,1), Bernoulli Nonlinear, and Verhulst model [113].

(vi) Other Models: There are other types of works where some characteristics of the previous techniques vary and new models are generated. Dong, Zheng, and LI define their method as predictive control, in which they seek to optimize inventory volumes in the supply chain [33]. Azimian and Aouni defined their method as Stochastic Goal Programming (SGP) with which they predict for supply chain decision-making [8]. Bildirici and Bakirtas called their model Autoregressive Distributed Lag Frontier (ARDL) with which they estimate growth in countries such as Brazil, Russia, India, China, Turkey, and South Africa with the ratio of coal, gas, and oil consumption [11]. Han, Chen, Long, and Cui used several models used for the estimation of the available

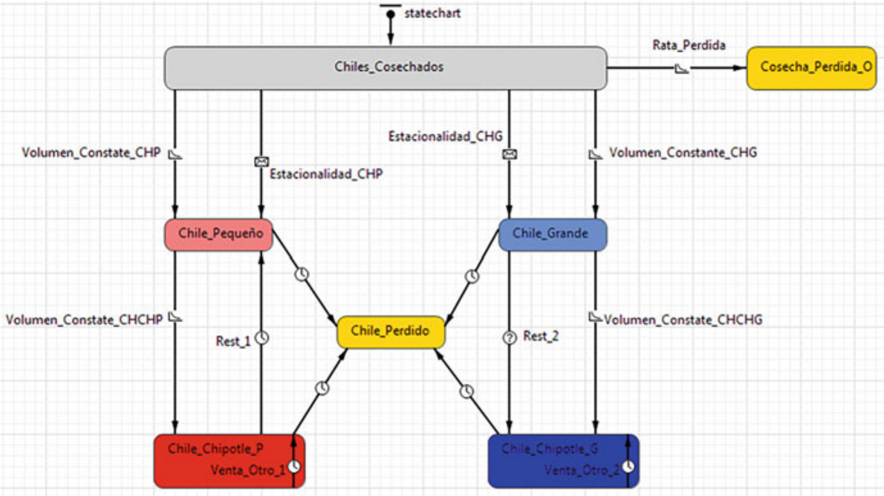


Fig. 2 Representation of an agent showing the evolution of a crop, own creation in AnyLogic 8.0 software

amounts of coal in China, where they evaluated methods such as Hubbert's Method, Gaussian Curves, Logistic Growth Model, and Modified Curve Fitting Model [44]. Nair, Rashidi, and Dixit used linear regression methods, structural equation modeling, and neural networks to predict the distribution of humanitarian aid to the most needy sectors based on the donations received [81]. Shih, Hsu, Yeh, and Lee made a modification to the traditional gray algorithm to achieve a better prediction of the stock market [101]. Zhemin, Ganqiong, and Wang used a method called the Short-Term Forecasting Model to estimate the egg price in the Chinese market [125].

In the documents investigated, we can see a great variety of methods: some very simple to apply but with great opportunities for precision, and others, with considerable mathematical complexity where it implies having more mathematical and programming skills than knowing the same process that is being evaluated. In addition, as you can see, all these methods work with static data, that is, we already have the historical data beforehand and only the mathematical operations are performed to make the predictions.

The methodology proposed in this research (Agent-Based Simulation) is a simple tool and is based on the behavior of the treated agent. In Fig. 2, you can see an example of the representation of an agent in the AnyLogic 8.0 software (this agent represents a chili cultivation area).

All the relations that exist in this representation show the states in which the chili harvest can be depending on how the simulation is defined. This is the important advantage that the software presents since we can include all the variables that we want with the conditions of change depending on time, seasonality, or other factors that can affect the model; the use of dynamic data allows this method to have all the flexibility required depending on the situation being modeled.

In the documentary analysis, 3 types of articles were found that generally associate research related to ABS applications at a general level. The first of them is related to publications that refer to a theoretical framework, in the other type of publications there is a focus on applications with respect to a specific area of study, and finally there are publications related to case studies.

2.1 Documentary Evaluation ABS Theoretical Framework

In this research, articles are considered where the theoretical concepts regarding agent-based simulation are the main topic.

The articles that were found in the research as a theoretical framework, in general, deal with topics with applications where they combine artificial intelligence [15, 90], treat simulation as a way to develop KPI's [35], methodologically talk about the characteristics around the key concepts to take into account when performing agent-based simulations [23, 26, 38, 47, 55, 65, 82, 98, 104, 109, 122], evaluate some mathematical models [61], evaluate production scheduling concepts for the manufacturing area [1, 62, 76], and conceptually validate some algorithms developed for optimization [2, 3, 28, 41, 43, 51, 53, 66, 88, 108].

2.2 Application Documentary Evaluation

Other general categories considered were those of a social and industrial scope, where theory and practice constitute two autonomous realities that manage knowledge in different ways and develop in different contexts, through the generation and application of new knowledge in the fields where common problems are faced, with specific and innovative solutions.

Within this review, we can see applications where simulation models issues related to the behavior of people in relation to the environment are seen [10, 14, 18, 19, 30, 34, 71, 72, 83, 96, 99, 100], applications to the industrial environment where processes are evaluated [6, 49, 73, 87, 110, 121], evaluations of new technologies such as the impact of autonomous vehicles when interacting with members of a society [12, 56, 57], evaluating environmental impacts [63, 64, 68, 84, 91, 117, 121], particular cases such as simulations of medieval battles [9], articles that validate market conditions for business analysis [25, 69, 74, 89], evacuation simulations to evaluate risk conditions [58, 70, 126].

2.3 Documentary Case Study Evaluation

In this review, there are articles such as studies on technologies and new technological developments in health sciences [27, 31, 52], evaluations of situations in the way businesses evolve [20, 40, 123], evaluations of conducts and behaviors analyzed from the perspective of the impact on the population in general [13, 59, 78, 79, 85, 94, 111, 112, 114, 120], evaluations of environmental situations [22, 77], medical emergencies, natural disasters, and disease proliferation [75, 107, 118].

2.4 Summary Documentary Analysis

In general, in the figures shown below, the distribution of research associated with ABS can be seen according to the documents that were consulted for the definition of the state of the art.

Figure 3 shows the timeline of the publication of all the articles consulted, grouped into the 3 different types" reviewed in points 2.1 (Theoretical framework), 2.2 (Applications), and 2.3 (Case Study).

In Fig. 4, the same timeline is observed depending on the type of industry to which the evaluated documents on ABS refer.

Figure 5 shows the percentage contribution of the articles based on the total consulted.

Finally, as the interest of this state of the art is to validate the research on the application of ABS in the Manufacturing Industry, in Fig. 6, the timeline of the publications focused on the Manufacturing Industry is observed and in Fig. 7 the specific applications in the manufacturing area.

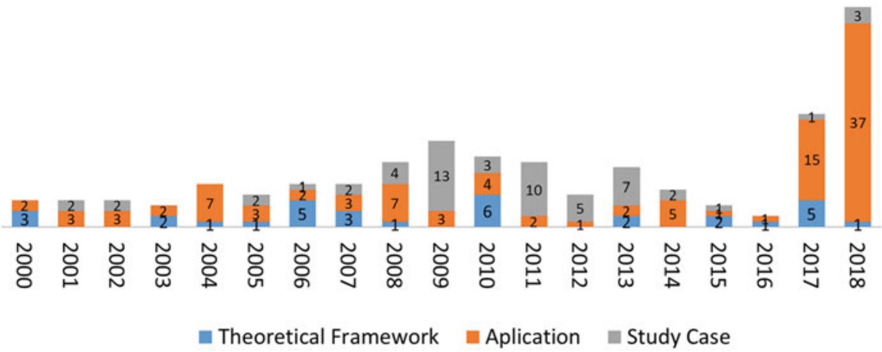


Fig. 3 Timeline research articles on agent-based simulation, own creation

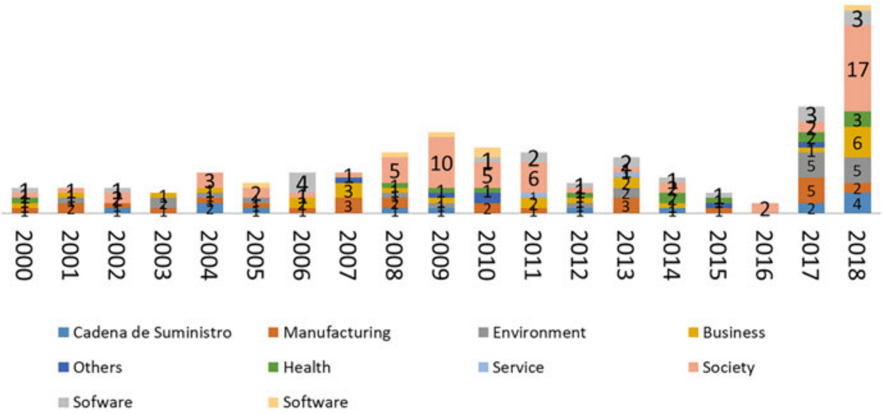
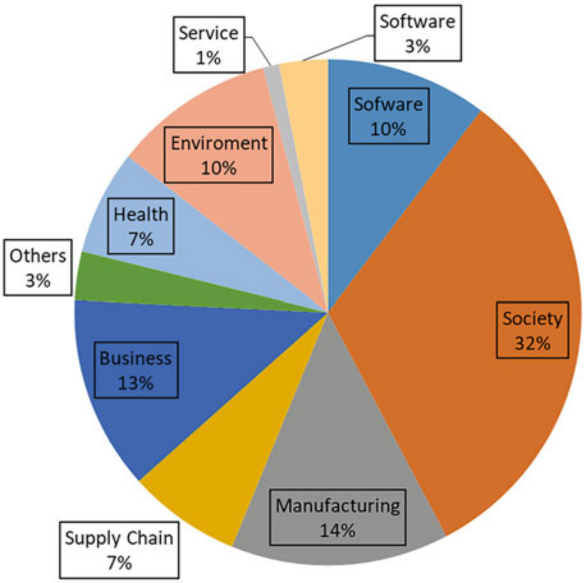


Fig. 4 Timeline of research articles on agent-based simulation according to industry (own creation)

Fig. 5 Distribution of articles by study area (own creation)



2.5 Methodology

Simulation is the imitation of the operation of a real system during a time interval. This simulation can be done either manually or computationally. The simulation is based on a model of reality that tells a story and by observing its behavior, and it allows us to obtain knowledge about the real system. The behavior of the simulation is determined by the simulation model or a set of assumptions concerning the real system, and these assumptions are expressed through logical and mathematical

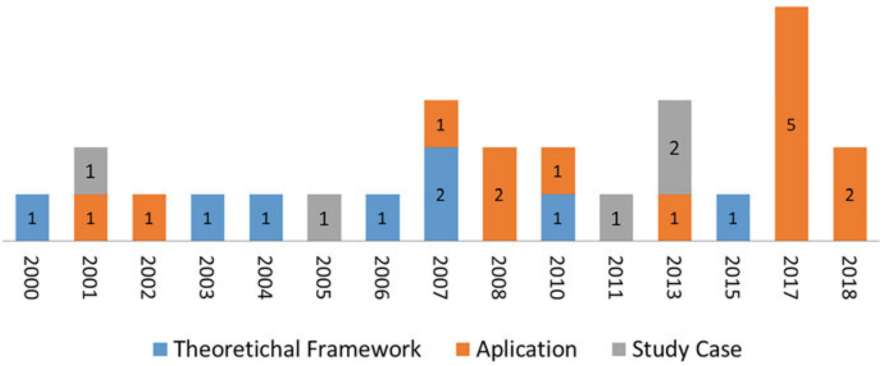


Fig. 6 Timeline research articles focused on manufacturing (own creation)

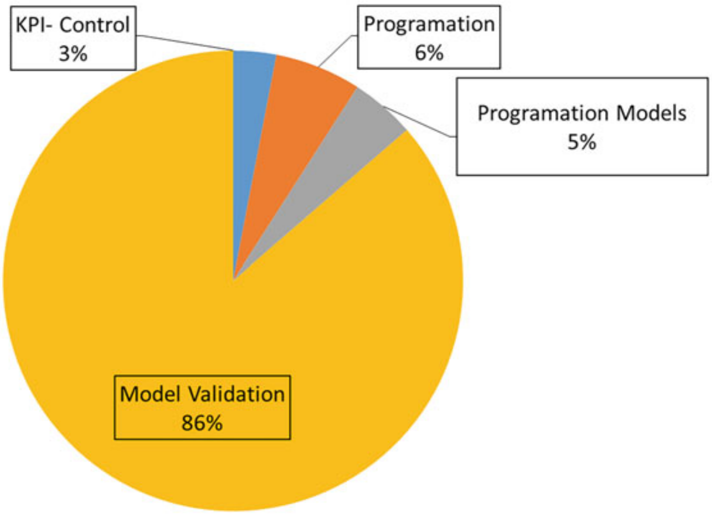


Fig. 7 Focus articles on manufacturing (own creation)

relationships between entities. By simulating the real situation in a controlled way on the computer, it is possible to answer what could be the causes of, for example, certain heart diseases, even, going one step further, computer simulation could help predict certain behaviors (the evolution of the disease that has been simulated or also the effect of a drug some time after its administration, the most optimal programming that could be achieved for a certain product or service, and the behavior of a market with respect to certain social variables, among others).

It can be said that simulation has the following advantages:

(i) Simulation allows solutions to be found where analytical methods fail.

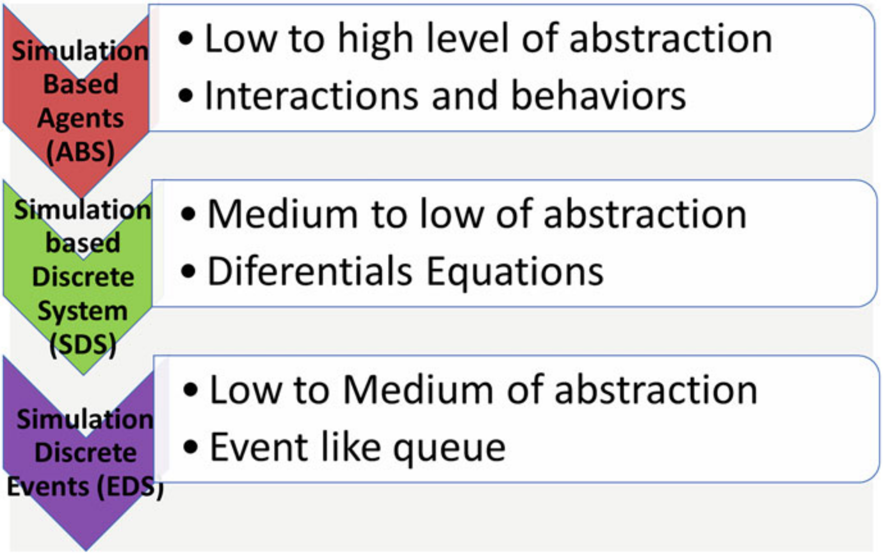


Fig. 8 Generalities different types of simulations, own creation based on [16]

(ii) When the model has been defined, it requires very little effort to search for alternatives, in addition to being scalable, modular, and incremental.
(iii) The structure of a model reflects the actual structure of the system.
(iv) Any variable can be measured at any time during the simulation.
(v) The ability to play with the behavior of the system in real time is one of the biggest advantages of simulation.
(vi) Viewing a simulation is more convincing than any method of modeling a process.

There are 3 types of simulation methods [16]:

(i) Simulation of Discrete Events (EDSs)
(ii) Simulation by Dynamic Systems (SDSs)
(iii) Agent-Based Simulation (ABS)

EDS is commonly used to model systems that change stochastically and dynamically on a process basis. SDS is one of the most widely used simulation methods for continuous systems and is based on differential equations, where its most important feature is feedback. ABS is the newer and more detailed method than EDS and SDS; it is based on organization and communication (Fig. 8).

ABS is the most recent simulation method, until the year 2000, and this method was more than an academic topic. The adoption of this method as an application began between the years 2002 and 2003, and this was due to the need to delve into systems that were not well defined by traditional approaches, the advancement of computer science, especially in object-oriented modeling, and the rapid growth

Fig. 9 Agent relations, own creation based on [16]

of CPU, power, and memory availability (agent-based simulation demands more memory than traditional methods such as discrete events and dynamic systems).

There is no standard language for ABS; the structure of an ABS is created using graphical editors or a script depending on the software. Agent behavior is specified in different ways. Normally, the agent has a notion of state, and his actions and reactions depend on his state. Therefore, their behavior is best defined as state charts; sometimes, behavior is defined in the form of rules that depend on a specific behavior. One of the most important features of ABS is the high level of abstraction it can achieve compared to other types of simulation where individualities are aggregated based on features (Fig. 9).

AB modeling is one of the newer simulation models. It was not until the year 2000 that AB ceased to be just an academic case study. However, AB became popular day by day as it is a flexible and efficient system, usable for complex and dynamic systems, which can be replicated in many disciplines (anthropology, consumer behavior analysis, economics, manufacturing, military, physical sciences, sociology, etc.).

The elements of AB are such as environmental conditions (housing, employment, transportation, etc.) Agents can be represented by living beings (people, animals, etc.), vehicles (cars, trucks, ships, etc.), equipment (machines, robots, etc.), material things (projects, ideas, etc.), and organizations (cities, companies, etc.).

Each of these agents has a state, attribute, behavior, object, and memory. The agents are intelligent, and they can make and change decisions, controlling themselves, which is why it is mentioned that these agents are decentralized.

They have different parameters and variables and communicate with each other and with their environment. The agents' reactions depend on their state. In a system, there is no requirement that all agents have to be the same object. So, one agent can be a car and another agent can be a car or the customer (Fig. 10).

ABS models demolished traditional top-down modeling. We can determine the interaction, behavior, and communication systems of agents, so top-down study cannot help solve complex problems.

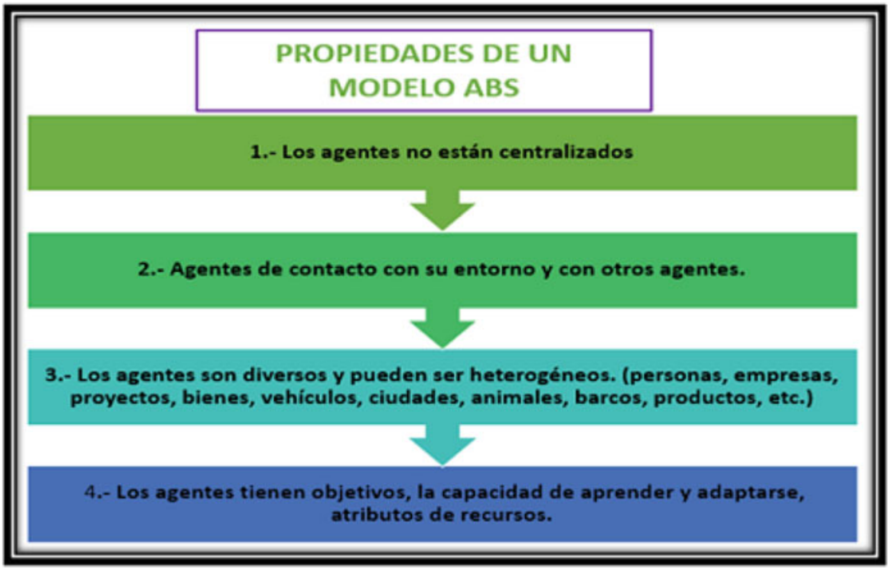


Fig. 10 Properties of an agent-based model, own creation based on [17]

2.6 Significant Differences Between Models

(i) The DE model is used for discrete systems, while AB and SD are used for continuous systems.
(ii) The DE model is based on processes, AB is based on agents, and SD is based on the changing results of equations instantly.
(iii) The most important part of SD is the feedback.
(iv) The DE model is usually identified by flow charts, while AB is usually represented by statutes.
(v) The AB simulation method is the most detailed of all the methods. Therefore, it helps to solve more complex systems that have higher variability.

2.7 Similarities Between AB, DE, and SD Models

(i) There are many visual techniques for all simulation methods and they are provided to make it easier to translate the conceptual model into a computerized model.
(ii) There are various simulation software programs for all methods, and some programs can be used for the same methods, for example, AB (Repast, NetLogo, AnyLogic, etc.), DE (Sand, Micro Saint Sharp, etc.), and SD (OptiSim, TRUE, AnyLogic, etc.).

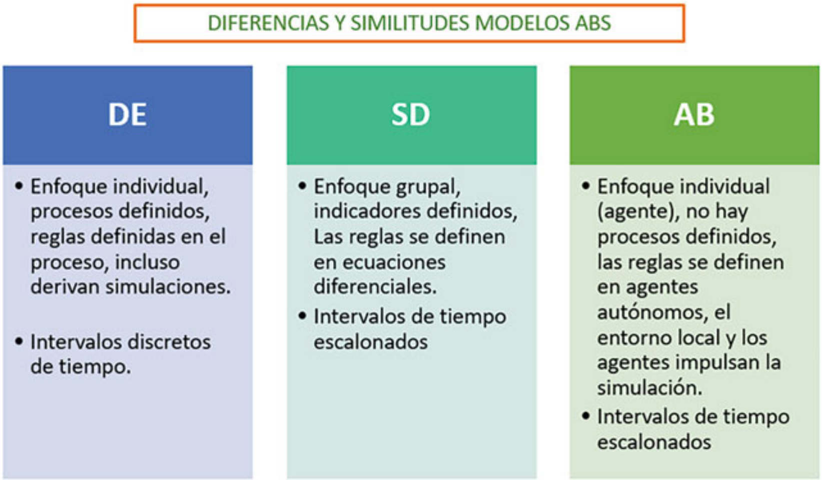


Fig. 11 Characteristics between simulation models, own creation based on [17]

(iii) AB and SD are usable for dynamic and continuous event systems. Also, they both have objects that have memory.

On the other hand, a table has been made on comparisons of DE, SD, and AB [17]. It is shown in Fig. 11.

2.8 Principle and Use of ABS Models

The working principle of ABS is that an agent is randomly selected, its behavior is executed, and then it affects other agents and their environment. All states of the agents are updated and it is repeated until the end of the simulation time horizon, and finally it is finished and a report is generated (Fig. 12).

2.9 Use of ABS Models

(i) For a problem that makes you a good candidate for an AB application
(ii) When the problem has a natural representation as agents
(iii) When the goal is to model the behaviors of individuals in a diverse population
(iv) When agents have relationships with other agents, especially dynamic relationships: relationships between agents form and dissipate, e.g., structured contact, social networks

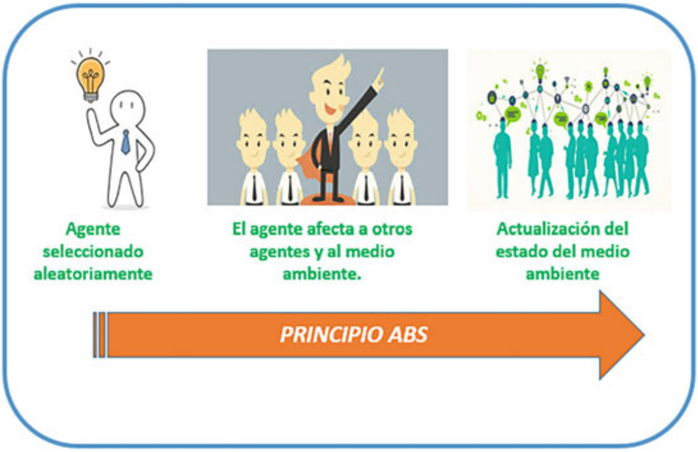


Fig. 12 ABS principle, own creation based on [37]

(v) When it is important that individual agents have spatial or geospatial aspects of their behaviors (e.g., agents move over a landscape)
(vi) When it is important for agents to learn or adapt or for populations to adapt
(vii) When agents engage in strategic behavior and anticipate the reactions of other agents when making their decisions
(viii) When it is important to model agents that cooperate, collude, or form organizations
(ix) When the past is not a predictor of the future (e.g., new markets that do not currently exist)
(x) When extension to arbitrary levels is important, that is, extensibility
(xi) When the structural change of the process must be an output of the model, rather than an input to the model (e.g., agents decide which process to move on to next)

2.10 ABS Model Simulation Stages

First, the data are collected and prepared for modeling. Then, the model is conceptualized; that step includes determining the agents, initial states of the agents, environment, and the relationship between the agents and the environment (Fig. 13).

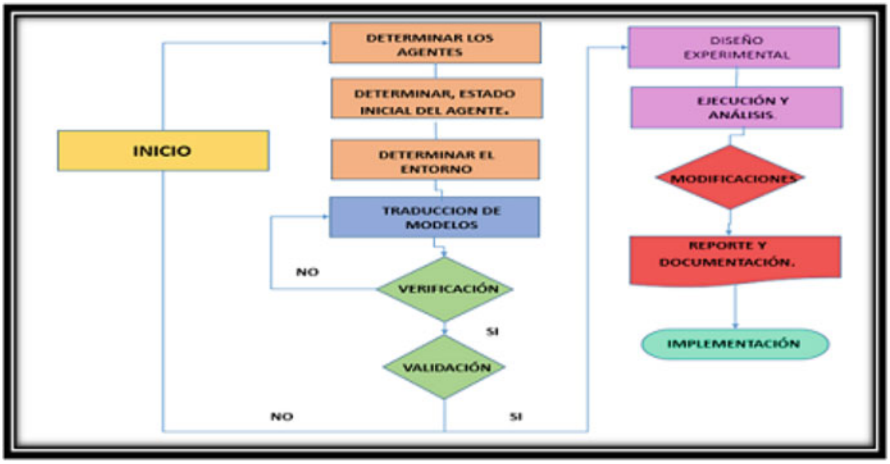


Fig. 13 Stages of an ABS model, own creation based on [118]

3 Results

Chili producers represent the physical places in Mexico where agriculture focused on the production of jalapeño peppers is developed. The jalapeño pepper is the most popular nationally and internationally. The fresh harvest is known as jalapeño or cuaresmeño. 60% of the production is used in the pickles industry, 20% is consumed fresh, and the rest, in a mature state, is processed by drying and smoking to obtain the chipotle chili [46].

The climate for the cultivation of jalapeño pepper is warm and with temperatures between 10 and 35 °C; the relief, the soil conditions, the doses of fertilizers, and the available water are the factors that determine the quantity and quality of the harvest obtained [36].

The main producing states are Baja California, Baja California Sur, Campeche, Chiapas, Chihuahua, Colima, Durango, Guanajuato, Guerrero, Hidalgo, Jalisco, State of Mexico, Michoacán, Morelos, Nayarit, Nuevo León, Oaxaca, Querétaro, Quintana Roo, San Luis Potosí, Sinaloa, Sonora, Tabasco, Tamaulipas, Veracruz, Yucatán, and Zacatecas. In Fig. 16, you can see the historical production of Jalapeño Peppers at the national level where we can see that by 2017 production has increased by almost 45.22% compared to 2008 (Fig. 14).

Figure 15 shows the National production adding the fraction of jalapeño pepper that is processed to turn it into chipotle pepper.

Figure 16 shows the historical production by States, where it can be seen that the States of Chihuahua, Sinaloa, Michoacán, Jalisco, and Tamaulipas have 80% of the historical production of jalapeño pepper [50].

For purposes of simplicity of the analysis, only the states that represent 80% of the Jalapeño pepper production were taken, as seen in the previous graph.

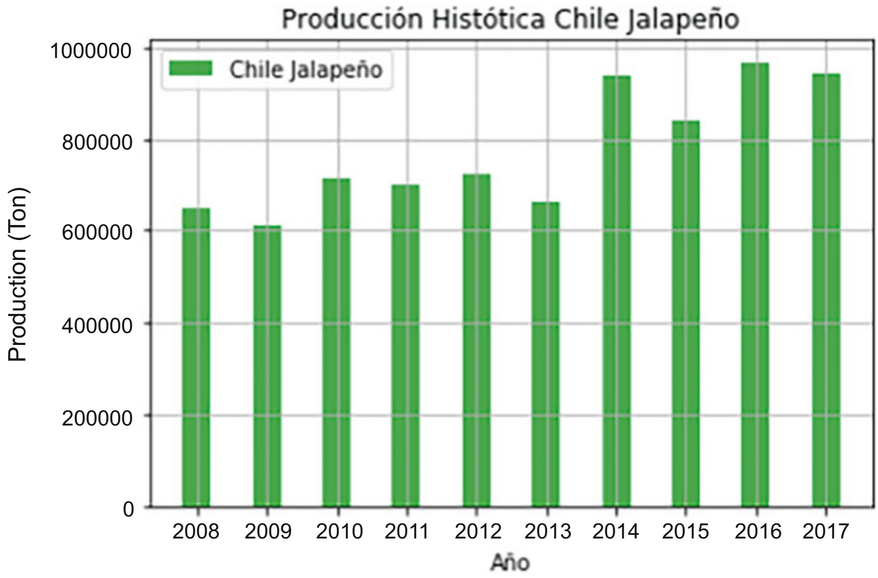


Fig. 14 Historical national production of Jalapeño pepper, own creation with Python 3.6 based on [50]

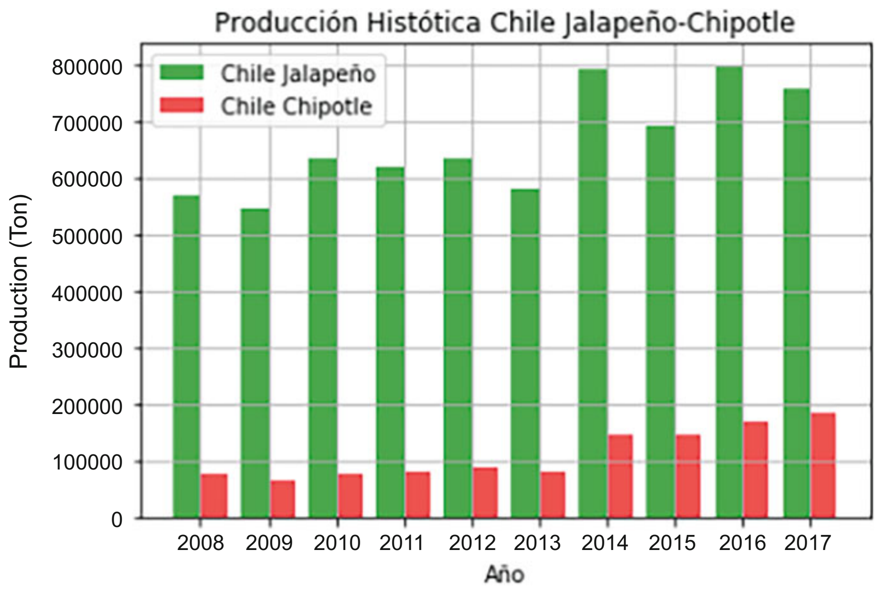


Fig. 15 Historical national production of Jalapeño pepper and Chipotle pepper, own creation with Python 3.6 based on [50]

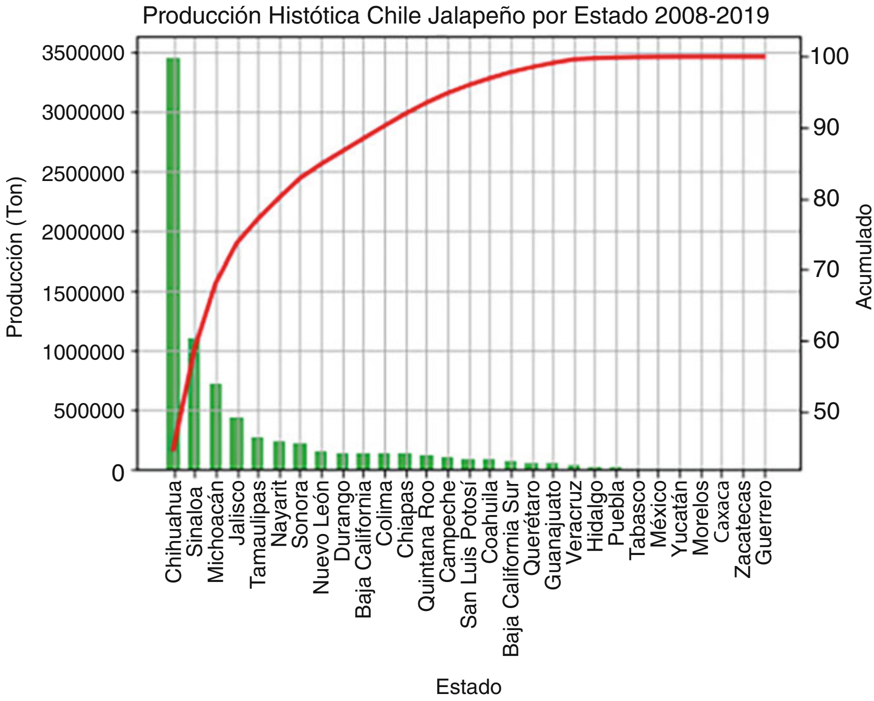


Fig. 16 Historical national production by state, own creation with Python 3.6 based on [50]

In Fig. 17, we can see the historical variations of the main monthly variables of the State of Chihuahua, where it can be clearly seen that the months of July and August are the months where there is the greatest production in this state, but it is also possible to notice a significant variation with respect to the other months, and the trend of temperatures tends to rise in the months of June–September where both maximum and minimum temperatures reach their maximum values; regarding the rains, the tendency of the months June–September to present the highest rainfall is very marked. In addition, in Table 2, you can see the numerical data with the basic statistics.

In Fig. 18, you can see the main variables for the State of Sinaloa. We can see that in the months of July and August the highest volume of production is presented, with respect to temperatures, and it is possible to observe the convergence of the values of both maximum and minimum toward an average of 30 °C and the range between their averages, the least of all other months. In relation to the rains, we see that around the months of June–October the greatest rainfall occurs with peaks in July and September. If we look at the production versus the amount of rain, they have a very similar behavior except for the months where the rains increase significantly.

In Fig. 19, you can see the main variables for the State of Michoacán. July and August are the months where the highest production values are presented, with

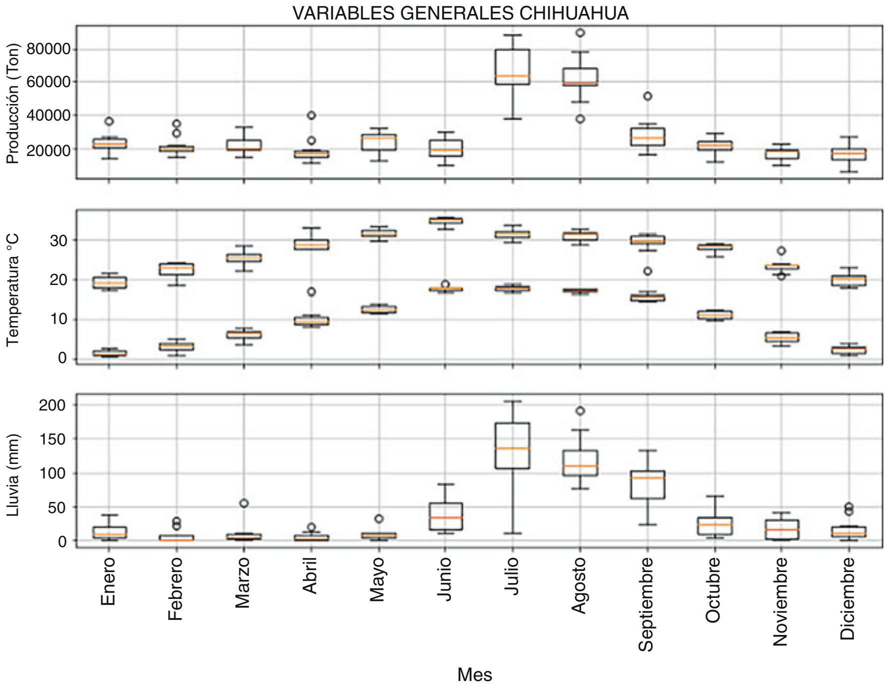


Fig. 17 Historical variations main variables Chihuahua, own creation with Python 3.6 based on [50, 105]

approximately double the dispersion in relation to the other months of the year. Heavy rains start from May to October and several atypical values can be seen throughout the year that do not seem to be related to temperatures. An interesting behavior of the highest temperature is observed from the month of June almost until the end of the year, although the minimum temperatures do not present such behavior. In Table 2, you can see the numerical values corresponding to the State of Sinaloa.

In Table 3, you can see the numerical data corresponding to the variables of the State of Sinaloa, where the review of the temperature ranges shows that there are variations around 4 °C for both maximum and minimum, as in the month of August, and the smallest variations of 1.7 °C and 1.5 °C, respectively, are presented.

In Fig. 20, you can see the main variables for the State of Michoacán where the trends of maximum production, rainfall has a similar trend to the previous states evaluated, except for the ascending and descending behavior of the maximum temperature between the months of May and August (Table 4).

In Table 5, you can see the numerical data corresponding to the State of Jalisco. We can observe that the temperature ranges have a greater variation reaching up to 5.2 °C in the maximum temperatures, and the minimum is around 2 °C.

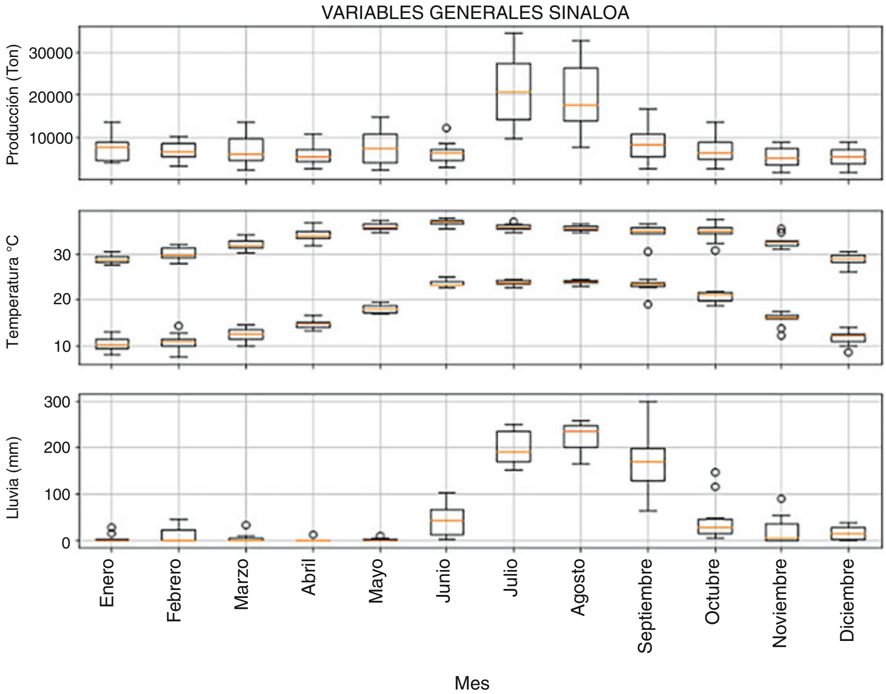


Fig. 18 Historical variations main variables Sinaloa, own creation with Python 3.6 based on [50, 105]

In Fig. 21, you can see the main variables for the State of Tamaulipas. The rainfall trend in this state does not present a trend as marked as the aforementioned states, although the coincidence of the June–October periods is preserved. The temperatures show a very similar pattern of both maximum and minimum and continuous throughout the year.

In Table 6, you can see the numerical data corresponding to the State of Tamaulipas.

In the analysis of the data by state, there is a clear and common trend of the months of greatest production and the coincidence with the months of greatest precipitation. All this is only what can be seen according to the graphs and the revised data, but for a better definition, a regression analysis was carried out with the data of each state, where in Figs. 22, 23, 24, 25 and 26 the relevance factors obtained can be seen. In Table 9 we can see the summary of all states with your prediction.

In Table 7, it can be seen that the fit of the regression is very poor since the maximum achieved was 0.51 and the minimum was 0.14, but the intention of this analysis is to understand the variables that weigh the most on the determination of the amount of Jalapeño pepper production, and as can be seen the amount of precipitation is very relevant in determining the amount of pepper produced, and

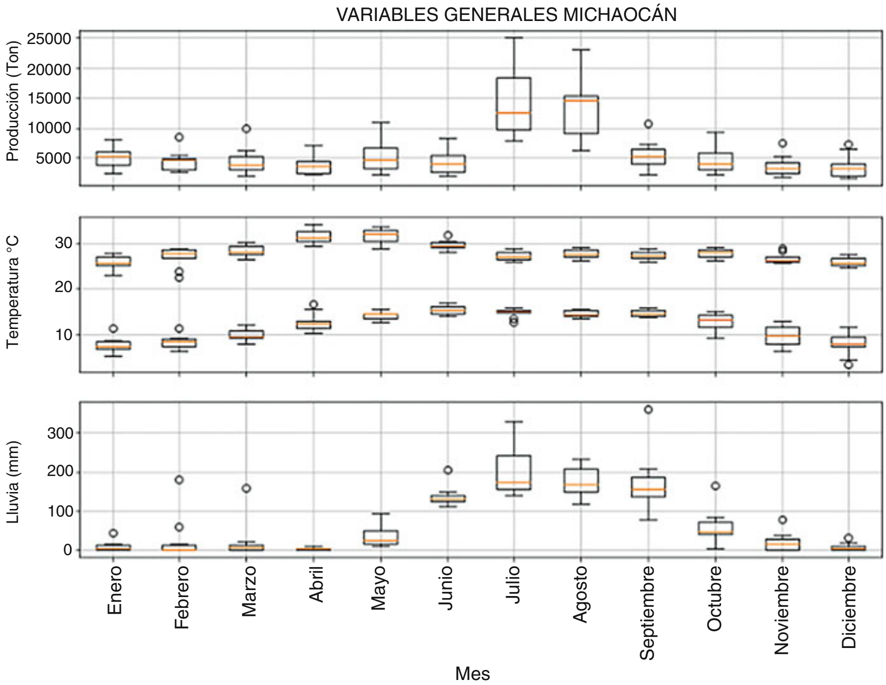


Fig. 19 Historical variations main variables Michoacán, own creation with Python 3.6 based on [50, 105]

then there is the lower temperature which is also very relevant in this determination, different from the maximum temperatures that do not have a significant representation compared to the other 2 variables.

Because precipitation is the most relevant factor for the jalapeño pepper harvest, but it is also clear that the production trends for each state have a similar behavior, the determination of the behaviors for the pepper harvest will be determined through 3 components: the first was called constant behavior where it refers to the constant minimum amounts that occur monthly according to the evaluated data, the second temporary behavior which refers to the periods where the increase in the chili harvest is evident, and finally a behavior of crop loss that refers to the amounts lost by affected areas.

Annual Analysis Behavior Constant Pepper Harvest As seen in the months of January–June and September–December, very constant harvest amounts can be seen, and therefore the probable distributions for each state were evaluated and thus define the behavior that will be used for the simulation. In Table 8, you can see the summary of the general parameters of the states; in general, it is evident that the average production of Chihuahua is well above that of the other states. In addition, we can see that the State of Chihuahua presents the least variation seen in the value

Table 2 General historical statistics monthly variables Chihuahua, own creation with Python 3.6 [50, 105]

Variable	Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Production	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	17,489	17,633	22,256	28,270	62,564	66,525	20,679	24,256	19,423	21,676	21,569	23,656
	Std	6,331	4,368	5,335	10,105	14,541	16,093	6,543	7,321	8,066	5,436	6,095	5,840
	Range	20,952	13,259	17,454	34,991	51,526	50,779	19,642	19,997	28,609	18,284	19,960	22,514
	Min	6,216	10,025	12,328	16,675	38,146	37,735	10,385	12,660	11,300	14,815	14,989	14,152
	25%	13,814	14,534	19,334	22,171	58,039	59,100	15,996	19,584	15,102	19,276	18,957	20,859
	50%	16,950	19,031	22,592	26,292	59,484	64,078	19,727	26,596	17,188	20,352	20,142	23,264
	75%	20,398	19,637	24,459	32,257	68,472	79,674	25,509	29,046	19,022	24,893	21,664	25,676
	Max	27,168	23,284	29,781	51,667	89,672	88,514	30,027	32,657	39,909	33,099	34,950	36,666
Max temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	19.2	22.5	25.4	29.3	31.6	34.8	31.6	31.2	29.9	28.1	23.3	20.0
	Std	1.5	1.9	2.0	1.8	1.0	1.1	1.2	1.3	1.3	1.1	1.8	1.6
	Range	4.1	5.8	6.6	5.5	3.5	3.1	4.2	4.0	4.3	3.6	6.5	5.3
	Min	17.3	18.6	22.0	27.5	29.8	32.9	29.6	28.9	27.3	25.7	20.8	17.8
	25%	17.9	21.3	24.5	27.8	31.1	34.4	30.8	30.2	29.2	27.5	22.6	18.6
	50%	19.2	23.0	25.6	28.9	31.7	35.2	31.6	31.6	29.8	28.4	23.6	20.2
	0.75	20.6	23.9	26.5	30.2	32.4	35.6	32.2	32.0	31.0	28.8	23.8	20.9
	Max	21.4	24.4	28.6	33.0	33.3	36.0	33.8	32.9	31.6	29.3	27.3	23.1

Min temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	1.4	3.0	6.1	10.2	12.5	17.7	17.7	17.2	16.1	11.1	5.3	2.3
	Std	0.8	1.2	1.4	2.6	0.9	0.6	0.8	0.5	2.3	1.0	1.3	1.1
	Range	2.2	4.2	4.1	9.0	2.6	2.2	2.4	1.5	7.6	2.8	3.6	3.1
	Min	0.5	0.8	3.6	8.1	11.4	16.6	16.5	16.2	14.5	9.5	3.2	0.8
	25%	0.8	2.3	5.4	8.7	11.8	17.3	17.3	17.0	14.7	10.2	4.5	1.3
	50%	1.0	3.2	6.7	9.4	12.5	17.8	17.8	17.3	15.7	11.1	5.4	2.7
	0.75	2.0	3.8	7.0	10.5	13.2	17.9	18.2	17.5	16.2	12.0	6.4	2.8
	Max	2.7	5.0	7.7	17.1	14.0	18.8	18.9	17.7	22.1	12.3	6.8	3.9
Rain	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	14.3	7.1	9.8	5.3	9.4	38.8	132.2	117.7	84.3	26.2	17.1	17.0
	Std	13.0	10.2	16.2	6.8	9.1	25.6	54.9	37.3	31.4	20.3	14.7	16.9
	Range	37.0	29.2	54.6	20.4	32.0	71.8	193.3	114.5	108.4	62.8	40.3	49.4
	Min	1.2	0.3	–	–	1.1	11.6	11.6	76.3	23.9	3.7	0.2	0.2
	25%	3.6	0.6	1.7	0.4	4.4	15.8	107.1	95.6	62.2	9.5	3.1	6.8
	50%	10.3	1.3	4.9	2.3	7.4	33.8	136.1	109.9	93.1	23.9	16.4	11.3
	0.75	19.8	7.9	9.3	6.9	11.2	55.6	173.8	132.8	103.7	33.9	29.9	20.9
	Max	38.2	29.5	54.6	20.4	33.1	83.4	204.9	190.8	132.3	66.5	40.5	49.6

Table 3 General historical statistics monthly variables Sinaloa, own creation with Python 3.6 [50, 105]

Variable	Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Production	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	7,692	6,886	7,108	5,984	7,710	6,382	21,248	19,791	8,798	7,039	5,562	5,414
	Std	3,177	2,203	3,428	2,357	4,167	2,735	9,037	8,669	4,440	3,381	2,563	2,589
	Range	9,520	6,826	11,170	8,267	12,416	9,206	24,664	24,929	14,125	10,902	6,989	7,075
	Min	4,158	3,392	2,485	2,647	2,532	3,053	9,724	7,822	2,663	2,703	2,044	1,820
	25%	4,692	5,584	4,713	4,401	4,046	4,668	14,219	13,857	5,617	4,973	3,635	3,948
	50%	7,929	6,574	6,112	5,629	7,473	6,339	20,654	17,591	8,336	6,314	5,174	5,430
	75%	8,798	8,570	9,677	7,171	10,767	7,375	27,466	26,205	10,797	9,020	7,648	7,353
	Max	13,678	10,218	13,656	10,913	14,948	12,258	34,388	32,752	16,788	13,605	9,033	8,895
Max temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	28.7	29.9	31.9	33.9	35.8	36.7	35.7	35.4	34.6	34.6	32.6	28.8
	Std	0.9	1.4	1.2	1.4	0.8	0.7	0.6	0.6	1.6	2.0	1.4	1.3
	Range	2.8	4.1	3.9	4.8	2.4	2.4	2.3	1.7	6.0	6.8	4.5	4.3
	Min	27.5	27.9	30.2	31.8	34.6	35.3	34.6	34.5	30.4	30.6	30.8	26.0
	25%	28.2	29.0	31.2	33.3	35.4	36.3	35.3	35.0	34.3	34.2	31.7	28.2
	50%	28.5	29.7	31.8	33.8	35.7	36.8	35.6	35.4	34.8	34.7	32.5	28.9
	0.75	29.3	31.1	32.6	34.7	36.3	37.0	35.9	35.7	35.6	35.6	32.7	29.7
	Max	30.3	32.0	34.1	36.6	37.0	37.7	36.9	36.2	36.4	37.4	35.3	30.3

Min temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	10.5	11.0	12.6	15.0	18.1	23.5	23.7	23.8	23.1	20.8	15.9	11.9
	Std	1.5	1.8	1.5	1.1	1.0	0.7	0.6	0.4	1.5	1.1	1.6	1.5
	Range	4.8	6.7	4.6	3.5	2.6	2.2	1.8	1.5	5.4	3.2	5.2	5.4
	Min	8.3	7.8	10.2	13.3	16.9	22.7	22.6	23.0	19.1	18.8	12.4	8.8
	25%	9.5	10.0	11.7	14.2	17.2	23.1	23.5	23.6	23.0	19.9	15.9	11.2
	50%	10.3	11.1	12.7	15.1	18.0	23.3	23.8	23.9	23.5	21.2	16.2	12.3
	0.75	11.5	11.6	13.7	15.3	19.0	23.9	24.2	24.1	23.8	21.8	16.8	12.6
	Max	13.1	14.5	14.8	16.8	19.5	24.9	24.4	24.5	24.5	22.0	17.6	14.2
Rain	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	6.0	12.0	6.2	1.9	2.7	44.4	198.2	224.6	166.6	46.5	23.1	17.3
	Std	9.5	17.7	10.8	4.0	3.7	35.7	37.4	31.0	67.4	48.0	30.4	15.9
	Range	29.4	47.0	35.0	13.1	11.6	101.5	97.8	93.0	234.1	142.7	91.1	40.5
	Min	0.1	–	–	–	–	3.1	153.1	165.5	64.5	5.5	–	–
	25%	0.8	0.2	0.4	0.1	0.3	13.5	169.4	201.3	129.8	17.0	1.4	2.8
	50%	2.0	1.5	1.8	0.4	0.9	44.9	191.4	234.4	169.2	29.5	7.3	16.1
	0.75	4.0	24.1	6.2	1.4	4.3	68.6	234.1	246.5	199.6	47.9	36.8	29.6
	Max	29.5	47.0	35.0	13.1	11.6	104.6	250.9	258.5	298.6	148.2	91.1	40.5

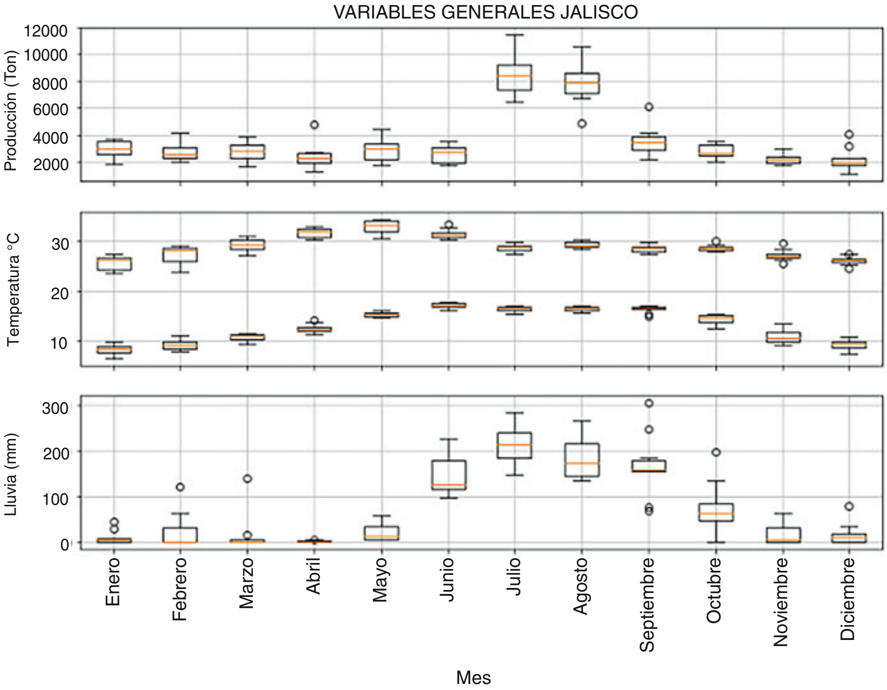


Fig. 20 Historical variations main variables Jalisco, own creation with Python 3.6 based on [50, 105]

of kurtosis, closely followed by the State of Jalisco. Regarding the asymmetry of the distributions, we can see that with the exception of Sinaloa and Tamaulipas, the other states present a symmetrically distributed distribution.

In Table 9, we can see the p-value result for different types of distribution. In the States of Chihuahua, Sinaloa, Michoacán, and Jalisco, it can be seen that the best p-values coincide with the Lognormal distribution, and for the State of Tamaulipas, the one that best fits is a Weibull.

In Table 10, we can see the key parameters to define the distributions previously defined for the States of Chihuahua, Sinaloa, and Michoacán. In Table 11, we can see the parameters for the States of Jalisco and Tamaulipas.

In Table 12, you can see the key parameters for the definition of the constant behavior of the chili harvest in the states analyzed with the information described above, where for the Lognormal distributions the first parameter is the mean, the second is the standard deviation, and the third is the minimum value of the data; for the Weibull distribution, the first parameter is a shape constant, the second parameter is the scale, and the third parameter is the minimum value of the production data [39].

Table 4 General historical statistics monthly variables Michoacán, own creation with Python 3.6 [50, 105]

Variable	Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Production	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	5,121	4,447	4,512	3,773	5,209	4,363	14,110	12,990	5,515	4,668	3,703	3,589
	Std	1,698	1,749	2,273	1,552	2,765	1,967	5,894	5,099	2,434	2,122	1,729	1,978
	Range	5,685	5,897	7,837	4,929	8,752	6,180	17,186	16,683	8,405	7,019	5,682	5,655
	Min	2,362	2,597	2,149	2,263	2,189	2,098	7,854	6,365	2,302	2,337	1,767	1,573
	25%	3,950	3,066	3,147	2,488	3,269	2,690	9,661	9,113	4,111	3,144	2,516	2,135
	50%	5,224	4,620	3,778	3,638	4,765	4,025	12,520	14,618	5,300	3,983	3,273	3,337
	75%	6,160	4,844	5,203	4,486	6,609	5,560	18,410	15,281	6,562	5,882	4,310	4,026
	Max	8,047	8,494	9,986	7,192	10,941	8,278	25,041	23,048	10,707	9,355	7,449	7,228
Max temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	25.7	26.9	28.3	31.5	31.5	29.6	27.2	27.6	27.2	27.8	26.7	25.9
	Std	1.4	2.1	1.2	1.6	1.6	1.0	1.0	1.0	0.9	1.0	1.1	1.0
	Range	4.6	6.2	3.7	4.9	4.8	3.8	2.8	2.9	2.7	2.7	3.1	3.1
	Min	23.1	22.5	26.5	29.3	28.7	28.0	25.9	26.2	26.0	26.3	25.7	24.5
	25%	25.1	26.8	27.5	30.4	30.3	29.2	26.5	26.9	26.6	27.0	25.9	25.2
	50%	25.6	27.7	28.2	31.3	32.0	29.5	27.0	27.4	27.1	27.9	26.2	25.6
	0.75	26.9	28.5	29.4	32.5	32.7	30.1	28.0	28.5	28.0	28.6	26.9	26.7
	Max	27.7	28.7	30.2	34.2	33.5	31.8	28.7	29.1	28.7	29.0	28.8	27.6

(continued)

Table 4 (continued)

Variable	Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Min temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	7.7	8.4	9.8	12.7	14.3	15.3	14.7	14.6	14.7	12.8	9.8	8.0
	Std	1.7	1.4	1.3	2.1	0.9	0.9	1.0	0.7	0.8	1.8	2.2	2.5
	Range	6.2	5.1	4.3	6.3	2.7	2.7	3.2	2.3	2.1	5.9	6.6	8.2
	Min	5.2	6.4	7.9	10.4	12.8	14.1	12.7	13.4	13.7	9.3	6.4	3.5
	25%	7.0	7.5	9.1	11.3	13.5	14.5	14.7	14.1	13.9	11.6	8.0	7.5
	50%	7.4	8.5	9.6	12.4	14.5	15.4	15.0	14.4	14.6	13.4	9.9	8.0
	0.75	8.5	8.9	10.8	13.1	14.7	16.1	15.3	15.2	15.4	14.2	11.5	9.6
	Max	11.4	11.5	12.2	16.7	15.5	16.8	15.9	15.7	15.8	15.2	13.0	11.7
Rain	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	9.3	26.2	22.2	4.1	36.4	137.3	202.1	175.5	168.2	59.0	20.9	8.1
	Std	13.6	56.8	48.5	3.4	29.2	26.0	64.5	39.6	79.3	43.8	24.7	10.4
	Range	43.0	179.1	158.7	10.3	84.4	90.9	190.4	114.5	282.5	161.7	77.4	32.1
	Min	0.1	–	–	0.9	10.3	112.9	139.7	118.2	76.9	2.9	–	–
	25%	0.7	0.6	1.2	1.7	15.3	124.2	154.8	148.7	135.9	40.2	0.8	0.4
	50%	3.2	1.1	6.5	2.9	26.0	131.9	175.0	166.8	155.7	47.1	16.8	5.4
	0.75	14.4	14.0	13.6	5.7	49.8	140.5	240.8	206.8	186.2	71.0	28.7	10.6
	Max	43.1	179.1	158.7	11.2	94.7	203.8	330.1	232.7	359.4	164.6	77.4	32.1

Table 5 General historical statistics monthly variables Jalisco, own creation with Python 3.6 [50, 105]

Variable	Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Production	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	2,956	2,699	2,762	2,404	2,897	2,584	8,488	7,930	3,498	2,754	2,207	2,157
	Std	637	687	685	920	838	707	1,627	1,649	1,113	556	407	864
	Range	1,920	2,159	2,266	3,463	2,709	1,810	4,967	5,728	3,979	1,554	1,247	3,020
	Min	1,811	1,975	1,650	1,258	1,695	1,743	6,488	4,880	2,134	1,969	1,724	1,069
	25%	2,554	2,232	2,276	1,917	2,198	1,894	7,360	7,104	2,894	2,422	1,896	1,698
	50%	2,963	2,488	2,776	2,227	3,018	2,677	8,454	7,850	3,410	2,607	2,134	1,894
	75%	3,528	3,078	3,256	2,598	3,358	3,082	9,272	8,575	3,834	3,281	2,371	2,222
	Max	3,731	4,135	3,916	4,721	4,404	3,552	11,455	10,608	6,112	3,523	2,971	4,088
Max temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	25.6	27.1	29.3	31.7	32.9	31.4	28.7	29.2	28.5	28.6	27.1	26.2
	Std	1.4	2.0	1.3	1.0	1.5	1.0	0.8	0.6	0.8	0.7	1.1	0.9
	Range	3.9	5.2	3.8	2.6	4.0	3.2	2.4	1.8	2.4	2.3	4.2	3.0
	Min	23.5	23.8	27.2	30.4	30.5	30.2	27.3	28.4	27.3	27.8	25.4	24.5
	25%	24.3	25.9	28.5	30.7	32.0	30.7	28.2	28.8	27.9	28.2	26.7	25.8
	50%	26.2	28.0	29.3	32.0	33.3	31.3	28.9	29.1	28.7	28.5	27.0	26.3
	0.75	26.6	28.6	30.3	32.5	34.2	31.8	29.2	29.7	28.9	29.0	27.4	26.5
	Max	27.4	29.0	31.0	33.0	34.5	33.4	29.7	30.2	29.7	30.1	29.6	27.5

(continued)

Table 5 (continued)

Variable	Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Min temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	8.3	9.1	10.6	12.4	15.3	17.0	16.4	16.4	16.3	14.2	10.9	9.1
	Std	1.1	1.0	0.7	0.9	0.5	0.6	0.5	0.5	0.7	1.0	1.5	1.0
	Range	3.4	3.2	2.2	2.9	1.5	1.7	1.7	1.6	2.3	3.0	4.3	3.4
	Min	6.4	7.8	9.3	11.2	14.5	16.1	15.4	15.5	14.8	12.4	9.1	7.4
	25%	7.6	8.4	10.2	11.9	14.9	16.7	16.0	16.1	16.2	13.6	9.8	8.5
	50%	8.4	9.1	10.9	12.3	15.3	17.0	16.6	16.5	16.5	14.6	10.5	9.2
	0.75	8.9	9.7	11.1	12.7	15.5	17.6	16.8	16.7	16.8	15.0	11.7	9.7
	Max	9.8	11.0	11.5	14.1	16.0	17.8	17.1	17.1	17.1	15.4	13.4	10.8
Rain	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	10.2	23.7	16.9	2.0	21.3	146.1	211.5	183.8	167.9	74.9	19.0	17.4
	Std	15.2	41.1	43.6	2.4	19.2	43.0	41.9	44.9	69.9	56.0	25.0	24.2
	Range	46.0	122.3	140.2	7.2	53.9	128.8	135.4	131.2	235.7	195.3	63.2	79.0
	Min	–	–	–	0.1	5.0	97.8	147.4	135.6	69.1	1.2	–	–
	25%	0.6	0.0	0.4	0.3	6.9	116.8	185.5	145.6	155.5	47.4	0.6	1.4
	50%	5.3	1.9	1.4	1.2	13.8	127.4	213.7	174.0	158.7	63.3	6.5	10.7
	0.75	7.9	32.9	5.4	2.4	34.9	180.3	238.4	215.7	179.9	85.5	31.3	18.0
	Max	46.0	122.3	140.2	7.3	58.9	226.6	282.8	266.8	304.8	196.5	63.2	79.0

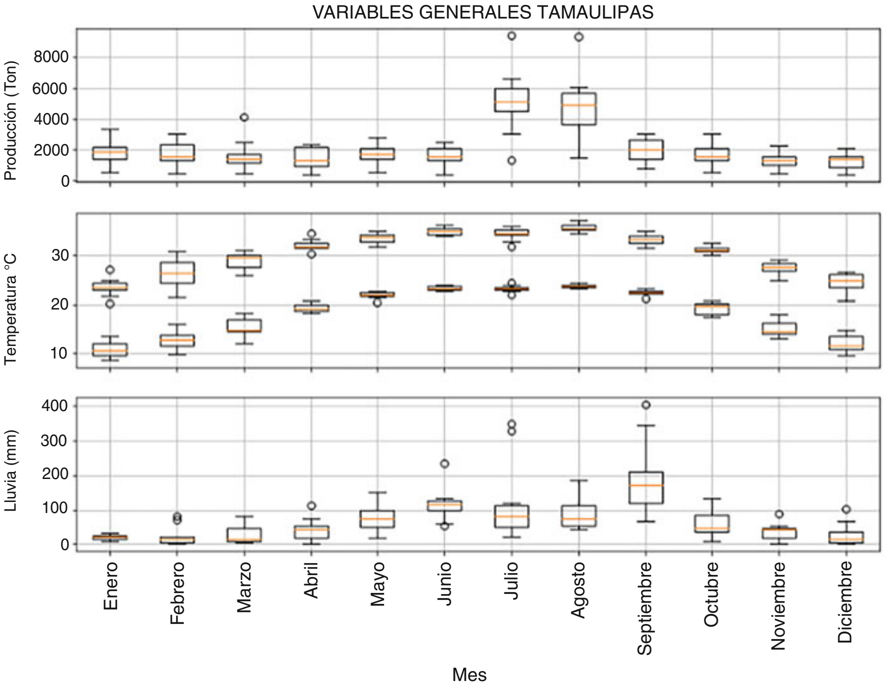


Fig. 21 Historical variations main variables Tamaulipas, own creation with Python 3.6 based on [50, 105]

Annual Analysis Seasonal Behavior Pepper Harvest In the months of July–August, you can see the chili production peaks, where compared to the other months there is a very significant variation; therefore, the probable distributions for each state were evaluated and thus define the behavior that will be used for the simulation. In Table 13, you can see the summary of the general parameters of the states for these months; we can see that the production of Chihuahua for these months is more than triple that of Sinaloa, which has the second highest value of production, and in addition, in the kurtosis values, we can see that Tamaulipas has a lower variation in relation to the others where its values are less than 0. The State of Michoacán presents an asymmetry to the right followed by Tamaulipas and the other states show good symmetry.

In Table 14, we can see the p-value result for different types of distribution. In the States of Chihuahua and Tamaulipas, we can see that the distribution that best fits is Normal since its p-values are the highest; for the States of Sinaloa and Jalisco, a Lognormal distribution is better adjusted; and finally, in the State of Michoacán, a Gamma distribution is better applied. For simplicity purposes, we could define a Normal distribution for all the states evaluated since its p-value is greater than 0.05, and statistically it would not be a bad approximation, but since we seek the greatest adherence to the behavior of the data, we will work with the largest values.

Table 6 General historical statistics monthly variables Tamaulipas, own creation with Python 3.6 [50, 105]

Variable	Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Production	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	1,888	1,724	1,676	1,488	1,721	1,585	5,197	4,804	2,007	1,662	1,318	1,264
	Std	871	775	990	708	653	633	2,164	2,137	740	700	504	516
	Range	2,795	2,625	3,616	1,982	2,233	2,124	8,136	7,848	2,261	2,541	1,834	1,733
	Min	523	442	485	373	567	346	1,295	1,480	777	495	418	349
	25%	1,424	1,343	1,175	943	1,361	1,349	4,547	3,648	1,417	1,286	1,017	880
	50%	1,857	1,575	1,405	1,328	1,677	1,557	5,155	4,902	2,033	1,513	1,302	1,364
	75%	2,189	2,314	1,718	2,152	2,081	2,074	6,030	5,723	2,616	2,067	1,569	1,572
	Max	3,318	3,067	4,101	2,355	2,800	2,470	9,431	9,328	3,038	3,036	2,252	2,082
Max temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	23.5	26.2	28.8	32.1	33.4	34.9	34.3	35.6	33.1	31.1	27.3	24.2
	Std	1.8	2.8	1.6	1.1	1.1	0.8	1.3	0.8	1.1	0.7	1.3	2.2
	Range	6.8	9.3	5.0	4.3	3.3	2.4	4.1	2.9	3.6	2.5	4.1	6.1
	Min	20.2	21.3	25.9	30.2	31.6	33.8	31.8	34.3	31.4	30.0	24.9	20.5
	25%	22.9	24.4	27.5	31.5	32.7	34.1	34.1	35.2	32.4	30.7	26.8	23.3
	50%	23.3	26.3	29.5	31.8	33.7	34.9	34.4	35.5	33.2	31.1	27.6	24.7
	0.75	24.3	28.4	30.0	32.4	34.2	35.5	35.1	36.1	33.8	31.4	28.2	26.0
	Max	27.0	30.6	30.9	34.5	34.9	36.2	35.9	37.2	35.0	32.5	29.0	26.6

Min temperature	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	10.8	12.9	15.4	19.3	21.8	23.2	23.1	23.6	22.3	19.2	15.1	12.1
	Std	1.6	2.0	1.8	0.8	0.7	0.4	0.6	0.4	0.6	1.3	1.7	1.7
	Range	5	6.1	6.0	2.3	2.2	1.0	2.3	1.4	2.1	3.2	4.9	5.0
	Min	8.6	9.8	12.1	18.2	20.3	22.7	21.9	23.0	21.1	17.5	12.9	9.6
	25%	9.6	11.5	14.6	18.7	21.5	22.9	22.9	23.4	22.1	17.9	14.0	10.9
	50%	10.6	12.8	14.8	19.2	22.0	23.2	23.2	23.6	22.3	19.7	14.5	11.6
	0.75	12.0	13.8	16.9	20.0	22.3	23.6	23.4	23.8	22.7	20.2	16.3	13.5
	Max	13.6	15.9	18.1	20.5	22.5	23.7	24.2	24.4	23.2	20.7	17.8	14.6
Rain	Count	10	10	10	10	10	10	10	10	10	10	10	10
	Mean	21.1	24.1	28.3	42.6	79.3	115.9	122.8	87.4	190.3	60.7	36.7	27.7
	Std	9.4	28.4	27.4	33.7	39.9	49.3	116.5	45.3	110.0	39.8	25.2	33.3
	Range	27.1	79.6	77.8	111.6	131.3	180.1	324.6	139.6	336.9	123.0	88.5	100.8
	Min	6.6	0.3	3.3	0.3	20.1	54.5	22.4	44.5	65.9	9.9	0.1	0.1
	25%	13.6	5.1	6.9	18.8	51.5	97.5	50.2	54.8	118.6	36.3	18.6	3.6
	50%	23.4	15.1	14.3	43.4	73.6	114.2	79.5	72.6	171.2	46.3	41.2	16.8
	0.75	26.5	22.0	47.1	52.3	99.2	126.2	113.0	112.9	210.8	85.7	45.5	37.1
	Max	33.7	79.9	81.1	111.9	151.4	234.6	347.0	184.1	402.8	132.9	88.6	100.9

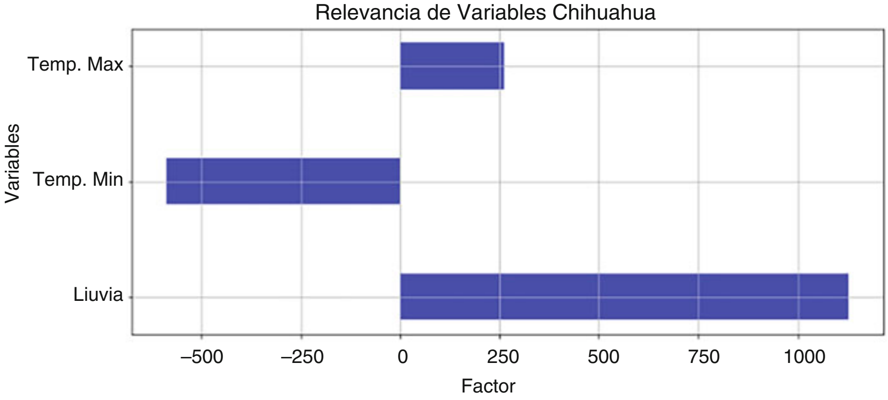


Fig. 22 Relevance of Chihuahua main variables by linear regression, own creation with Python 3.6 based on [50, 105]

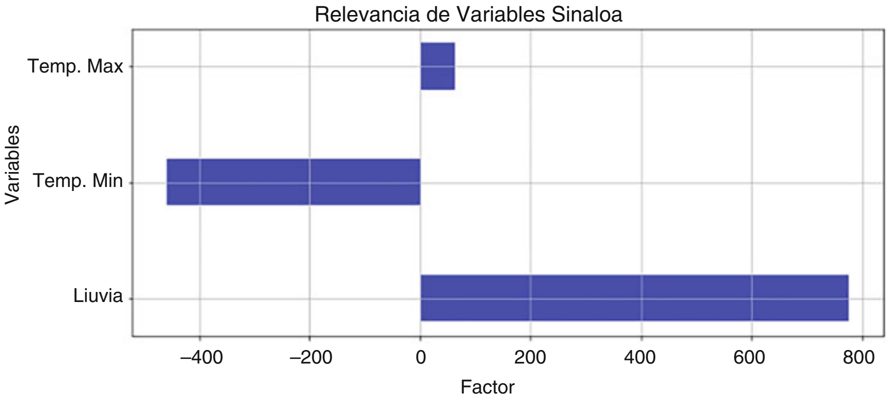


Fig. 23 Relevance of Sinaloa main variables by linear regression, own creation with Python 3.6 based on [50, 105]

In Table 15, we can see the key parameters to define the distributions previously defined for the States of Chihuahua, Sinaloa, and Michoacán. These parameters are key to defining the distributions, since in the cases of Normal distributions they are only the mean and the standard deviation, for the Lognormal distribution, apart from the mean and the standard deviation, the minimum value of the observed data is required, and finally for the gamma distribution these parameters are called alpha and beta which correspond to the shape and scale parameter as well as the minimum value.

In Table 16, we can see the parameters for the States of Jalisco and Tamaulipas.

In Table 17, you can see the key parameters for the definition of the stable behavior of the chili harvest in the states analyzed for the months of July–August with the information described above, where for the Lognormal distribution the first

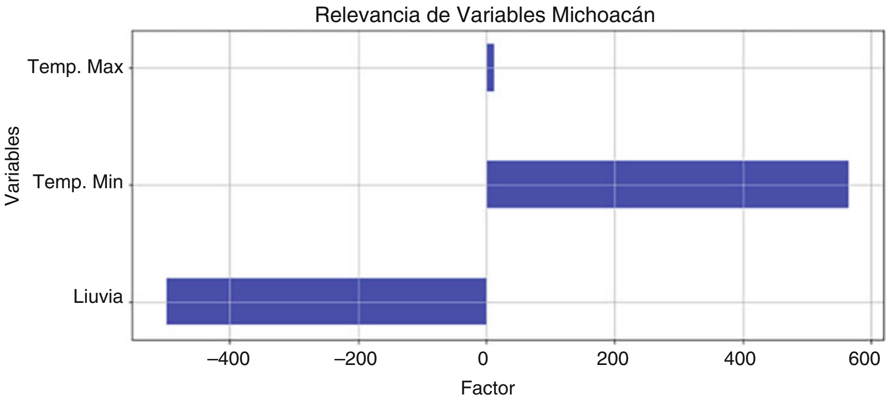


Fig. 24 Relevance of Michoacán main variables by linear regression, own creation with Python 3.6 based on [50, 105]

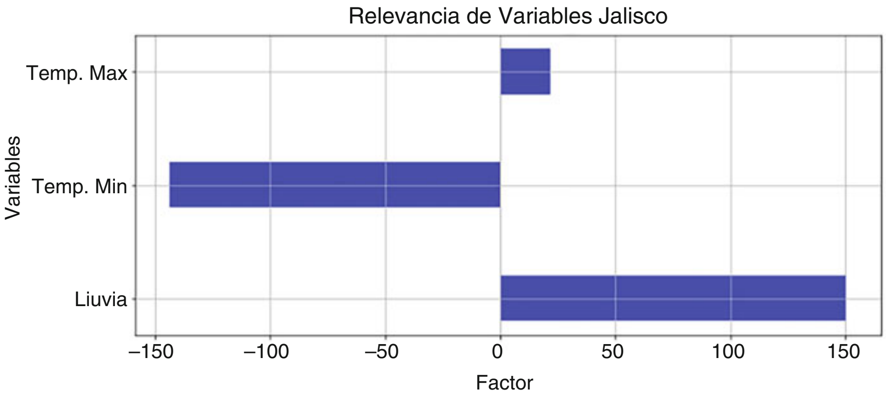


Fig. 25 Relevance of Jalisco main variables by linear regression, own creation with Python 3.6 based on [50, 105]

parameter is the mean, the second is the standard deviation, and the third is the minimum value of the data; for the gamma distribution, the first parameter is a shape constant (alpha), the second parameter is the scale (beta), and the third parameter is the minimum value of the production data; and finally for normal distributions, the first parameter is the mean and the second the standard deviation [39].

Annual Analysis Behavior of Damaged Pepper Harvest Damaged harvests are the quantities of chili that are damaged due to various effects such as frost, floods, and pests, among other factors that directly affect the yield of the quantities of chili planted. In Fig. 27, we can see the historical amounts of jalapeño pepper sinister for the States of Chihuahua, Sinaloa, Michoacán, Jalisco, and Tamaulipas; it can be seen that there is no clear pattern of the behavior of the annual amounts lost in pepper, only evaluating the annual productions.

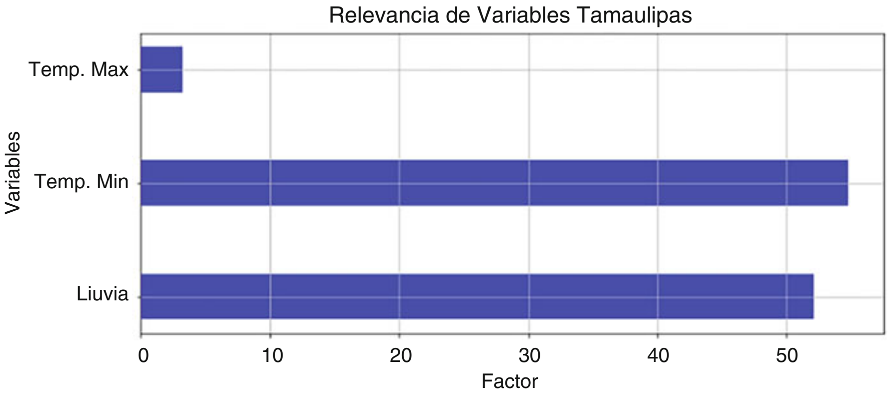


Fig. 26 Relevance of Tamaulipas main variables by linear regression, own creation with Python 3.6 based on [50, 105]

Table 7 Summary statistics by state and linear regression adjustment, own creation with Python 3.6 [50, 105]

Estate	Variable	Importance	R2
Chihuahua	Max temperature	262.70	0.51
	Min temperature	−589.10	
	Rain	1129.42	
Sinaloa	Max temperature	63.25	0.47
	Min temperature	−459.30	
	Rain	777.32	
Michoacán	Max temperature	11.79	0.31
	Min temperature	565.61	
	Rain	−498.97	
Jalisco	Max temperature	21.78	0.45
	Min temperature	−143.85	
	Rain	150.57	
Tamaulipas	Max temperature	3.25	0.14
	Min temperature	54.77	
	Rain	52.10	

Table 8 Summary statistics constants by state for evaluation of distributions, own creation with Minitab 18 [50]

Estate	N	Mean	Std	Median	Min	Max	Asymmetry	Kurtosis
Chihuahua	100.00	21,691	7,115	20,352	6,216	51,667	1.01	2.36
Sinaloa	100.00	6,858	3,200	6,398	1,820	16,788	0.78	0.34
Michoacán	100.00	4,490	2,062	4,039	1,573	10,941	0.99	0.79
Jalisco	100.00	2,692	820	2,517	1,069	6,112	0.96	1.94
Tamaulipas	100.00	1,633	724	1,519	346	4,101	0.58	0.56

Table 9 Summarizes constant p-value by state for evaluation of distributions, own creation with Minitab 18 [50]

	Chihuahua			Sinaloa			Michoacán			Jalisco			Tamaulipas		
Distribution	AD	P	LRT P	AD	P	LRT P	AD	P	LRT P	AD	P	LRT P	AD	P	LRT P
Normal	0.94	0.017		1.15	0.005		1.825	<0.005		1.133	0.006		0.664	0.081	
Box-Cox transformation	0.297	0.584		0.276	0.651		0.428	0.306		0.469	0.244		0.441	0.284	
Lognormal	0.224	**0.819**		0.37	**0.42**		0.428	**0.306**		0.469	**0.244**		1.732	<0.005	
3-parameter Lognormal	0.163	*	0.182	0.238	*	0.27	0.45	*	0.373	0.479	*	0.92	0.347	*	0
Exponential	22.161	<0.003		13.462	<0.003		14.649	<0.003		23.444	<0.003		14.366	<0.003	
2-parameter exponential	14.652	<0.010	0	5.452	<0.010	0	3.555	<0.010	0	12.023	<0.010	0	8.462	<0.010	0
Weibull	1.207	<0.010		0.463	>0.250		1.067	<0.010		1.338	<0.010		0.446	**>0.250**	
3-parameter Weibull	0.655	0.077	0.01	0.168	>0.500	0.01	0.241	>0.500	0	0.599	0.113	0	0.498	0.216	0.383
Smallest extreme value	5.269	<0.010		3.907	<0.010		4.881	<0.010		4.749	<0.010		3.285	<0.010	
Largest extreme value	0.228	>0.250		0.266	>0.250		0.65	0.089		0.48	0.236		0.635	0.095	
Gamma	0.204	>0.250		0.194	>0.250		0.576	0.154		0.558	0.169		0.73	0.062	
3-parameter gamma	0.205	*	0.99	0.188	*	0.432	0.304	*	0.015	0.464	*	0.318	0.521	*	0.236
Logistic	0.577	0.092		0.884	0.012		1.407	<0.005		1.102	<0.005		0.537	0.127	
Loglogistic	0.161	>0.250		0.441	0.233		0.627	0.066		0.71	0.038		0.915	0.009	
3-parameter loglogistic	0.161	*	0.572	0.395	*	0.647	0.662	*	0.22	0.688	*	0.698	0.323	*	0.012
Johnson transformation	0.123	0.987		0.159	0.949		0.22	0.83		0.432	0.3		0.313	0.544	

*that indicate that the software don't get a specific value for that data
Bold values indicate the best parameter for each prediction

Table 10 Summarizes parameters Chihuahua, Sinaloa and Michoacán constant for allocation of distributions, own creation with Minitab 18 [50]

	Chihuahua				Sinaloa				Michoacán			
Distribution	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold
Normal[a]	21,690.770		7,114.860		6857.5037		3200.24469		4489.9683		2061.82788	
Box-Cox transformation[a]	145.393		23.607		80.5963		19.11528		8.31009		0.44899	
Lognormal[a]	9.932		0.329		8.72129		0.48913		8.31009		0.44899	
3-parameter Lognormal	10.236		0.240	−7,015.022	8.96752		0.37814	-1559.12329	8.11693		0.53979	633.70747
Exponential			21,690.770				6857.50216				4489.96725	
2-parameter exponential			15,631.024	6,059.730			5088.7663	1768.73234			2946.33876	1543.62661
Weibull		3.135	24,162.055			2.29981	7764.00021			2.33363	5085.58639	
3-parameter Weibull		2.409	18,448.446	5,304.579		1.69708	5922.4859	1566.52786		1.46833	3292.3395	1508.2348
Smallest extreme value	25,515.099		9,174.617		8554.02212		3627.26146		5599.96457		2435.74468	
Largest extreme value	18,444.658		5,878.891		5386.14843		2530.69829		3564.91957		1534.45196	
Gamma		9.750	2,224.765			4.63197	1480.47193			5.18548	865.87315	
3-parameter gamma		9.809	2,217.542	−61.151		3.56961	1722.36602	709.26487		2.50548	1309.20118	1209.73463
Logistic	21,176.305		3,858.251		6603.56237		1799.41442		4286.88735		1146.07485	
Loglogistic	9.938		0.182		8.73857		0.28163		8.30847		0.26446	
3-parameter loglogistic	10.111		0.153	−3,833.356	8.85956		0.24781	-761.00084	8.00471		0.36219	985.28326
Johnson transformation[a]	−0.001		1.028		0.03697		1.00371		0.04667		0.90682	

[a]Scale: Adjusted ML estimate

Table 11 Summarizes Jalisco and Tamaulipas constant for allocation of distributions own creation with Minitab 18 [50]

	Jalisco				Tamaulipas			
Distribution	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold
Normal[a]	2691.9022		820.0408		1633.1433		723.67017	
Box-Cox transformation[a]	7.85398		0.29845		39.35899		9.21204	
Lognormal[a]	7.85398		0.29845		7.28465		0.51423	
3-parameter Lognormal	7.88034		0.28921	-65.91645	8.17449		0.19744	-1985.89157
Exponential			2691.9022				1633.1433	
2-parameter exponential			1639.57633	1052.32424			1300.24445	332.89756
Weibull		3.36702	2987.3841			2.40782	1842.32643	
3-parameter Weibull		2.19138	1925.73266	985.88685		2.14848	1668.42919	153.21767
Smallest extreme value	3130.32572		1044.34704		2010.38785		819.2534	
Largest extreme value	2317.09508		659.88932		1290.93758		624.75705	
Gamma		11.52152	233.64117			4.56109	358.06006	
3-parameter gamma		6.7032	311.58387	603.29258		6.1853	298.95522	-216.00426
Logistic	2634.89288		457.54564		1595.47615		404.82552	
Loglogistic	7.85321		0.17219		7.32898		0.27673	
3-parameter loglogistic	7.72199		0.19662	308.33467	8.09796		0.1215	-1725.37695
Johnson transformation[a]	0.01163		1.05082		-0.00378		1.01225	

[a]Scale: Adjusted ML estimate

Table 12 Summary of constant behavior statistics by state (own creation) [50]

Estate	Behavior
Chihuahua	Lognormal (9.93248, 0.32915, 6216.04)
Sinaloa	Lognormal (8.72129, 0.48913, 1819.62)
Michoacán	Lognormal (8.31009, 0.44899, 1573.09)
Jalisco	Lognormal (7.85398, 0.29845, 1068.72)
Tamaulipas	Weibull (2.40782, 1842.32643, 345.9)

Table 13 Summary of seasonal statistics by state for evaluation of distributions, own creation with Minitab 18 [50]

Estate	N	Mean	Std	Median	Min	Max	Asymmetry	Kurtosis
Chihuahua	20	64544.2	15065.5	61493.6	37735.1	89672	0.0112875	−0.525387
Sinaloa	20	20519.5	8651.41	17818	7822.25	34387.7	0.270512	−1.33734
Michoacán	20	13549.8	5394.66	14610	6365.33	25040.5	0.59072	−0.421737
Jalisco	20	8209.06	1619.88	8109.28	4880.47	11455.2	0.202934	−0.0405646
Tamaulipas	20	5000.49	2103.06	4983.15	1295.26	9431.13	0.372593	0.646392

* that indicate that the software don't get a especific value for that data

In Fig. 28, the analysis of the relevance of the variables of maximum temperature, minimum temperature, and precipitation with respect to the amounts of damaged pepper where we can see that when temperatures tend to be very low, there is a high probability that high quantities of jalapeño pepper; the rain factor can also be seen when there is little precipitation they tend to lose a lot of harvest and the maximum temperature does not significantly affect the harvest.

For the analysis of damaged chili in the State of Chihuahua, in Fig. 29, we can see the quantities of chili that have been damaged between the years 2008 and 2017, where in the last 4 years there have been no damaged amounts compared to the previous years. In Fig. 30, in the State of Sinaloa, there is no significant trend, since in 3 years there was a behavior of almost zero amounts claimed, but in 2016 the highest occurred in the last 10 years. In Fig. 31, for the State of Michoacán, in general, there are no significant amounts claimed, but in 2014 the amounts lost were very relevant. In Fig. 32, we can see the State of Jalisco, and there is a behavior where for a period of 3 years the amounts lost are low, but when many losses occur there is a tendency to increase over time. Finally, in Fig. 33, in the State of Tamaulipas, a downward trend can be seen over the years with respect to the amounts claimed and with a frequency of 5 years.

With the data previously evaluated, in Table 18, we can see the summary of statistics for each state, where we can see that the asymmetry is very loaded toward the minimum value in all the data of each state, and in the States of Chihuahua and Michoacán, the coefficient of kurtosis tells us about the amplitude of the distribution, due to the existence of many years without events of harvest loss, but with some very significant events.

In Table 19, we can see the distribution evaluation summary, where we can see that in all the states the behavior is a 3-Parameter Weibull distribution since,

Table 14 Summarizes Seasonal p-value by state for evaluation of distributions own creation with Minitab 18 [50]

	Chihuahua			Sinaloa			Michoacán			Jalisco			Tamaulipas		
Distribution	AD	P	LRT P	AD	P	LRT P	AD	P	LRT P	AD	P	LRT P	AD	P	LRT P
Normal	0.343	**0.455**		0.556	0.132		0.584	0.113		0.242	0.737		0.42	**0.295**	
Box-Cox transformation	0.343	0.455		0.441	0.261		0.471	0.219		0.2	0.865		0.466	0.226	
Lognormal	0.415	0.304		0.403	**0.325**		0.471	0.219		0.203	**0.856**		0.844	0.024	
3-parameter Lognormal	0.359	*	0.311	0.449	*	0.852	0.51	*	0.715	0.208	*	0.508	0.384	*	0.053
Exponential	5.537	<0.003		3.137	<0.003		3.592	<0.003		6.078	<0.003		3.317	<0.003	
2-parameter exponential	2.005	<0.010	0	0.954	0.076	0	0.741	0.149	0	2.718	<0.010	0	2.122	<0.010	0.002
Weibull	0.395	>0.250		0.501	0.205		0.517	0.189		0.361	>0.250		0.44	>0.250	
3-parameter Weibull	0.367	0.394	0.505	0.404	0.381	0.226	0.476	0.246	0.055	0.241	>0.500	0.306	0.447	0.251	0.908
Smallest extreme value	0.577	0.131		0.752	0.044		0.89	0.02		0.605	0.105		0.976	0.012	
Largest extreme value	0.489	0.216		0.467	0.237		0.535	0.172		0.286	>0.250		0.537	0.17	
Gamma	0.374	>0.250		0.454	>0.250		0.501	**0.223**		0.197	>0.250		0.571	0.159	
3-parameter gamma	0.668	*	1	0.511	*	1	0.486	*	0.534	2.033	*	1	0.638	*	0.565
Logistic	0.359	>0.250		0.591	0.082		0.598	0.078		0.229	>0.250		0.306	>0.250	
Loglogistic	0.367	>0.250		0.454	0.213		0.545	0.112		0.171	>0.250		0.575	0.09	
3-parameter Loglogistic	0.347	*	0.504	0.445	*	0.853	0.552	*	0.604	0.18	*	0.78	0.3	*	0.113

*Scale: Adjusted ML estimate
Bold values indicate the best parameter for each prediction

Table 15 Summarizes Chihuahua, Sinaloa, and Michoacán seasonal parameters for allocation of distributions, own creation with Minitab 18 [50]

	Chihuahua				Sinaloa				Michoacán			
Distribution	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold
Normal[a]	64,544.243		15,065.514		20,519.482		8651.40879		13,549.81		5394.65542	
Box-Cox transformation[a]	64,544.243		15,065.514		140.09666		30.64924		9.4381		0.40343	
Lognormal[a]	11.048		0.246		9.83749		0.4504		9.4381		0.40343	
3-parameter Lognormal	14.714		0.006	−2.39E+06	9.98229		0.37948	−2.68E+03	9.17911		0.50776	2.59E+03
Exponential			64,544.243				20,519.48002				13,549.80849	
2-parameter exponential			28,219.942	3.63E+04			13,365.42374	7.15E+03			7562.56321	5.99E+03
Weibull		4.941	70,378.753			2.67807	23,181.439			2.7871	15,265.38859	
3-parameter Weibull		3.072	45,560.149	2.39E+04		1.5888	15,118.77761	6.91E+03		1.28426	7918.5281	6.17E+03
Smallest extreme value	71,893.127		13,749.225		24,814.86096		7968.83779		16,301.81603		5516.51643	
Largest extreme value	57,222.483		13,906.636		16,440.23625		7056.03814		11,074.49565		4209.11836	
Gamma		18.274	3532.051			5.6174	3652.84443			6.73857	2010.78478	
3-parameter gamma		507.645	652.267	−2.69E+05		15.20946	2163.18929	−1.24E+04		4.74514	2427.64703	2.03E+03
Logistic	64,317.416		8564.611		20,142.2846		5166.84313		13,206.56435		3086.00308	
Loglogistic	11.059		0.136		9.84839		0.26636		9.44173		0.23771	
3-parameter Loglogistic	12.556		0.030	−2.20E+05	9.70712		0.30811	2.32E+03	9.0222		0.36676	4.00E+03

[a]Scale: Adjusted ML estimate

Table 16 Summarizes Jalisco and Tamaulipas seasonal parameters for allocation of distributions, own creation with Minitab 18 [50]

	Jalisco				Tamaulipas			
Distribution	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold
Normal[a]	8209.063		1619.882		5000.49		2103.058	
Box-Cox transformation[a]	90.17874		8.99458		69.05082		15.64318	
Lognormal[a]	8.99395		0.20261		8.41351		0.50751	
3-parameter Lognormal	9.95787		0.0745	−1.30E+04	9.81707		0.11069	−1.35E+04
Exponential			8209.063				5000.49	
2-parameter exponential			3503.76	4.71E+03			3900.218	1.10E+03
Weibull		5.62499	8864.547			2.60618	5623.567	
3-parameter Weibull		3.05762	4966.092	3.77E+03		2.50307	5429.441	1.76E+02
Smallest extreme value	9013.6935		1570.823		6061.2579		2176.824	
Largest extreme value	7434.49072		1489.151		4003.5204		1885.105	
Gamma		26.42605	310.6428			4.97855	1004.406	
3-parameter gamma		2408.04952	32.1414	−6.99E+04		494.27051	91.85358	−4.08E+04
Logistic	8146.80195		903.7426		4949.5814		1124.522	
Loglogistic	8.99711		0.11125		8.47167		0.25973	
3-parameter Loglogistic	9.40849		0.07353	−4.09E+03	9.99227		0.05142	−1.69E+04

[a]Scale: Adjusted ML estimate

Table 17 Summary of seasonal behavior statistics by state, own creation

Estate	Behavior
Chihuahua	Normal (64,544.24, 15,065.51)
Sinaloa	Lognormal (9.83749, 0.4504, 7822.25)
Michoacán	Gama (6.73857, 2010.78478, 6365.33)
Jalisco	Lognormal (8.99395, 0.20261, 4880.47)
Tamaulipas	Normal (5000.49, 2103.06)

Fig. 27 History of Chile affected in the states of Chihuahua, Sinaloa, Michoacán, Jalisco, and Tamaulipas, own creation with Python 3.6 based on [50, 105]

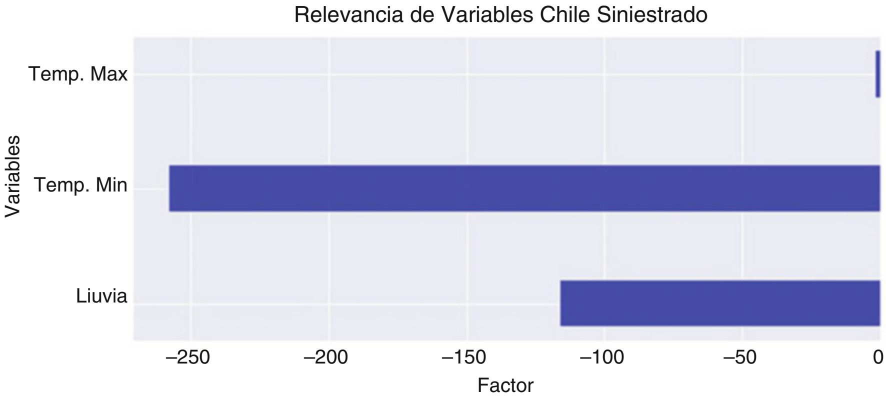


Fig. 28 Relevance of main variables by affected harvest by linear regression, own creation with Python 3.6 based on [50, 105]

Fig. 29 Historical quantities of damaged harvest, state of Chihuahua, own creation with Python 3.6 based on [50]

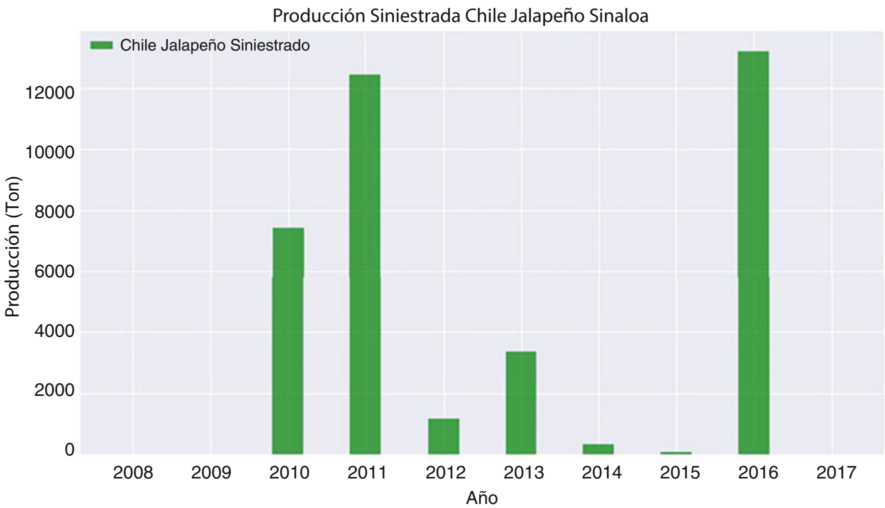


Fig. 30 Historic amounts of damaged harvest, state of Sinaloa, own creation with Python 3.6 based on [50]

according to the previously evaluated data, the greatest possibilities were in values equal to or close to zero. In the case of Tamaulipas, the p-value is less than 0.05, which would imply not assuming this type of distribution, but for analysis purposes, since it is the best value of all, this distribution will be considered.

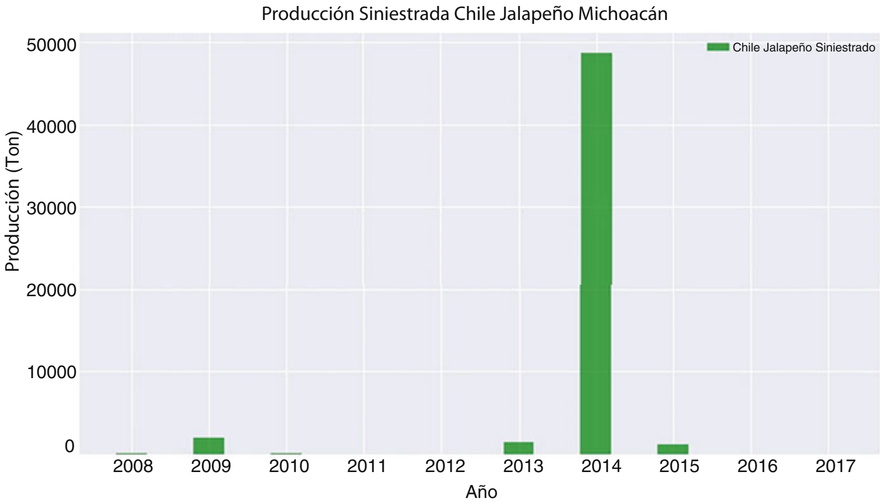


Fig. 31 Historic amounts of damaged harvest, state of Michoacán, own creation with Python 3.6 based on [50]

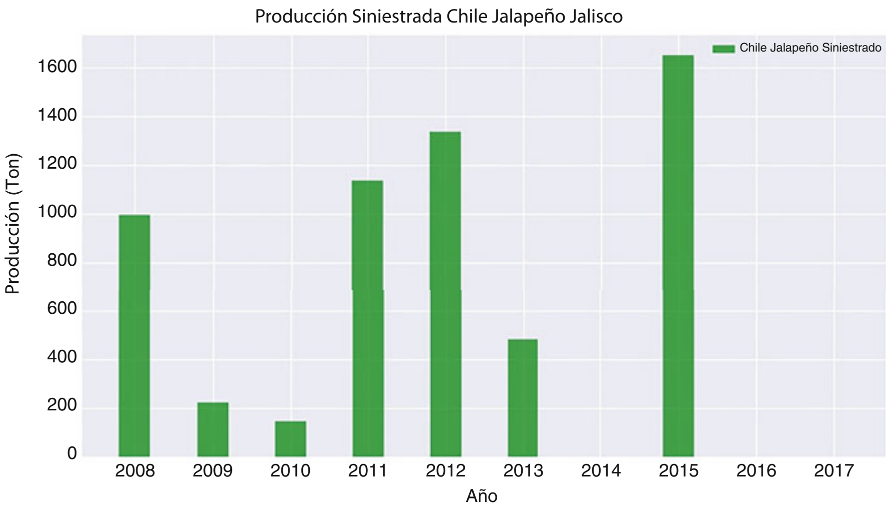


Fig. 32 Historic amounts of damaged harvest, state of Jalisco, own creation with Python 3.6 based on [50]

In Tables 20, 21 and 22, we can observe the parameters for each type of distribution evaluated where the 3 parameters of the Weibull distributions resulting from the test will be taken.

With the previously generated data, Table 23 shows the distributions that represent the behavior of the amounts claimed by each state for the development

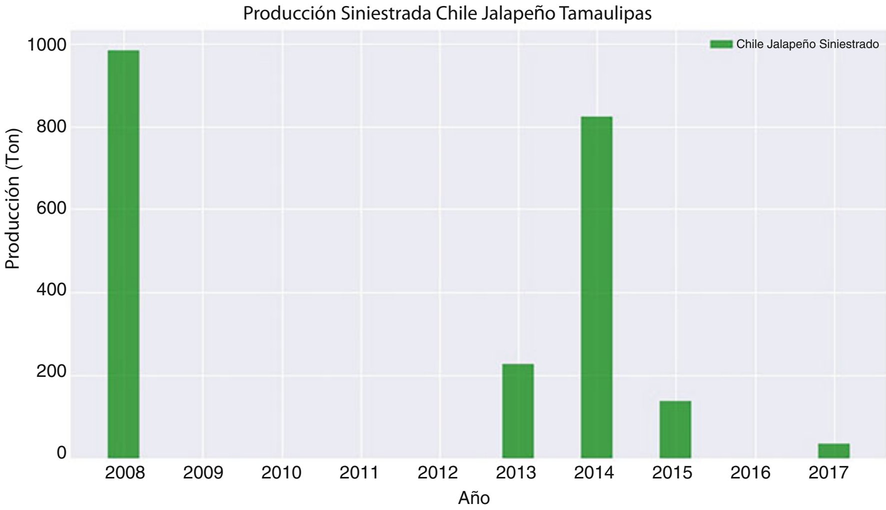


Fig. 33 Historic amounts of damaged harvest, state of Tamaulipas, own creation with Python 3.6 based on [50]

Table 18 Summary statistics behaviors claimed by state, own creation with Minitab 18 [50]

Estate	N	Mean	Std	Median	Min	Max	Asymmetry	Kurtosis
Chihuahua	10	4879.18	9539.29	1127.84	0	30,970.8	2.75122	7.94905
Sinaloa	10	3794.77	5288.25	749.695	0	13,170.3	1.16549	-0.302163
Michoacán	10	5342.53	15,222	127.87	0	48,615	3.14875	9.93642
Jalisco	10	3794.77	5288.25	749.695	0	13,170.3	1.16549	-0.302163
Tamaulipas	10	220.465	369.275	17.65	0	982.26	1.65497	1.31063

of the simulation. Where the first parameter represents the shape of the distribution, the second represents the scale according to the probabilities of the data, and finally there is the threshold that is the tendency to the minimum of the data.

4 Discussion

The analysis of results will be seen in 3 stages: the first stage refers to the simulation of the production of chili where it can be seen in Fig. 34, the response of the AnyLogic software to the simulation of the production of jalapeño pepper according to the data evaluated, where the upper graph represents the amounts of simulated annual chili, and the lower graph shows the behavior for each month.

In Fig. 35, we can see the location of each of the states evaluated, where the green bars correspond to the production of chili for each state and the red bars the amount damaged in each state.

Table 19 Summarizes seasonal p-value by state for evaluation of distributions, own creation with Minitab 18 [50]

	Chihuahua		Sinaloa		Michoacán		Jalisco		Tamaulipas	
Distribution	AD	P	AD	P	AD	P	AD	P	AD	P
Normal	1.776	<0.005	1.14	<0.005	2.907	<0.005	1.14	<0.005	1.576	<0.005
3-parameter Lognormal	1.708	*	0.598	*	1.588	*	0.598	*	1.655	*
2-parameter exponential	1.805	<0.010	1.64	<0.010	5.732	<0.010	1.64	<0.010	2.653	<0.010
3-parameter Weibull	0.554	**0.16**	0.542	**0.173**	0.534	**0.182**	0.542	**0.173**	0.988	**0.015**
Smallest extreme value	2.04	<0.010	1.228	<0.010	2.819	<0.010	1.228	<0.010	1.646	<0.010
Largest extreme value	1.324	<0.010	1.226	<0.010	3.053	<0.010	1.226	<0.010	1.614	<0.010
3-parameter gamma	2.297	*	0.551	*	1.481	*	0.551	*	2.006	*
Logistic	1.362	<0.005	1.116	<0.005	2.799	<0.005	1.116	<0.005	1.519	<0.005
3-parameter Loglogistic	1.537	*	0.51	*	1.43	*	0.51	*	1.47	*
Johnson transformation	0.478	0.182	0.274	0.58	N/A	N/A	0.274	0.58	N/A	N/A

*that indicate that the software don't get a specific value for that data
Bold values indicate the best parameter for each prediction

Table 20 Summarizes parameters Chihuahua, Sinaloa, and Michoacán amounts claimed for allocation of distributions, own creation with Minitab 18 [50]

	Chihuahua				Sinaloa				Michoacán			
Distribution	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold
Normal[a]	4879.1835		9539.28904		3794.77004		5288.25368		5342.531		15,221.95412	
3-parameter Lognormal	-9.45064		21.73256	0	5.86511		3.0597	-5.92907	-10.19059		21.16748	0
2-parameter exponential			5421.26076	-542.12608			4216.36897	-421.6369			5936.08617	-593.60862
3-parameter Weibull		0.43418	1950.79431	-26.13128		0.49926	2073.75327	-41.30654		0.30615	560.91713	-2.57181
Smallest extreme value	10,230.02832		12,544.38197		6511.35298		5460.32337		14,036.6836		20,718.61434	
Largest extreme value	1749.11625		4024.15742		1605.86035		3204.8022		999.88359		4908.48057	
3-parameter gamma		0.04887	99,839.72952	0		0.19789	19,176.80189	-0.06826		0.04688	1.14E+05	0
Logistic	2777.50813		3679.70806		2933.7076		2859.23992		1622.83802		4856.52199	
3-parameter Loglogistic	-7.52599		13.77753	0	6.41154		1.51029	-30.86196	-8.36513		13.41547	0
Johnson transformation[a]	0.08686		0.77879		-0.00337		1.08791					

[a]Scale: Adjusted ML estimate

Table 21 Summarizes parameters Jalisco and Tamaulipas amounts claimed for allocation of distributions, own creation with Minitab 18 [50]

	Jalisco				Tamaulipas			
Distribution	Ubication	Shape	Escale	Threshold	Ubication	Shape	Escale	Threshold
Normal[a]	3794.77004		5288.25368		220.465		369.27495	
3-parameter Lognormal	5.86511		3.0597	−5.92907	−15.27041		20.79127	0
2-parameter exponential			4216.36897	−421.6369			244.95866	−24.49587
3-parameter Weibull		0.49926	2073.75327	−41.30654		0.30922	44.97265	−0.21286
Smallest extreme value	6511.35298		5460.32337		415.91963		411.79953	
Largest extreme value	1605.86035		3204.8022		78.61077		192.76601	
3-parameter gamma		0.19789	19,176.80189	−0.06826		0.04291	5138.17754	0
Logistic	2933.7076		2859.23992		141.29973		179.27123	
3-parameter Loglogistic	6.41154		1.51029	−30.86196	−15.28843		13.46422	0
Johnson transformation[a]	−0.00337		1.08791					

[a]Scale: Adjusted ML estimate

Table 22 Summary statistics behavior amounts claimed by state, own creation with Minitab 18 [50]

Estate	Behavior
Chihuahua	3-Weibull (0.43418, 1950,79,431, −26.13128)
Sinaloa	3-Weibull (0.49926, 2073.75327, −41.30654)
Michoacán	3-Weibull (0.30615, 560.91713, −2.57181)
Jalisco	3-Weibull (0.49926, 2073.75327, −41.30654)
Tamaulipas	3-Weibull (0.30922, 44.97265, −0.21286)

Table 23 Prediction results of each method, own creation with Minitab 18 [50]

Prediction method	Prediction accuracy (%)
Moving averages	49.07%
Weighted moving averages	32.02%
Exponential smoothing	61.18%
Simple exponential smoothing	66.03%
Multiplicative model	60.60%
Models based on periodic data	46.22%
Regression analysis	63.86%
Regression analysis with time series	61.62%
ABS model	72.87%

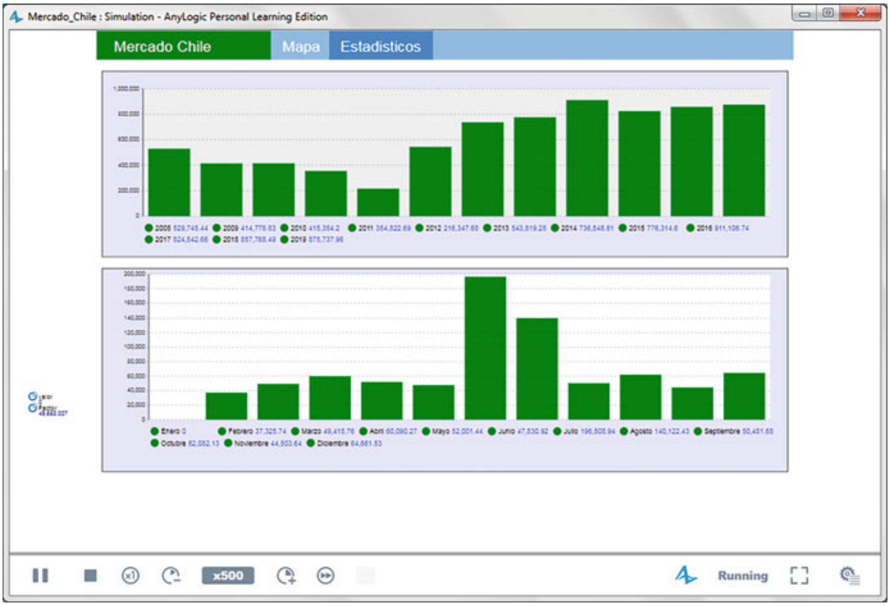


Fig. 34 Graphs of quantities of Jalapeño peppers harvested by state, own creation according to AnyLogic 8.0

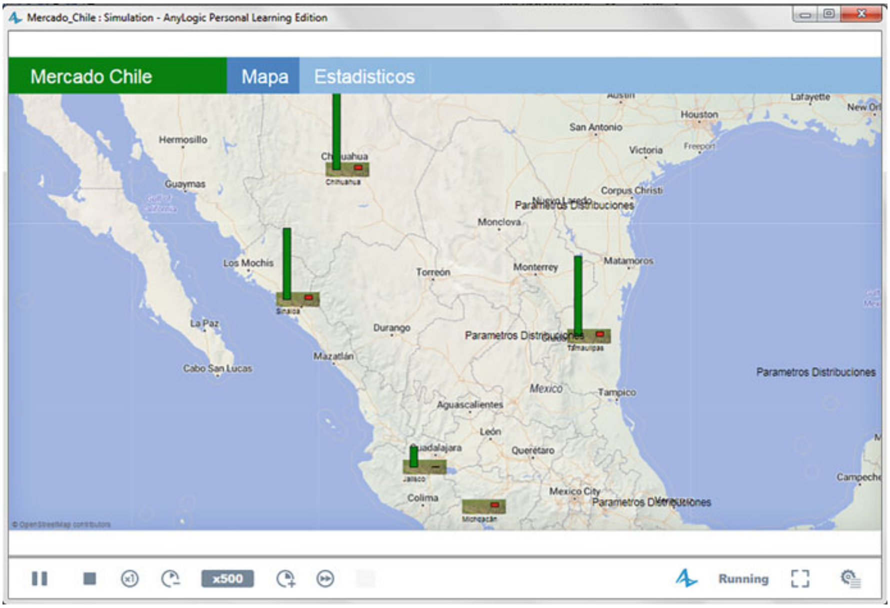


Fig. 35 Simulation of quantities of Jalapeño pepper harvested and damaged by state, own creation according to AnyLogic 8.0

Due to the lack of updated data in the pages of INEGI and SEMARNAT, the comparison of the results was carried out in the periods from 2015 to 2017; the performance of the ABS model has been tested and the differences between the most common mathematical models have been compared. According to the results in Table 23, the mathematical model that has a better prediction obtains 66.03% compared to 72.87% for the ABS model, with this better results are obtained and without the complexity of estimating seasonality values among others.

For future work and to improve the model obtained, it is necessary to look for more direct and reliable sources of information on some variables that were not considered due to the lack of available information, such as fertilizer applications, quality of cultivated land, and technological level of the farms, among others.

5 Conclusion

In this chapter, a tool for predicting the amount of jalapeño peppers that will be available in the market was proposed, so that with this information, companies producing metal containers can have a simple and low complexity estimation tool, compared to models based on mathematical equations. First, the necessary general variables that impact the jalapeño pepper harvest were established, where variables such as temperature and amount of rainfall were the ones that best explained the harvest obtained in each period; the relevance was calculated according to the historical data of jalapeño pepper harvest, and their behaviors were established according to the statistical results obtained.

The ABS model proposed in the chapter can be applied to the prediction of the amount of jalapeño peppers that would be available in the market.

References

1. Aburukba, R., Ghenniwa, H., Shen, W.: Agent-based dynamic scheduling approach for collaborative manufacturing. In: Proceedings of the 2007 11th International Conference on Computer Supported Cooperative Work in Design. Ontario, Canada (2007)
2. Aksyonova, K., et al.: Architecture of the multi-agent resource conversion processes extended with agent coalitions. Procedia Comput. Sci. **105**, 221–226 (2017)
3. Albadawi, Z., et al.: Agent-based control of manufacturing processes. Int. J. Manufact. Res. **1**(4), 466–481 (2006)
4. Alpaydin, E.: Introduction to Machine Learning, 3rd edn. MIT Press, Cambridge (2014)
5. Anderson, D., Dennis, J., and Thomas, A.: Quantitative Methods for Business, 11th edn. Cengage Leaning, Boston (2016). ISBN-13: 978-607-481-697-6, ISBN-10: 607-481-697-2
6. Athanassios, N., Avramidis, M., Gendreau, P.: Optimization of agent scheduling in multiskilling call centers. In: Submitted to: 2007 Industrial Simulation Conference. Montreal, Canada (2007)
7. Aurrand, L., Jarri Fournier, P., et al.: Application of fuzzy control for ISIS vehicle braking. In: Proceedings of Fuzzy and Neuronal Systems, and Vehicle applications'91. (1991)
8. Azimian, A., Aouni, B.: Supply chain management through the stochastic goal programming model. Ann. Oper. Res. **251**, 351–365 (2017). https://doi.org/10.1007/s10479-015-2007-1
9. Bandyopadhyay, S.: Tarantula mating based multi-agent routing strategy for manufacturing problems. In: Computer Science and Information Technology (ACSIT), vol. 2, Num. 5, pp. 402–406 (2015). www.krishisanskriti.org
10. Ben-Dora, G., Ben-Lia, E., Benenson, I.: Assessing the impacts of dedicated bus lines on Urban traffic congestion an modal split with an agent-based model. Procedia Comput. Sci. **130**, 824–829 (2018)
11. Bildirici, M., Bakirtas, T.: The relationship among oil, natural gas and coal consumption and economic growth in BRICTS (Brazil, Russian, India, China, Turkey and South Africa) countries. Energy **65**, 134–144 (2014)
12. Bischoft, J., Maciejewski, M.: Agent-based simulation of electric taxi-cab fleets. Transport. Res. Procedia **4**, 191–198 (2014)
13. Bjorn, R., Franziska, K.: Qualitative validation from within–Extension of a simulation framework and test." In: Department of Technology Orebro University, Karlskoga, Sweden (2009)
14. Bok, M., Tavsazzy, L.: An Empirical agent-based simulation system for urban goods transport. Procedia Comput. Sci. **130**, 126–133 (2018)
15. Boloni, L., et al.: Agent based scientific simulation and modeling. Concurrency Practice Exper. **200**(12), 845–861 (2000)
16. Borshchev, A.: The Big Book of Simulation Modeling: Multimethod Modeling with AnyLogic 6. AnyLogic North America, Oakbrook Terrace (2013)
17. Borshchev, A., Filippov, A.: From system dynamics and discrete event to practical agent based modeling: Reasons, techniques, tools. In: The 22nd International Conference of the System Dynamics Society. (2004)

18. Bosse, T., Gerritsen, C., Treur, J.: Agent-Based Simulation of Episodic Criminal Behaviour (2008). http://www.few.vu.nl
19. Caramuta, C., et al.: Survey of detention techniques, mathematical models and simulation software in pedestrian dynamics. Transport. Res. Procedia **25**, 551–567 (2017)
20. Cavalcantea, R.A., Roorda, M.: Freight market interactions simulation (FREMIS): an agent-based modeling framework. Procedia Comput. Sci. **19**, 867–873 (2013)
21. Chang, C., Lin, J., Jin, P.: A grey modeling procedure based on the data smoothing index for short-term manufacturing demand forecast. Comput. Math. Organ. Theory **23**, 409–422 (2017). https://doi.org/10.1007/s10588-016-9234-0
22. Chappin, M.S., Dijkema, G., Vries, L.J.: Agent-based simulation of carbon policies and power generation. In: Presented at the IAEE Conference. San Francisco, pp. 21–24 (2009)
23. Charles, M., Michael, J.: Agent-based modeling and simulation: Desktop ABMS. In: Proceedings of the 2007 Winter Simulation Conference. Argonne, IL (2007)
24. Chena, Y., Tana, H., Song, X.: Day-ahead forecasting of non-stationary electric power demand in commercial buildings: hybrid support vector regression based. Energy Procedia **105**, 2101–2106 (2017)
25. Cheng, S.F., Lim, Y.P.: An agent-based commodity trading simulation. In: School of Information Systems Singapore Management University (2009). http://www.aaai.org/
26. Chick, S., et al.: Agent-based modeling and simulation in construction. In: Proceedings of the 2003 Winter Simulation Conference Fulton School of Engineering Arizona State University, USA (2003)
27. Christley, S., et al.: Multiscale agent-based simulation for chondrogenic pattern formation in vitro. Cyber. Syst. Int. J. **38**(7), 707–727 (2012). http://doi.org/10.1080/01969720701534125
28. Craenen, B.G.W., Theodoropoulos, G.K.: Ubiquitous computing and distributed agent-based simulation. In: School of Computing University of Birmingham. Birmingham, United Kingdom (2010)
29. Crouzet, M.: General History of Civilizations: The 19th Century. Destiny Edition 4 (1969)
30. Devendran, A., Lakshmanan, G.: Urban growth prediction using neural network coupled agents-based cellular automata model for Sriperumbudur Taluk, Tamil Nadu, India. Egypt. J. Remote Sens. Space Sci. **21**, 353–362 (2018)
31. Dionisio, A., Rocha, R., Barata, J.: An agent based monitoring architecture for plug and produce based manufacturing systems. In: University of Nova, Lisboan (2012)
32. Domit, S., Monti, F., Sokol, A.: Forecasting the UK economy with a medium-scale Bayesian VAR (2019). https://doi.org/10.1016/j.ijforecast.2018.11.004
33. Dong, H., Zheng, H., Li, Y.: Model predictive control for inventory management in supply chain planning. Procedia Eng. **15**, 1154–1159 (2011)
34. Druckenmiller, D., William, A., Marvin, T.: Agent based modeling and simulation of causal maps. In: Americans Conference on Information Systems (AMCIS), Proceedings. Paper 217. USA (2004)
35. Dumitrache, L., Caramihai, S., Stanescu, A.M.: Intelligent agent-based control systems in manufacturing. In: 11 Proceedings of the 15th IEEE international Symposium on Intelligent Control (ISIC2000). Rio Patras, Greece (2000)
36. Edmons, B., Moss, S., Davidsson, P.: A crop overview of pepper. In: Agrifood and Fisheries Information Service Magazine, pp. 1–20 (2001)
37. Edmons, B., Moss, S., Davidsson, P.: The use of models - making MABS actually work. In: Multi-Agent-Based Simulation. Lecture Notes in Artificial Intelligence 1979. Berlin, pp. 15–32 (2001)
38. Elsevier, B.V.: Simulation-based design and evaluation of multi-agent systems. Simulat. Modell. Practice Theory **18**, 1425–1427 (2010)
39. Flowers, J., Rivera, F.: Evaluation of selection criteria for probabilistic models: Validation with series of simulated. In: Technology and Water Sciences, vol. V, no. 5, pp. 189–197 (2014)
40. Gabardoa, F., Pereimab, B., Porcilec, G.: The dynamics of sectoral labour reallocation: an agent-based model of structural change and growth. Economia **24**, 867–873 (2018)

41. Grazziela, P., Aickelin, U., Siebers, P.: Systems dynamics or agent-based modelling for immune simulation?' In: Intelligent Modelling and Analysis Research Group, School of Computer Science. UK (2010)
42. Guoa, X., Lia, D., Zhang, A.: Improved support vector machine oil price forecast model based on genetic algorithm optimization parameters. AASRI Procedia **1**, 525–530 (2012)
43. Haltas, I., et al.: Anaerobic digestion: a prime solution for water, energy and food nexus challenges. Energy Procedia **123**, 22–29 (2017)
44. Han, S., et al.: Peak coal in China: a literature review. Resour. Conservat. Recycling **129**, 293–306 (2018)
45. Hernandez, E., Muñoz, V.: Chile as food. Science Mag. July–September **2015**, 16–23 (2015)
46. Hernández, E., Ocotero, V.: Pepper like food. Science Mag. July–September **2015**, 16–23 (2015)
47. Heywooda, P., et al.: Data-parallel agent-based microscopic road network simulation using graphics processing units. Simulat. Modell. Practice Theory **83**, 188–200 (2018)
48. Hu, Y., Jiang, P.: Forecasting energy demand using neural-network-based grey residual modification models. J. Oper. Res. Soc. **68**, 556–565 (2017)
49. Hu, Y., et al.: System dynamics of oxyfuel power plants with liquid oxygen energy storage. Energy Procedia **142**, 3727–3733 (2017)
50. INEGI (2017). https://www.inegi.org.mx/
51. Iorio, S., Chen, F.: Agent-based simulation as an implementation of methodological individualism. In: National Chengchi University, Taipei, Taiwan (2013)
52. Jiang, W., et al.: Research on cloud manufacturing resource integrating service modeling based on cloud-agent. In: Zhengzhou, China (2011)
53. Jianming, Y., Krishna, K.: A conceptual framework for agent-based agile manufacturing cells. In: Info Syst. J. **14**, 93–109 (2004)
54. Julong, D.: Grey System theories and measurement futurics. Future CRND Develop. **3**, 3–20 (1983)
55. Julyan, H., Cartwright, E.: Agent-based social simulation: A dynamical-systems viewpoint. Cyber. Syst. Int. J. **41**, 281–286 (2010)
56. Jung, T., Shan, P., Weyrich, M.: Dynamic co-simulation of internet-of-things-components using a multi-agent-system. Procedia CIRP **72**, 874–879 (2018)
57. Kamel, J., et al.: Exploring the impact of user preferences on shared autonomous vehicle modal split: a multi-agent simulation approach. Transport. Res. Procedia **37**, 115–122 (2018)
58. Kaserekaa, S., et al.: Agent-based modelling and simulation for evacuation of people from a building in case of fire. Procedia Comput. Sci. **130**, 10–17 (2018)
59. Kazemi, S.: Computational Simulation Of Linguistic Change: An Agent-Based Model. A Thesis Presented to the Faculty of San Diego State University (2014)
60. Killian, M., Mayer, B., Kozek, M.: Hierarchical fuzzy MPC concept for building heating control. In: 19th IFAC World Congress (2014)
61. Koen, B., Magnus, B.: Agent based social simulation in markets. In: Symposium Dept of Business Administration University of Namur, Belgium (2000)
62. Kornienko, S., Kornienko, O., Levi, P.: Flexible manufacturing process planning based on the multi-agent technology. In: Proceedings of the 21st IASTED International Conference Applied Informatics. Innsbruck, Australia (2003)
63. Kraana, O., et al.: Jumping to a better world: an agent-based exploration of criticality in low-carbon energy transitions. Energy Res. Soc. Sci. **47**, 156–165 (2018)
64. Krajzewiez, D., Heinrich, M., Beige, S.: Embedding intermodal mobility behavior demand model. Procedia Comput. Sci. **130**, 865–871 (2017)
65. Leitao, P., Baratan, J.: Agent-based disturbance handling architecture in manufacturing polytechnic. In: Institute of Braganza, Portugal (2006)
66. Leitao, P., Barata, J.: An agent-based disturbance handling architecture in manufacturing control. In: Polytechnic Institute of Braganca. Portugal (2007)
67. Lia, S., Cheng, Y.: Analyzing of renewable and non-renewable energy consumption via Bayesian inference. Comput. Math. Appl. **53**, 1904–1920 (2007)

68. Lin, H., et al.: Agent based modeling for estimating energy-saving potential of offices under different pricing mechanisms. Energy Procedia **105**, 3881–3888 (2017)
69. Liu, P., Wu, S.: An agent-based simulation model to study accountable care organizations. Health Care Manag. Sci. (2009). https://doi.org/10.1007/s10729-014-9279
70. Makinoshimaa, F., Imamura, F., Abe, A.: Enhancing a tsunami evacuation simulation for a multi-scenario analysis using parallel computing. Simulat. Modell. Practice Theory **83**, 36–50 (2018)
71. Malleson, N., Heppenstall, A., See, L.: Crime reduction through simulation: An agent-based model of burglary. In: Computers, Environment and Urban Systems, vol. 34. United Kingdom, pp. 236–250 (2010)
72. Manleya, E., Chengb, T.: Exploring the role of spatial cognition in predicting urban traffic flow through agent-based modelling. Transport. Res. Part A **109**, 14–23 (2018)
73. Marzouki, B., Belkahla, O., Ghedira, K.: Agent model based on chemical reaction optimization with greedy algorithm for flexible job shop scheduling problem. Procedia Comput. Sci. **112**, 81–90 (2017)
74. Mason, R., et al.: Distributed agent-based simulation of construction projects. In: Proceedings of the 2008 Winter Simulation Conference, University of Alberta Edmonton. AB, Canada (2004)
75. Matthew, C., Martyn, A.: Sim Zombie: A case-study in agent-based simulation construction. In: School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University. Manchester (2009)
76. Moergestel, L., Puik, E., Telgen, D.: Decentralized autonomous-agent-based infrastructure for agile multiparallel manufacturing. In: Faculty of Science and Technology Utrecht University of Applied Sciences. Netherlands (2010)
77. Moncada, J., et al.: Exploring policy options to spur the expansion of ethanol production and consumption in Brazil: an agent-based modeling approach. Energy Policy **123**, 619–641 (2018)
78. Montanola-Sales, C. et al.: Agent-based simulation validation: A case study in demographic simulation. In: The Third International Conference on Advances in System Simulation. Barcelona, Spain (2011)
79. Montanola-Sales, C. et al.: Demographic agent-based simulation of Gambians immigrants in Spain. In: Departament dÉstadistica Investigació Operativa (DEIO), Barcelona, España, vol. 106 (2013)
80. Nadimi, R., Tokimatsu, K.: Analyzing of renewable and non-renewable energy consumption via Bayesian inference. Energy Procedia **142**, 2773–2778 (2017)
81. Nair, D., Rashidi, T., Dixit, V.: Estimating surplus food supply for food rescue and delivery operations. Socio-Econ. Plann. Sci. **57**, 73–83 (2017)
82. Page, C., et al.: Agent-based modelling and simulation applied to environmental management. In: Simulating Social Complexity, Understanding Complex Systems (2013). https://doi.org/110.1007/978-3-540-93813-2_19
83. Paulsena, M., Rasmussena, T., Nielsena, O.: Output variability caused by random seeds of a multi-agent transport simulation model. Procedia Comput. Sci. **130**, 850–857 (2018)
84. Pereira, A., Duarte, P.: Agent-based simulation of ecological models. In: Proceedings 5th Workshop on Agent-Based Simulation. Porto, Portugal (2004)
85. Petri, Y.: Agent-based simulation and sociological understanding. In: Department of Social Research University of Helsinki (2011)
86. Pulido, A., Perez, J.: Econometric Models. Madrid (2001)
87. Raabe, B., et al.: Collaboration platform for enabling industrial symbiosis: application of the by-product exchange network model. Procedia CIRP **61**, 263–268 (2017)
88. Ramírez, S., Castañeda, W., Velásquez, J.: Representation of unlearning in the innovation systems: a proposal from agent-based modeling. Estudios Gerenciales **33**, 366–376 (2017)
89. Reaidy, J., Massotte, P., Daniel, D.: Comparison of negotiation protocols in dynamic agent-based manufacturing systems. In: Laboratoire de Genie Informatique et Ingenierie de Production, Ecole des Mines. France (2004)

90. Remondino, M., Correndo, G.: Data mining applied to agent based simulation. In: Department of Computer Science University of Turin, Italy (2005)
91. Remondino, M., et al.: Periodical insects and prime numbers: An agent based simulation. In: University of Turin, Italy (2004)
92. Rioja, I.: Mexico Annual Report 2018 (2018). https://www.industriaalimenticia.com/
93. Robert, C.P.: Data Analytics for More Accurate Forecasts. Springer, Berlin (2001)
94. Romano, D., Lawrence, L.: NARCSIM an agent-based illegal drug market simulation. In: University of Sheffield. Sheffield, UK (2009)
95. SAGARPA (2016). https://www.gob.mx/agricultura
96. Sabine, T.: Appropriating places in public spaces a multi-agent simulation. In: Geoinformatics, Department of Geography University of Augsburg (2009). sabine.timpf@geo.uni-augsburg.de
97. Salas, R., Bauza, A.: Data analytics for more accurate forecasts. Emphas. Logist. July-August **2018**, 58–59 (2018)
98. Schenk, T.: Qualitative social science methods for multi-agent models. In: Institut fur Geographie, Universitat Leipzig, Johannisallee. Leipzig, Germany (2008)
99. Shaabana, K., Abdel-Waritha, K.: Agent-based modeling of pedestrian behavior at an unmarked midblock crossing. Procedia Comput. Sci. **109C**, 26–33 (2017)
100. Shahabpoor, E., Pavic, A., Racic, V.: Identification of walking human model using agent-based modelling. Mech. Syst. Signal Process. **103**, 352–367 (2018)
101. Shih, C., et al.: Grey number prediction using the grey modification model with progression technique. Appl. Math. Modell. **35**, 1314–1321 (2011)
102. Silva, M., et al.: Control Problems of grey systems. Syst. Control Lett. **1**(5), 211–215 (1982)
103. Silva, M., et al.: Assessment of forecasting models for patients arrival at emergency department. Oper. Res. Health Care **18**, 112–118 (2018)
104. Smith, N., Capiluppi, A., Fernandez, J.: Agent-based simulation of open source software evolution. In: International Software Process Workshop and International Workshop on Software Process Simulation and Modelling, Shanghai, China, pp. 20–21 (2006)
105. SMN (2019). https://www.smn.cna.gob.mx/
106. Songa, H., et al.: Energy consumption combination forecast of hebei province based on the IOWA operator. Energy Procedia **5**, 2224–2229 (2011)
107. Stainsby, H., Taboada, M., Luque, E.: Towards an agent-based simulation of hospital emergency departments. In: University Autonomy de Barcelona, Spain (2009)
108. Sumari, S., et al.: Comparing three simulation model using Taxonomy: System dynamic simulation. Int. J. Manag. Excellence **1**(3), 54–59 (2013)
109. Tanrikulu, O.G.: Advancing rationality with sustainability: an analysis of agent-based simulation. Hatfield Graduate J. Public Affairs **2**, Article 3 (2017)
110. Telgen, D., et al.: Automatic structured decomposition of manufacturing actions in an agent-based manufacturing system. In: University of Applied Sciences Utrecht. The Netherlands (2013)
111. Tibor, B., Charlotte, G., Michel, K.: Agent-based simulation of social learning in criminology. In: Amsterdam, The Netherlands (2009)
112. Timo, K., Simo, H., Ehtamob, H.: Counterflow model for agent-based simulation of crowd dynamics. In: Conference submitted to Safety Science, Aalto, Finland (2010)
113. Tsai, S., Lee, Y., Guo, J.: Using modified grey forecasting models to forecast the growth trends of green materials. J. Eng. Manufact. **228**(6), 931–940 (2014). https://doi.org/10.1177/0954405413509079
114. Vizzari, G.: Situated agents based pedestrian simulation: When reactive agents are not enough. In: Complex Systems and Artificial Intelligence Research Center. University of Milan-Bicocca, Italy (2008)
115. Wang, X.: Application research on electricity demand forecasting based on gaussian quadrature formula. Procedia Eng. **15**, 5574–5578 (2011)
116. Wang, X.: Forecasting modeling and analysis of power engineering in China based on Gauss-Chebyshev formula. Syst. Eng. Procedia **5**, 131–136 (2012)

117. Wang, Q., et al.: Potential of energy saving in a data center application of an agent-based modelling. Energy Procedia **105**, 3903–3908 (2017)
118. Wein, A., Labiosa, W.: Serious games experiment toward agent-based simulation. U.S. Geological Survey, Reston, Virginia. Open File Report 112, pp. 2013–1152 (2013)
119. Widowatia, et al.: Implementation of ARIMA Model to Assess Seasonal Variability Macrobenthic Assemblages. Aquatic Procedia **7**, 277–284 (2016)
120. William, M., Matthew, W.: Agent-based simulation for human-induced hazard analysis. Risk Analy. **31**(2), 205–217 (2011)
121. Woltmann, A., et al.: Agent based simulation model of virtual power plants for greener manufacturing. Procedia CIRP **69**, 377–382 (2018)
122. Yasini, S.H., Naghibi-Sistani, B., Karimpour, A.: Agent-based simulation for blood glucose control in diabetic patients. Int. J. Appl. Sci. Eng. Technol., **5**, 5–12 (2015)
123. Young, L., Wilkinson, I., Held, F.: Modelling of marketing and business systems: agent-based simulation, discipline of marketing. In: The University of Sydney Business School. Western, Sydney (2011)
124. Zhanga, X., Wanga, J., Zhang, K.: Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. Electric Power Syst. Res. **146**, 270–285 (2017)
125. Zhemin, L., Ganqiong, L., Yuting, W.: Construction of short-term forecast model of eggs market price. Agricult. Agricult. Sci. Procedia **1**, 396–401 (2010)
126. Zhua, Y., et al.: Hurricane evacuation modeling using behavior models and scenario-driven. Procedia Comput. Sci. **130**, 836–843 (2018)

Personalised Emotion Detection from Text Using Machine Learning

A. V. Bhavya, R. H. Dhanush, J. Sangeetha, and Arun Cyril Jose

Abstract This research discusses an emotion recognition system, which is an important component of many effective computing technologies for natural language processing, based on open-source platforms with automatic speech recognition (ASR) and text analysis, and which is used for user-based customised sentiment analysis. PocketSphinx as ASR and Word2vec model, K-means clustering, and TfidfVectoriser for text automatic analysis are used to design the proposed framework. Further, the dataset that is used for testing and training the model is from International Survey on Emotion Antecedents and Reactions (ISEAR). This research yields a user-dependent system that will function as a tailored assistant for identifying emotional responses and discovering innovative applications. The suggested model greatly outperforms the prior models, with an efficiency of 81% and an f-measure of 89%.

Keywords Emotion detection · Text analysis · Machine learning · Word2vec model

1 Introduction

Models and classifiers fail to generalise when training and testing settings differ, which is one of the major challenges in emotion recognition [1]. This challenge is seen when the data from a person in the testing dataset is not confined when training. In reality, earlier studies have shown that speaker-dependent classifiers perform better than speaker-independent classifiers [1]. This statement implies that speaker dependencies are present in the articulation of emotions. Although there

A. V. Bhavya · R. H. Dhanush · J. Sangeetha · A. C. Jose (✉)
Department of Computer Science and Engineering, Indian Institute of Information Technology, Kottayam, Kerala, India
e-mail: avbhavya18bcs@iiitkottayam.ac.in; sangeethaj2017@iiitkottayam.ac.in; aruncyril@iiitkottayam.ac.in

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_10

are common patterns among speakers, these traits are insufficient to create reliable emotion identification systems [2]. The issue is that practical procedures have strong classifiers that can generalise to the suggestive speech of unseen speakers. The paper discusses the issue of adapting an emotion identification system to a specific user. The use of feature and/or model adaptation for emotion recognition is an intriguing strategy [2]. The fact that the model adaption frame requests data from the user with emotional labelling limits their usage in many practical operations [3]. This study introduces an unsupervised model to adapt an emotion recognition system to a particular user in this setting. The suggested technique seeks to reduce the discrepancy between aural information retrieved from the targeted speaker utilised during testing and those features used during training. A validated emotion model is required to increase the realism of the recognition process since it can effectively handle the variety of emotions that can arise from different scenarios depending on the individual [4]. The accuracy of traditional styles in the environment of SER (speech emotion recognition) is comparatively low [4]. Supervised learning requires manually annotated data, which is frequently time-consuming and not always possible. We're proposing a hybrid model addressing these issues. The ISEAR (International Survey on Emotion Antecedents and Reactions) dataset is used to build the emotion detection system. Videos from a well-liked videotape-participating website, containing colourful interviews from a targeted subject, are downloaded, and 1.5 hours of speech from the targeted speaker is extracted [5]. The experimental results indicate that the proposed system gives accuracy of 80% (approx.). This paper has the following contents: In Sect. 2, the related works for this project are discussed. Section 3 contains the architecture of the proposed model. Section 4 discusses the technique used to implement the architecture. Section 5 details about the dataset, and Sects. 6 and 7 discuss how the implementation is done and the results obtained.

1.1 Contributions

Here, we implement unsupervised training, without any prior assumptions. We have used transfer learning method (Word2vec algorithm), which increased the accuracy to 80%. In our work, we made emotion personalised to each individual. In this work, primarily, text is used for emotion detection, and if the provided speech as input, it is handled by converting speech to text.

2 Literature Review

Semantic information may be extracted using semantic analysis, ontologies can be created using emotion models, and case-based reasoning can be used to adopt new keywords [5]. Their suggested architecture aims to offer improved flexibility for various domains and systematic processing of text input.

Semantic analysis: To retrieve extra information about phrases, comparable to dependency trees, semantic analysis uses methodologies of natural language processing such as statistic-based parsing. On the basis of sentence verbal information, keyword extraction was suggested [5].

Sentiment models: It defines the underpinning research required to link the results of semantics to plausible emotions. An OCC framework, which covers 22 emotion types, is employed. Implementing these algorithms into semantics resulted in a more methodical framework for analysing various linguistic feeds [5].

Case-based reasoning: It was advised for embracing new concepts and applications. The source is compared with every case stored in storage in a case-based reasoning system, and the distances of both the intake and every case are calculated using techniques to determine the amount of resemblance. The approximation approach was not created, and the information and instance criteria were not completed manually, which impacted the efficiency [5].

The keyword-based component of a hybrid architecture [6] is mostly based on the knowledge engineering approach for information extraction. Using the sentence separator, tokeniser, and POS tagger to capture syntactically and semantically data, hidden phrase patterns may be discovered and contextual analysis improved. Furthermore, the training module builds selected features and predicts accuracy using the LibSVM standard. The machine learning method is split into two stages. The initial stage is to develop an emotion prediction training model. The learnt model will be used to predict the sentiment class on a testing dataset. However, they then separated people based on industry-specific prejudice.

The goal of normalising features is to reduce speaker variability while keeping the capacity to differentiate between various emotional classes. The discrepancies between the audio characteristics utilised to train and test the emotion identification system were reduced [6], resulting in the anticipated outcome. They also used the front-end of the iterative feature normalisation (IFN) approach. Even after using the IFN, however, many samples were wrongly labelled. Shaheen and Hajj [14] performed a difficult syntactic and semantic analysis of the text and utilised a range of ontologies, including WordNet and ConceptNet, to detect emotions. WordNet and ConceptNet are used to help their classifier generalise the training data, which enhances emotion coverage. Their classifier is context sensitive due to syntactic and semantic examination of the phrase.

Based on the process of detection, it is separated into three types:

Approach according to keywords: Methods according to keywords are used at the fundamental word level. They require a lexicon of emotional words that pairs words with labels for the associated emotions [7]. Approach based on rules: Using this strategy, different rules are developed to form a linguistic structure that is helpful in determining emotion [8]. Approach based on learning: This technique articulates the challenge of identifying sentiment from message as a set of input texts with feelings as categories. The supervised technique trains a classifier using hand-labelled datasets, which is then applied to other groups. Unsupervised methods do not require any labelled data [8].

Based on the data source used, it is separated into three types:

Knowledge-based approach: These approaches use synonyms of lexical resources to ascertain the mood or polarity in terms, phrases, and documents [9]. Approach based on corpora: It is a repository of vast and organised sets of information, typically tied to a specific topic or author [10]. Fusion approach: It is a type of hybrid method which employs both the previously mentioned approaches [10].

Based on the user perspective, it is divided into two types:

Reader perspective: From a reader's perspective, one particular text segment can evoke multiple emotions [11]. From a writer's perspective: In the perspective of the writer where one text segment portrays only one emotion [11].

To recognise the effect as six fundamental emotions, Anusha and Sandhya [1] combined machine learning and natural language processing methods. But lengthy sentences cannot be utilised with it. Feature selection and extraction, feature categorisation, acoustic modelling, unit-based recognition, and language-based modelling are some of the components that make up a SER system [12]. The different nonlinear parts that make up deep learning techniques carry out computation in parallel. To address the drawbacks of other procedures, these approaches still need to be more deeply layered in their framework. Convolutional neural networks (CNN), deep belief networks (DBNs), recurrent neural networks (RNN), recursive neural networks (RvNN), deep Boltzmann machines (DBMs), and auto encoders (AE) are some examples of deep learning techniques, used for SER [13] that significantly improves the overall performance of the designed system.

Deep Boltzmann machines (DBMs) are made up of a variety of hidden layers and are primarily derived from Markov random fields. These layers are based on variables that were selected at random and linked to stochastic elements. The primary benefits of DBM are its propensity to learn quickly and provide useful representation. This is accomplished through layer-by-layer pretraining [13].

An RNN is a type of neural network that uses sequential information and has interdependent outputs and inputs. This interdependency is typically helpful in anticipating the input's future state. The RNN's susceptibility to gradient disappearance is the primary issue affecting its overall performance [14].

Recursive neural network (RvNN) is a hierarchical deep learning algorithm that does not rely on a tree-structured input sequence. By breaking up the input into manageable bits, it may quickly learn the parse tree of the supplied data [14].

Deep belief network (DBN) is constructed from cascading RBM structures and has a significantly more complex structure. RBMs are extended into DBNs, where RBMs are bottom-up taught layer by layer. Due to their capacity to learn the recognition parameters fast, regardless of how many parameters there are, DBNs are typically utilised for speech emotion recognition. Additionally, it prevents layer nonlinearity.

These layer-wise frameworks for deep learning algorithms are succinctly created based on the classification of various natural sentiments. These methods provide simple model training in addition to the effectiveness of added weights. Deep learning approaches include limitations such as being their huge internal layer-by-layer architecture, poorer effectiveness for temporally variable input data, and overlearning during layer-by-layer memory [15].

Iliev and Peter [5] applied data retrieval based on artwork by using sentiment from voice. The ability to extract sentiment from speech and use it to identify and recommend art has been proven to be quite useful. To locate and suggest digital content, they used speaker sentiment. They started the process of creating a conceptual model of an ecosystem for digital culture that serves several purposes. It is founded on pre-defined text queries for metadata. For multiclass text analysis and classification, Ramya H. R. and Dr. Mahabaleswara Ram Bhatt employed PocketSphinx [16] as an automatic speech recogniser, human-computer interaction (HCI) [16], and linear support vector machine (LineaSVM). Future directions of the study could involve applying ASR based on neural networks and convolution neural network (CNN) techniques to noisy environments. More data, including a database of annotated emotions, Twitter data, and other sources, should be used to train the text analyser. By dividing sentiment into positive and negative categories, CNN with sentiment analysis can be utilised to broaden research investigation. Using hybrid classifiers, such as NB-SVM and long-/short-term memory techniques, improves accuracy. A dialog-based or interaction-based emotion recognition system can be developed for usage in practical applications to execute the HMI cycle. Sentiment analysis levels, different emotion models, and the process of sentiment analysis and emotion identification from text were all employed by Pansy Nandwani and Rupali Verma [9]. Deep learning approaches include limitations such as their huge internal layer-based design, lower efficacy for temporally variable input data, and overlearning during layer-based information memory. Another significant issue is that it might be challenging to infer polarity from comparison phrases.

3 Methodology

We provide a customised emotion detection model to improve the realism of the recognition process. The model's flow is as follows: If the source exists in the shape of voice signals, it tries to classify emotional reactions collected via ASR in conjunction with textual analysis, and if the intake is text, it may be categorised straight by this concept.

In this paper, PocketSphinx is used for speech recogniser, while a word embedding model is used for categorisation. PocketSphinx is optimum lightweight recogniser developed in C language, and it promotes portable, memory optimisation, ease of use, and wide vocabulary continuous voice recognition. PocketSphinx can be used in android and in offline mode too. PocketSphinx ASR requires an audio input in .wav file format, and it is transformed into feature sets of MFCC. This MFCC are utilised in a knowledge base to map and create specialised text output, which includes a phonetic thesaurus, learning algorithm, and acoustics models.

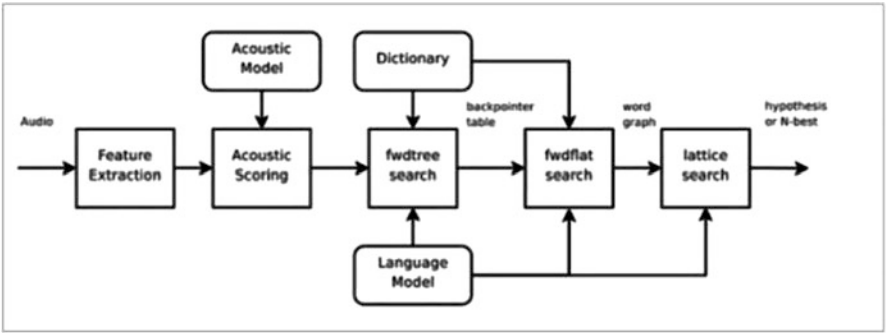


Fig. 1 PocketSphinx [11]

The three primary characteristics of the ASR model are as follows:

A phonetic dictionary in the system provides the semantics mapping terms to patterns of phonemes. A dictionary should have all of the terms you're looking for, or else the classifier will be unable to do so to recognise them. Even putting the terms in a thesaurus is insufficient. The recognition system looks up a word in the lexicon and the learning algorithm. Even if a term is in the dictionary, it will not be detected without the learning algorithm [17] (Fig. 1).

The linguistic model plays a crucial component of the structure since it tells the decoder which word sequences are feasible to identify. There are various varieties of models, each with its own set of capabilities and performance characteristics. Any decoding mode that meets your needs can be chosen, and you can switch between them at any moment [17].

Acoustic model is a hidden Markov framework statistical model. Again, the Gaussian distribution is used to represent the states' HMM output. PocketSphinx offers a variety of acoustic adaption techniques, including maximum a posteriori probability (MAP) and maximum likelihood linear regression (MLLR), which were used for our investigation. We can see that both MAP and MLLR are capable of designing speech systems. Furthermore, as the size of data expands, MAP adaptability outperforms MLLR. MAP is defined in this work as in equation:

$$\mathrm{MAP} {=}_{\theta}{}^{\arg_{\max}} f\ (x|\theta)\ g\ (\theta) \tag{1}$$

Word2vec learns word associations from a vast corpus of text using a neural network model. It can identify synonyms after being educated. Each word is represented by a specific collection of figures known as a vector by Word2vec. Word2vec supports two models [18]. The design of the CBOW predicts the present term from a window of words in the surrounding. The arrangement of the environmental terms has a little impact on prediction. Given the current word, the model in the skip structure estimates the neighbouring window of surrounding terms.

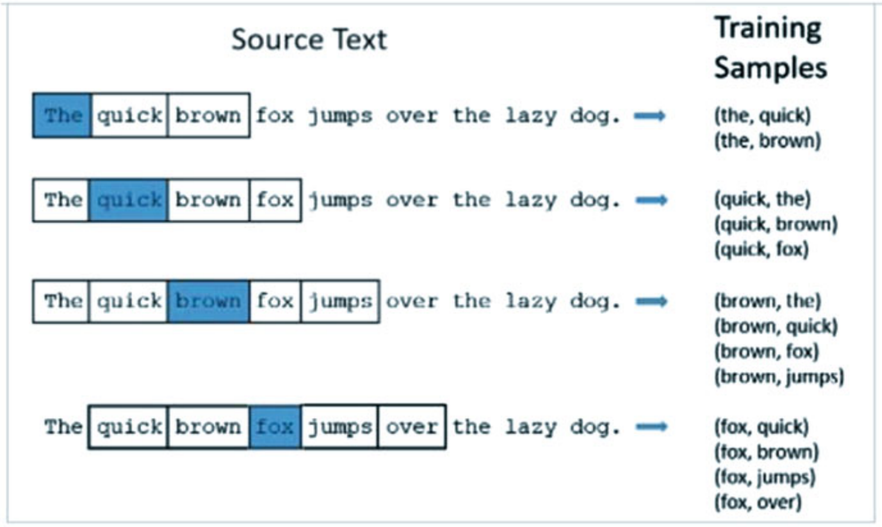


Fig. 2 Word2vec [8]

The fake task: We will train the neural network to look at the words around a certain word in the centre of a phrase (the input word) and choose one at random. The system will provide the likelihood of each term in our lexicon containing the "proximate term" that was selected. This is taught to neural network by feeding it terms from our training content. The picture below depicts some of the phrase-derived training instances (word pairs). The window size is set to 2. The entered word is highlighted in blue [18].

The neural network architecture:

Because it is difficult to feed a word to a neural network as a text string, a method for representing the words to the network is required. To do this, a lexicon of terms from our training manuals is constructed. Assume we have a vocabulary of 10,000 distinct words [18]. A single-word input, such as "books", is represented as a one-hot vector. The above vector will have roughly 20,000 components, with a "1" corresponding to the phrase "books" but a "0" in any other locations. The output of the network is a single vector expressing the chance that a randomly chosen neighbouring word fits each word in our lexicon (Figs. 2 and 3).

Following the Word2vec model, K-means clustering is conducted on the previously generated word vectors, and the assigned values are based on the cluster towards which they relate. By multiplying the values by their closeness to the cluster, the weighted sentiment coefficient is calculated. Each word now has two vectors: one for tf-idf score and one for weighted sentiment ratings. Finally, the tf-idf score is computed, and the detected emotion is determined by identifying the dot product of the two vectors for each utterance. K-means is employed because it scales well to big datasets, will always converge, and adjusts to new examples and

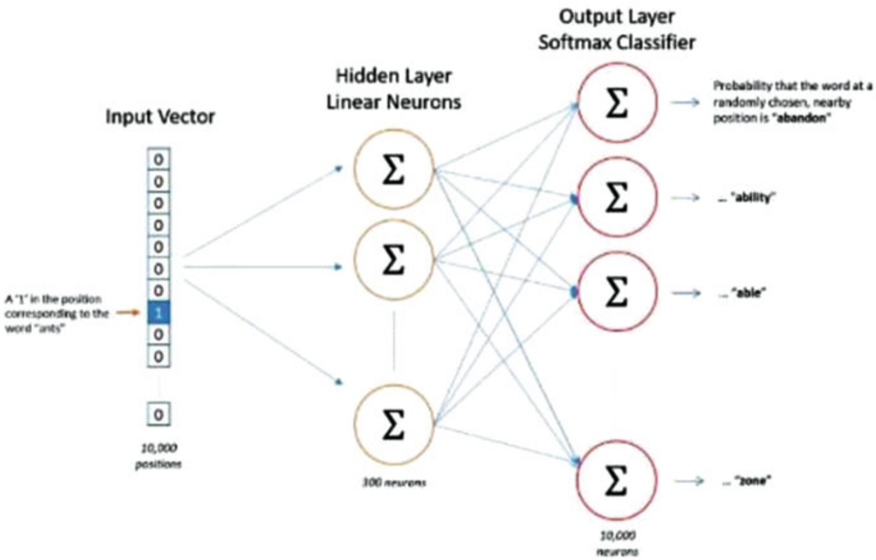


Fig. 3 Neural network [8]

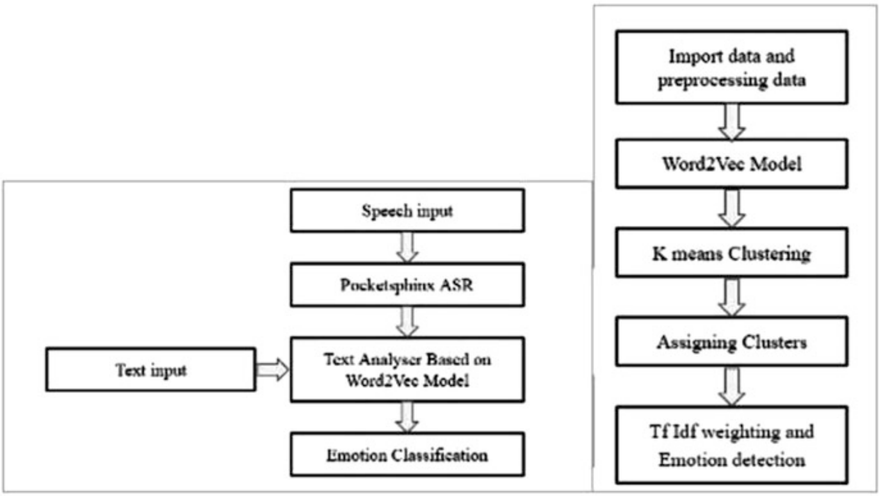


Fig. 4 Proposed architecture

TF. Idf is utilised because it assigns a value to a term depending on its significance in a paper, which is then scaled by its significance through all sources. Figure 4 depicts the suggested model.

4 Dataset

ISEAR is used to evaluate the proposed architecture. The ISEAR (International Survey on Emotion Antecedents and Reactions) has seven basic emotion groups. Anger, contempt, fear, guilt, pleasure, embarrassment, and grief are among them. There is just one categorisation label for each sentence. In the ISEAR project, coordinated by Klaus R. Scherer and Harald Wallbott, a varied collection of psychologists throughout the world provided data.

5 Implementation

The set of data that was utilised in this study was ISEAR. The Word2vec algorithm's gensim implementation with skip gram architecture was used. It was trained with a 6-word lookup window, 20-word negative sampling, 1e-5 subsampling, and a learning rate declining from 0.03 to 0.0007. The K-means method was implemented using sklearn, using 50 repeated beginning points and 1000 rounds of re-assignments of the points to clusters. To get the score, we multiply it by the distance between them and their cluster centres. Following these procedures, a comprehensive thesaurus (as in shape of a pandas DataFrame) was created, for each term given its own score. The tf-idf score of every term in each phrase was then obtained using sklearn's TfidfVectoriser. This was done to examine how each word differed in each phrase. The intersection of these two-word vectors showed whether the overall feeling was positive or negative.

6 Results and Discussion

To anticipate human emotions, we introduced an unsupervised and tailored emotion detection architecture in this work. We are training and evaluating the suggested sentiment recogniser with a combination of ASR and text categorisation, and we are doing so with the ISEAR datasets. Table 1 is a summary of the outcomes. We proposed a tailored emotion detection architecture for emotion categorisation. This algorithm is fed a human phrase as feed, which is subsequently fed into the classification algorithm, which produces the anticipated emotion. Figure 5 shows an

Table 1 Various performance measures

Performance analysis	Model
Accuracy measure	0.80
Precision measure	0.99
Recall measure	0.79
F1 score measure	0.89

```
Enter User text:
 For instance, giving a kiss to your younger sibling daily after waking up
 in the morning and showing him how much you love them. For some
 happiness means loving life and seeing others happy. While some finds
 happiness in writing stories. Some conquer happiness in being
 simple yet the best person they can ever be. Everyone has their own
 unique way to feel happy by finding things that they never expected to find.
 PREDICTION: Happy
Enter User text:
 I am so angry at you!
 PREDICTION: Anger
Enter User text:
 I think i'm gonna be sick
 PREDICTION: Sad
```

Fig. 5 Predicted emotion

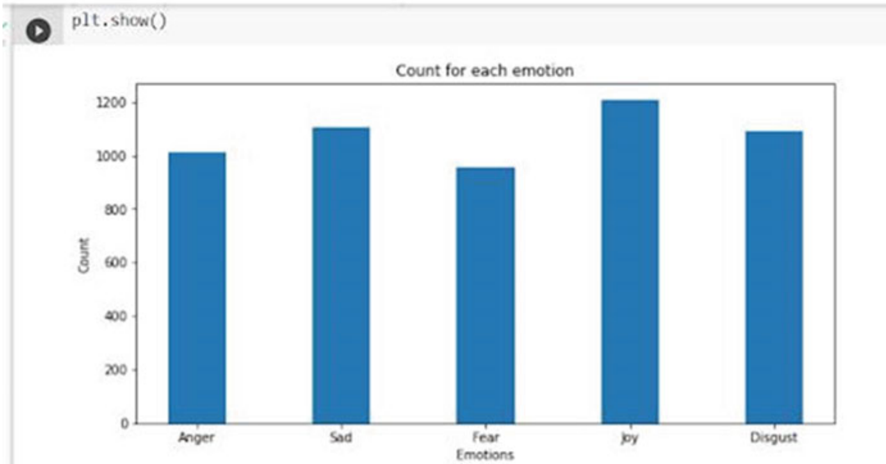


Fig. 6 Count for emotions

example of the obtained result. Figure 6 depicts the number of emotions we were able to achieve. The resulting prediction accuracy using the Word2vec approach was 80%, which is 10% higher than the model provided using the unsupervised feature adaptation scheme [8]. The achieved ac-curacy indicates that Word2vec produces the best results when applied with the appropriate settings as described previously. As it turned out, our model had a precision of 99%, indicating that it was quite effective at distinguishing negative sentiment observations. The algorithm also attained over 80% recall, implying that most of the positive samples were correctly recognised as positive, and an F1 score of 0.88, which is among the best. Based on the findings, we anticipate that when the data source grows more acquainted with user information, the feeling categorisation from text will be expanded to incorporate all of the users' feelings. We want to utilise it in an interactive journal, and it could also be used as a virtual assistant in the future, with the obtained emotions functioning as a driver for the recommendation system, making the suggestion process more particular.

7 Conclusion

This research provides a model for emotion detection that combines ASR and text categorisation. For voice to text conversion, the open-source application PocketSphinx-based ASR and Word2vec-based text analyser were utilised. ISEAR (International Survey on Emotion Antecedents and Reactions) is the dataset used for emotion text analysis since it offers a global collection of emotion data. K-means and TfIdfVectoriser algorithms were employed. Our algorithm takes user text as input and predicts user emotion based on the content. According to the experimental data, the suggested system has an accuracy of (approx.) 80%. The supervised learning technique necessitates manually labelled data, which may be time-consuming and not always practical. Second, for underutilised NLP languages, such as Polish vocabulary, because there are no pretrained models to work with, repositories that have previously trained on a large prediction model are inaccessible. These concerns are addressed by our suggested hybrid algorithm.

8 Future Work

In the future, this algorithm might be integrated into a recommendation tool that uses emotion. Improvements include a third neutral cluster or assigning certain words that end up midway between the positive and negative clusters with a score of 0. Additionally, hyperparameter adjustment of the Word2vec method, depending on, for example, F1 score attained on dataset (albeit this would involve separating the dataset into train and test datasets, since the training would become supervised), may be done to enhance the model's accuracy above 80%.

References

1. Anusha, V., Sandhya, B.: A learning based emotion classifier with semantic text processing. In: Advances in Intelligent Informatics. Springer International Publishing (2015)
2. Binali, H., Wu, C., Potdar, V.: Computational approaches for emotion detection in text. In: 4th IEEE International Conference on Digital Ecosys-tems and Technologies (DEST) (2010)
3. Rachman, F.H., Sarno, R., Fatichah, C.: Corpus-based of emotion for emotion detection in text document. In: 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE) (2016)

4. Aydin, C.R., Güngör, T.: Combination of recursive and recurrent neural networks for aspect-based sentiment analysis using inter-aspect relations. IEEE Access. **8**, 77820–77832 (2020)
5. Iliev, A., Stanchev, P.: Information retrieval and recommendation using emotion from speech signals. In: IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (2018)
6. Cho, K.H., Raiko, T., Ilin, A.: Gaussian-Bernoulli deep Boltz-mann machine. In: The 2013 International Joint Conference on Neural Networks (IJCNN) (2013)
7. Khalil, R.A., Jones, E., Babar, M.I., Jan, T., Zafar, M.H., Alhussain, T.: Speech emotion recognition using deep learning techniques: a review. IEEE Access. **7**, 117327–117345 (2017)
8. McCormick, C.: Word2vec tutorial – the skip-gram model (2016)
9. Nandwani, P., Verma, R.: A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining. **11**(1), 81 (2021)
10. Pantazoglou, F., Papadakis, N., Kladis, G.: Implementation of the generic Greek model for CMU Sphinx speech recognition toolkit (2017)
11. Rahman, T., Busso, C.: A personalized emotion recognition system using an unsupervised feature adaptation scheme. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2012)
12. Ramya, H.R., Bhatt, M.R.: Personalised emotion recognition utilising speech signal and linguistic cues. In: IEEE 11th International Conference on Communication Systems Networks (COMSNETS) (2019)
13. Rashid, U., Iqbal, M.W., Skiandar, M.A., Raiz, M.Q., Naqvi, M.R., Shahzad, S.K.: Emotion detection of contextual text using deep learning. In: IEEE 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (2020)
14. Shaheen, S., El-Hajj, W., Hajj, H., Elbassuoni, S.: Emotion recognition from text based on automatically generated rules. In: IEEE International Conference on Data Mining Workshop (ICDMW) (2014)
15. Kao, E.C.-C., Liu, C.-C., Yang, T.-H., Hsieh, C.-T., Soo, V.-W.: Towards Text-based emotion detection a survey and possible improvements. In: IEEE International Conference on Information Management and Engineering (2009)
16. He, Y., Udochukwu, O.: A rule-based approach to implicit emotion detection in text. In: Natural Language Processing and Information Systems. NLDB 2015. Lecture Notes in Computer Science, vol. 9103. Springer, Cham (2015)
17. Hua, Y., Guo, J., Zhao, H.: Deep belief networks and deep learning. In: IEEE Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things (2015)
18. Mikolov, T., Kombrink, S., Burget, L., Černocký, J., Khudanpur, S.: Extensions of recurrent neural network language model. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2011)

A State of the Art of Non-fungible Tokens: A Literature Review

Joan Cattori Krähenbühl and José Antonio Marmolejo-Saucedo

Abstract Non-fungible tokens, also known as NFTs, are based on the blockchain and are a type of proprietary digital asset registered on the blockchain, making them unique, irreplaceable, and indivisible assets. This main feature makes up the nascent and emerging phenomenon that has revolutionized the way digital assets are and will be handled in the future. Its success and importance has resonated particularly in the first and second quarter of 2022, and however, the full application and reach of NFTs through its technology in human life are still yet to be discovered. The main objective of this chapter is to inform the audience of developments in the research field, discuss the importance and significance of the issues investigated to analyze the relevance and suitability of the topic in order to inform the audience about its operations, and generate, through the NFTs, economical value and further knowledge for future applications. This is the reason why a systematic review of the literature on the subject of "non-fungible tokens" is carried out based on publications of Scopus indexed journals, with a collection of information from 2017 to 2022.

Keywords Non-fungible token · Blockchain · Literature review

1 Introduction

An NFT is a digital asset that represents real-world objects such as art, music, game elements, and videos. They are bought and sold online, with cryptocurrencies such as Ethereum or Solana, and are encrypted with the same underlying software as

J. Cattori Krähenbühl (✉)
Universidad Panamericana, Facultad de Ingeniería, Ciudad de México, Mexico
e-mail: jcattori@up.edu.mx; 0201355@up.edu.mx

J. A. Marmolejo-Saucedo
Facultad de Ingenieria, Universidad Nacional Autonoma de Mexico, Ciudad de Mexico, Mexico
e-mail: jose.marmolejo@fi.unam.edu

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_11

many cryptocurrencies. Although they have been around since 2014, NFTs are currently gaining notoriety because they are becoming an increasingly popular way to buy and sell digital artwork. In 2021 alone, the NFT market was worth a staggering 41 trillion USD, an amount approaching the total value of the entire global fine art market [7]. NFTs are also unique and have individual identification codes. "Essentially, NFTs create digital scarcity," says Arry Yu, chairman of the Cascadia Blockchain Council of the Washington Tech Industry Association and CEO of Yellow Umbrella Ventures. This is in stark contrast to most of the digital creations, which are mostly in endless supply. Hypothetically, cutting off supply should increase the value of a given asset, assuming there is demand. But many NFTs, at least in these early days, have been digital creations that already exist in some form elsewhere, like iconic video clips of NBA games or securitized versions of digital art already floating around Instagram [13].

Famed digital artist Mike Winklemann, better known as "Beeple," created a composite of 5000 daily drawings to create perhaps the most famous NFT of 2021, "EVERY-DAYS: The First 5,000 Days," which sold at Christie's for 69.3 million USD. This piece exhibited in Fig. 1 holds the record for the highest NFTs bought by a single person ever [19].

Anyone can view the individual images or even the entire collage of images online for free. So why are people willing to spend millions on something they could easily capture or download? Because an NFT allows the buyer to own the only original item ever. Not only that, but it also contains built-in authentication, which serves as proof of ownership [11]. Once the economic impact of NFT is understood, it is necessary to change the focus of vision to a research perspective. This is extremely important since the study of this topic gives rise to advances and new innovations.

As previously mentioned, NFTs are an incisive technology, and therefore their areas of knowledge and research lack sufficient reliable sources and publications. To support this statement, Table 1 is shown to exemplify the publications that have been published regarding the research topics of NFTs, Cryptocurrency, and Blockchain.

Table 1 compiles information extracted from the database of the Scopus indexed journal between the years of 2017 and 2022 and is graphically displayed by topic and the number of publications. As can be seen, there is a big difference between the three topics in the graph, since the topic with the most publications is blockchain, followed by cryptocurrency, and in last place NFTs. This is to such an extent that there is practically no column for non-fungible tokens in the table.

Likewise, the interrelationships that exist within the publications with the previously raised topics can be identified as well. Figures 2 and 3 show this correlation under the researched keyword of "Blockchain" and "Cryptocurrency." There are no hints of "non-fungible tokens" in these articles.

The objective of having presented the last three figures is to reveal the opportunities in the NFT research area and how it is imperative to develop investigations and publications to teach the advancement of this topic. Cryptocurrency and blockchain have opened a door of new economic, financial, and digital applications. NFTs follow the same path, but this subject is just beginning.

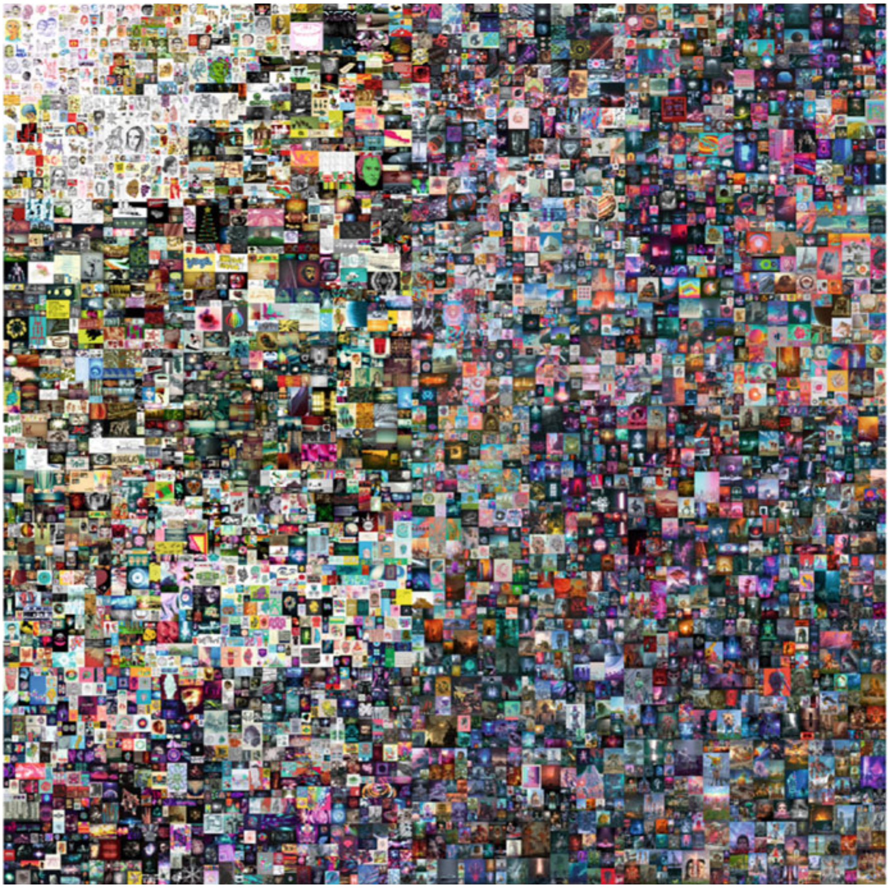
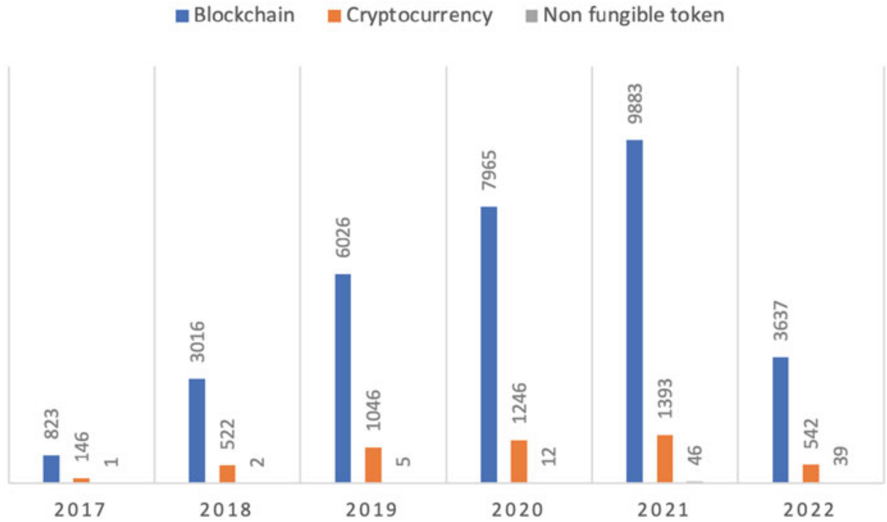


Fig. 1 EVERY-DAYS: the first 5000 day

This literature review's objective is to analyze the state of the art regarding the literature that has been written on the subject of non-fungible tokens, and this is to show the scientific and technological progress and thus encourage the analysis of this topic in later articles. The article will include a study of what has been researched regarding NFTs, their actual need or importance in the research field will be question, and relevant doubts will be presented in order to enrich the research content of NFTs.

Structurally, this research is organized as follows. After the summary and introduction, the following parts are presented in the next order: Methodology and Special Considerations in Sect. 2, Results and finally in Sect. 3, Discussions Around NFTs in Sect. 4, and Conclusions in Sect. 5.

Table 1 Amount of publications per year regarding following researched topics: NFTs, Cryptocurrency, and Blockchain



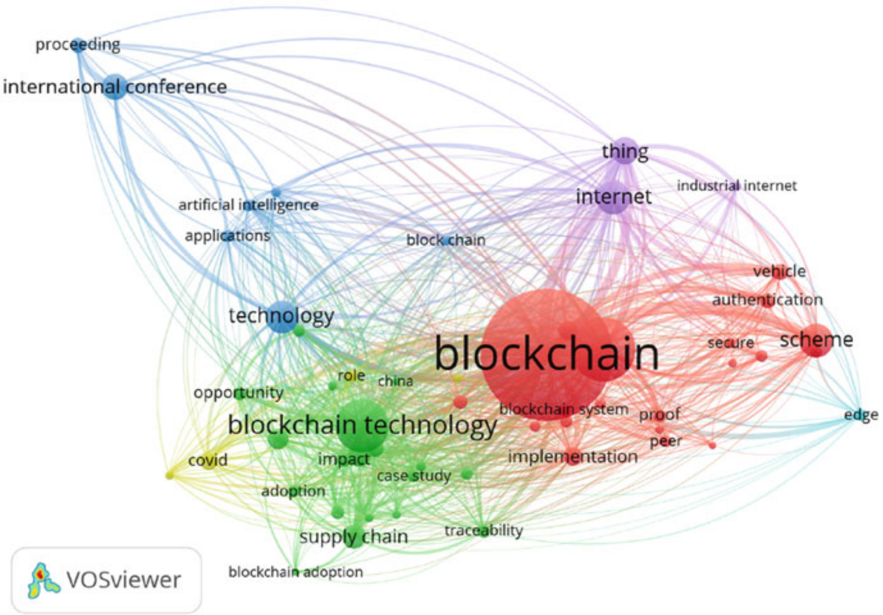


Fig. 2 Keyword's correlation under "Blockchain" using VOSViewer

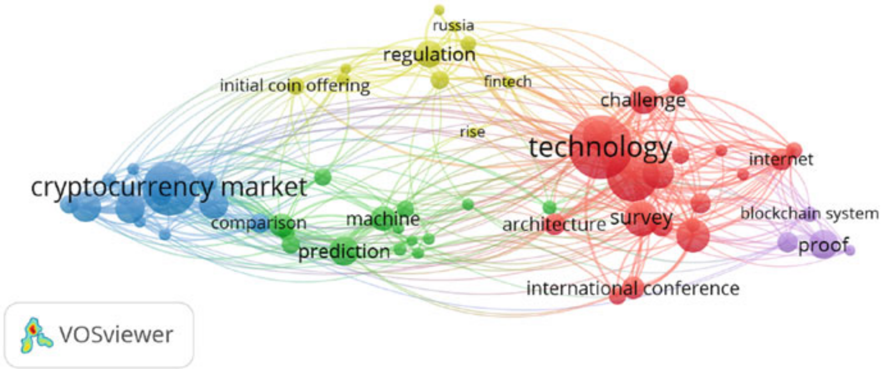


Fig. 3 Keyword's correlation under "Cryptocurrency" using VOSViewer

2 Methodology

Any investigation requires a clear and logical methodology in order to present and generate a process that obtains new results and innovative advances. For this work, the following methodology is used.

The process utilized is divided into two parts. The first is the cleaning and filtering of articles. This stage is vital for the investigation because it guaranties that the information is complete, correct, reliable, important, and truthful. For this, the PRIMSA methodology is used, an acronym for Preferred Reporting Items for Systematic Reviews and Meta-analyses. PRISMA consists of a 27-item checklist and a four-stage flowchart and is intended to directly help authors improve reporting of systematic reviews and meta-analyses. It may also be useful for critical appraisal of up-to-date systematic reviews [18]. Its flow chart will be presented in Chapter 3.

Exactly, this technique is chosen, as the use of checklists such as PRISMA improves the quality of systematic review reporting and provides substantial transparency in the process of selecting articles in a systematic review. Likewise, to complement this stage, the software called "VOSviewer" is utilized. These networks can include, for example, individual journals, researchers, or publications, and can be built on the basis of citation, bibliographic storage, co-citation, or co-authorship relationships. VOSviewer also offers text mining functionality that can be used to build and visualize co-occurrence networks of important terms extracted from a body of scientific literature [10]. The network for this literature review is portrayed in Fig. 4.

Once the database of relevant articles and publications has been obtained, phase 2 begins. This focuses on presenting the data and information collected, that is, displaying the most relevant conclusions based on the articles obtained.

Before continuing with the results of this literature review, it is necessary to establish its limitations and special considerations.

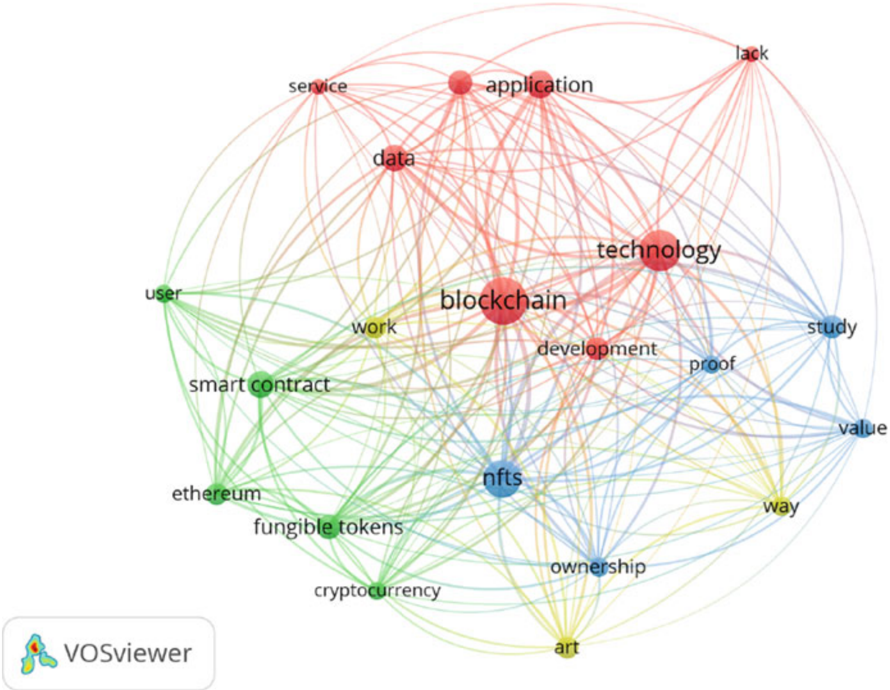


Fig. 4 Non-fungible tokens networks using VOSViewer

(a) The articles and journal publications used in this research are exclusively from the Scopus database under the principal keyword of "non-fungible token" and "NFTs."
(b) The date range was established to accept articles from the year 2017 to May 30, 2022.
(c) All topics within the processed article base will be addressed as long as they are mainly related to NFTs, and this includes fintech topics, innovation, entrepreneurship, cryptocurrency, applications in blockchain, and intellectual property, among others.

3 Results

The mission of this literature review must not be forgotten. As said in the previous sections, NFTs are currently under an early phase of development and excitement due to their vast applications; therefore, the need of further planned research is essential and indispensable. This chapter is focused to synthesize the most important research work regarding NFTs in terms of finance, legality, opportunities and challenges, and cryptocurrencies.

The use of the PRISMA Methodology resulted in a filtration of 105 publications to 10 reconciled papers. Its process can be detailed and identified in Table 2, where the execution and quantity of data depuration is shown step by step. In order to achieve a full understanding of the selected articles, Table 3 summarizes the content that is discussed and analyzed in this investigation.

The chosen papers were classified into four different chapters: these are opportunities presented by NFTs in a digital world, the benefits of integrating NFTs in a portfolio, the correlation between non-fungible tokens and cryptocurrencies, and the legal challenges for NTFs. Nevertheless, some chapters are subsequently divided into subsections.

Table 2 Structured PRISMA flow diagram with step-by-step data deputation

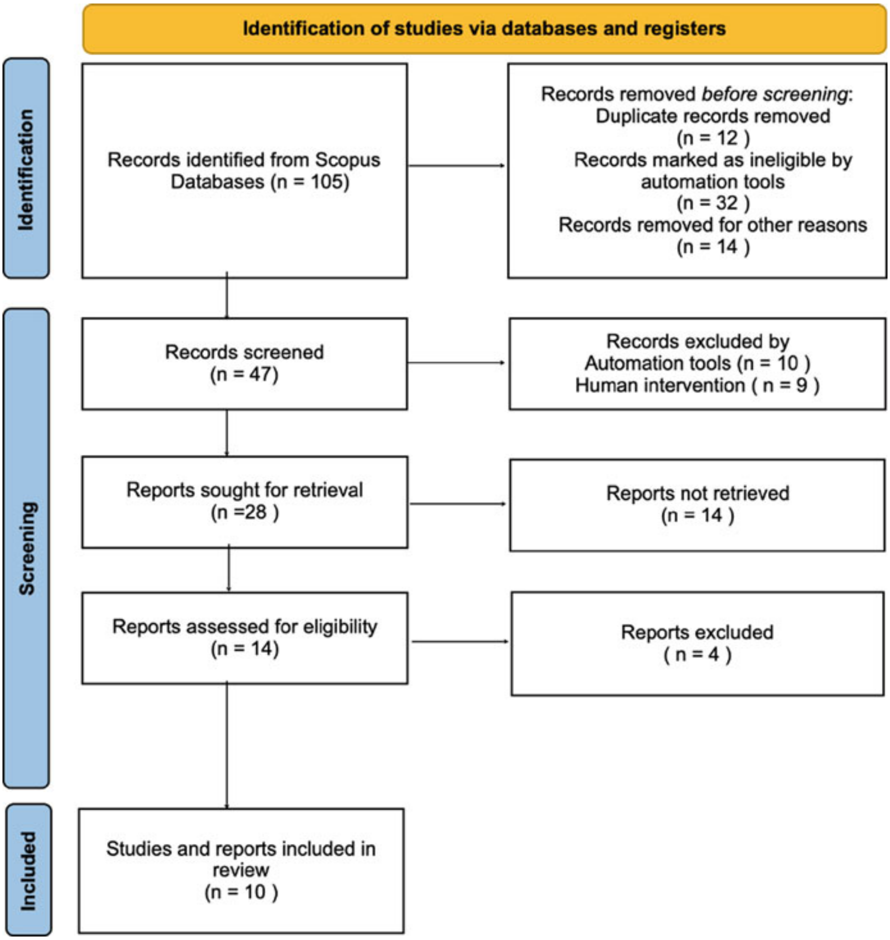


Table 3 Details of the selected literature's summary

Authors	Title	Year	Source title	Methology
Bamakan S.M.H., Nezhadsistani N., Bodaghi O., Qu Q.	Patents and intellectual property assets as non-fungible tokens; key technologies and challenges	2022	Scientific Reports	VAR and Volatility spillover index
Basu S., Basu K., Austin T.H.	Crowdfunding Non-fungible Tokens on the Blockchain	2022	Communications in Computer and Information Science	Crowdsourcing mechanism in blockchain
Chalmers D., Fisch C., Matthews R., Quinn W., Recker J.	Beyond the bubble: Will NFTs and digital proof of ownership empower creative industry entrepreneurs?	2022	Journal of Business Venturing Insights	Empirical research
Dowling M.	Is non-fungible token pricing driven by cryptocurrencies?	2022	Finance Research Letters	Volatility spillover methodology and wavelet coherence analysis
Ko H., Son B., Lee Y., Jang H., Lee J.	The economic value of NFT: Evidence from a portfolio analysis using mean–variance framework	2022	Finance Research Letters	Empirical research
Pinto-Gutiérrez C., Gaitán S., Jaramillo D., Velasquez S.	The NFT Hype: What Draws Attention to Non-Fungible Tokens?	2022	Mathematics	Vector autoregressive models and wavelet coherence analysis
Ali M., Bagui S.	Introduction to NFTs: The Future of Digital Collectibles	2021	International Journal of Advanced Computer Science and Applications	Empirical research
Dowling M.	Fertile LAND: Pricing non-fungible tokens	2021	Finance Research Letters	AVR, AP and DL tests
Mackenzie S., Bērziņa D.	NFTs: Digital things and their criminal lives	2021	Crime, Media, Culture	Empirical research
Nadini M., Alessandretti L., Di Giacinto F., Martino M., Aiello L.M., Baronchelli A.	Mapping the NFT revolution: market trends, trade networks, and visual features	2021	Scientific Reports	Machine learning algorithms

3.1 *Opportunities Presented by NFTs in a Digital World*

Non-fungible tokens are gaining more attention day after day. Although the current situation regarding this topic may be too early and undeveloped, the valued opportunities that NFTs offer can come in handy if handled well. This chapter firstly introduces important terms surrounding NFTs and secondly communicates the advances, opportunities, the future of digital assets, and market trends of NFTs.

3.1.1 The Future of Digital Collectibles: Concepts and Identities of NFTs

In this section, the basics of blockchain and Ethereum are presented. This is to achieve a better understanding of how NFTs work and what their benefits are.

Blockchain Technology To understand how the blockchain technology works, one needs to first understand what is stored in a single block. Each block contains three types of information: the data, its hash, and a pointer to the hash of the previous block. The hash of the block is a unique identifier for the block. The addition of the previous block's hash is the main reason behind the success of blockchain technology and this is also why the alteration of data stored in the blockchain becomes infeasible. If one wants to alter the blockchain by changing the information stored in a block, this will also lead to a change in the hash of the block, which will consequently not match with the hash that is in its next block, causing a chain reaction.

Specifically talking, there are different types of blockchains. In this section, the Ethereum blockchain construction will be explained, for the reason that almost all NFT transactions or transfers are operated with Ethereum tokens.

The blockchain state transitions when there are direct transfers of value and information between accounts or objects. An Ethereum account has four parts:

(a) The nonce: a nonce is a counter set to make sure each transaction can only be processed once.
(b) The current balance of the Ether account
(c) The contract code of the account
(d) The storage of the account, which is empty by default

As it was previously said, in order to pay for most NFTs, Ether is the used cryptocurrency of Ethereum trade. It is used to pay transactional or operational fees [1].

Smart Contracts Smart contracts are computer programs designed to be executed automatically as the people or companies involved in an agreement comply with the clauses of the agreement. They are based on blockchain technology and promise to transform the traditional way of doing business in the not too distant future, eliminating the need to interpret whether a clause has been executed or not, thus making smart contracts deterministic and executed autonomously.

Smart contracts provide stronger and more secure verification for access to the property. A good example is digital security systems for automobiles. The protocol defines the conditions to give authority of the property based on who is the rightful owner at a certain time. Considering cryptographic keys as means to access a car, we can determine conditions that will make the owner change and therefore change the conditions in which the car can be used. A smart lien protocol can be created, described as follows: if the owner fails to make the scheduled payments, the smart contract invokes the lien protocol. This would give back authority of the car keys to the bank. A further step would probably remove the lien in the event the loan gets paid off [1]. A graphical representation of this example can be seen in Fig. 5.

Ethereum Standards As presented above, smart contracts are basically pre-programmed or intelligent programs that get executed when certain conditions are met or certain events happen. Since the process is automated, they have no downtime. They also allow for the creation of decentralized applications. The Ethereum Virtual Machine (EVM), which runs on an Ethereum network, simplifies the process of building blockchain applications. EVM allows the developers of each application to build new blockchains from scratch. And the developers do not have to use any specific programming language. This allows Ethereum to technically build the first decentralized world computer on a public blockchain. The native currency for Ethereum is Ether (ETH). The platform is mainly maintained through transactional or operational fees. One of the most useful advantages of Ethereum is its ability to create unique tokens, with various functionalities, and operate on the Ethereum blockchain. These tokens have multiple utility-like purposes. Ethereum

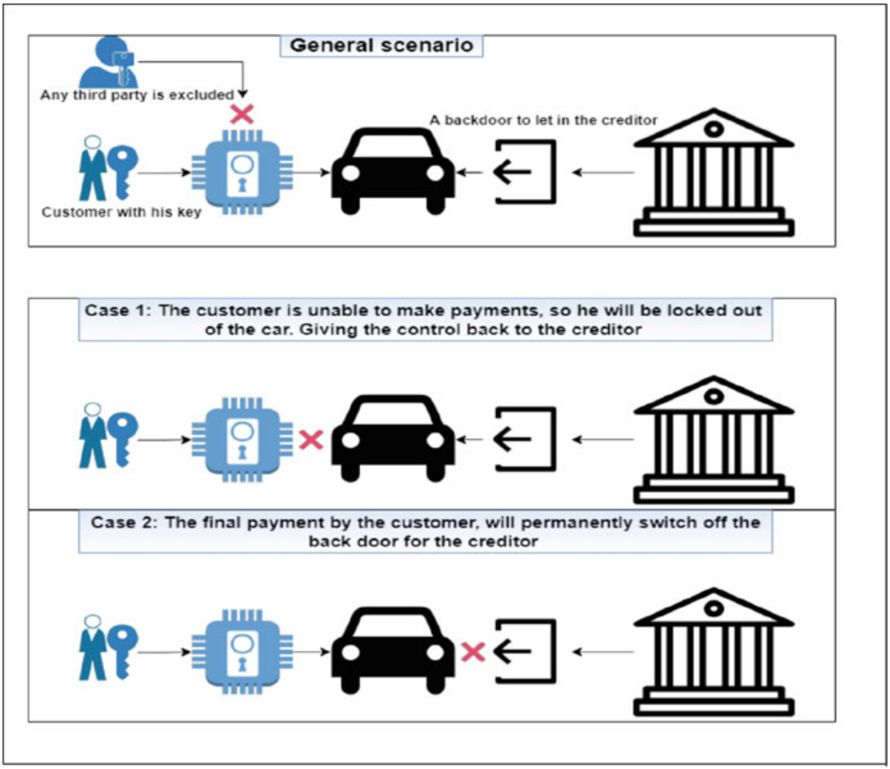


Fig. 5 Digital security system for automobiles that relies on a smart contract

Request for Comments (ERCs) are application-level standards for Ethereum. ERCs include token standards, name registries, library or package formats, and other features. An ERC can be created by anyone, but it falls on the creator to clearly explain the standard. This will also help the creator gain support from within the community. There are common ERC standards for functions of each token type, allowing applications and smart contracts to interact with them in pre-determined or pre-planned ways. ERC-20, which allows for the simple creation, use, and exchange of Ethereum-based tokens, is the most common ERC standard used to date [1].

3.1.2 The Opportunities for Creative Industry Entrepreneurs

This section focuses on presenting the value of NFTs for creative industry entrepreneurs by first inspecting the digital advances of the digital technology, second we study NFTs throughout the emerging Initial Coin Offering boom and bust, and finally we establish a long-term perspective to diagnose if past trends have influenced present NFT economy.

NFT markets are enjoying a notable increase in the digital market, especially as it tries to solve the legitimacy related to the ownership of digital assets. Whether non-fungible tokens actually solve this problem through this new technology remains a question, and the legal status of the assets, the veracity of the data that is delivered to the public, and the generalized fraud are still being questioned. Still, the question of whether the use of NFTs will support entrepreneurs remains to be answered. For this, it must be answered in two parts: short-term and long-term implications. Indeed, entrepreneurs can use NFTs to their advantage. There is currently a profitable market, and as long as this is the case, these people can ride the wave to take advantage of this situation.

The answer about the life of the NFT bubble is uncertain. When will it explode? No one knows, and however NFT entrepreneurs should be prepared. Since what is suggested as true is how ephemeral this bubble is expected to be. This is based on the fact that a large part of the NFTs created are not sold, only a small group. Its application would still be limited and not supported since the legality and protection of NFTs is not yet protected or monitored by the authorities. But as said before, this does not limit artists, musicians, or entrepreneurs to generate income through NFT.

A special case of great interest to entrepreneurs is the issue of crowdfunding to raise capital. Each time it is more complicated to receive financing to grow the business operation of an entrepreneur, that is why with a capitalization campaign of NFTs is possible to obtain funds for growth. This is based on offering NFTs in the form of art in specialized markets for non-fungible tokens, such as Opensea, Rarible, Nifty Gateway, or Superare among others. These NFTs can cost on average between 32 USD and 1,000 USD, in which the seller, that is, the entrepreneur, receives the money to finance his company, while the buyer receives a unique collectible digital asset that may also have an inherent value.

Talking about the long-term implications, the most likely case is anticipated. And it is based on the evolution of the digital technology infrastructure in general and proposes that NFTs will continue to evolve and be replaced by other technological advances. According to the market trend of NFTs for creative entrepreneurs in the industry, a cool down is expected, which would cause a closure of money flow and financing, which would force entrepreneurs to adapt at the technology level or disappear.

It can be said that no matter what there are important signs that there are entrepreneurs who will take advantage of some new technology trend. There is a huge opportunity in the decentralized digital asset market.

Finally, speaking of NFT, two futures are proposed around these assets and entrepreneurs. The first asset is where the exciting bubble of NFTs bursts, thus affirming the hypothesis of passing technology as other technology takes its place or where entrepreneurs find the errors and faults caused by NFTs to propose new technologies and real applications that add value to companies [5].

3.1.3 Crowdfunding Non-fungible Tokens on the Blockchain

As presented in the previous section, the NFT technology presents itself as an alternative to raise capital in exchange of valuable digital assets, not only for entrepreneur but artists as well. In this section, we discuss the details of a fundraising mechanism that will help artists to obtain financial support for their business proposition, while the investors can receive a share of the profits in exchange for their support.

Non-fungible tokens have been one proposed solution for artists to create their work and then sell it online, with ownership tracked on the blockchain to determine who owns the unique copy of the work of art.

The process of this idea is as follows: an artist posts a transaction to the blockchain advertising their project in different digital platforms such as Instagram, YouTube, Twitter, or Discord. Investors, such as NFT collectors or even NFT beginners, can contribute to the project, in exchange obtaining a fraction of the coins from the initial sale. Once the artist then launches the NFT on the blockchain, the funding campaign is tied to the NFT itself. The artist may then sell the NFT as they see fit, and the artist's backers are compensated automatically [3].

Questions and specifications regarding this chapter are shared by the authors of "Crowdfunding Non-fungible Tokens on the Blockchain" detailing the full execution details for their crowdfunding algorithm in the next website: https://github.com/taustin/spartan-gold-nft.

3.2 The Consequence of Integrating NFTs in a Portfolio

NFTs can be classified as digital assets and are likewise financial tools to build financial portfolios, which at the end create a wealthy investment portfolio.

To better understand the need of integrating NFTs in such portfolios, we need first to understand the urge to utilize them.

An investment portfolio is a conglomerate of different investments, that is, a group of securities that together make up the total amount of an investment and its main objective, which pushes investors to build portfolio inflows, are:

(a) Prevent financial resources from being placed in a single investment instrument.
(b) Diversify risk.
(c) Ensure at least a minimal performance.
(d) Use diversification as a risk management mechanism.

A constant improvement caused by NFTs, in any given portfolio, shows a benefit of integrating non-fungible tokens into investment portfolios [4].

Statistical experiments, such as the Gerber Statistic and Person's correlation, conclude the low reciprocity between traditional assets and NFTs. Moreover, in terms of volatility transmission supported by DY and TVP-VAR indices, there is also a divergence between non-fungible and traditional marketplaces. Additionally, the inclusion of adequate quantities of NFTs has had a statistical upturn in the risk-adjusted performance. Proving that, NFTs actually affect attractively the diversification specially EW and tangency portfolio strategies. While the previous conclusions bring a broader and optimized view of an investment strategy while adopting NFTs as a diversifying agent, there is still a demand to evaluate and measure this result's error [12].

3.3 *The Correlation Between NFTs and Cryptocurrencies*

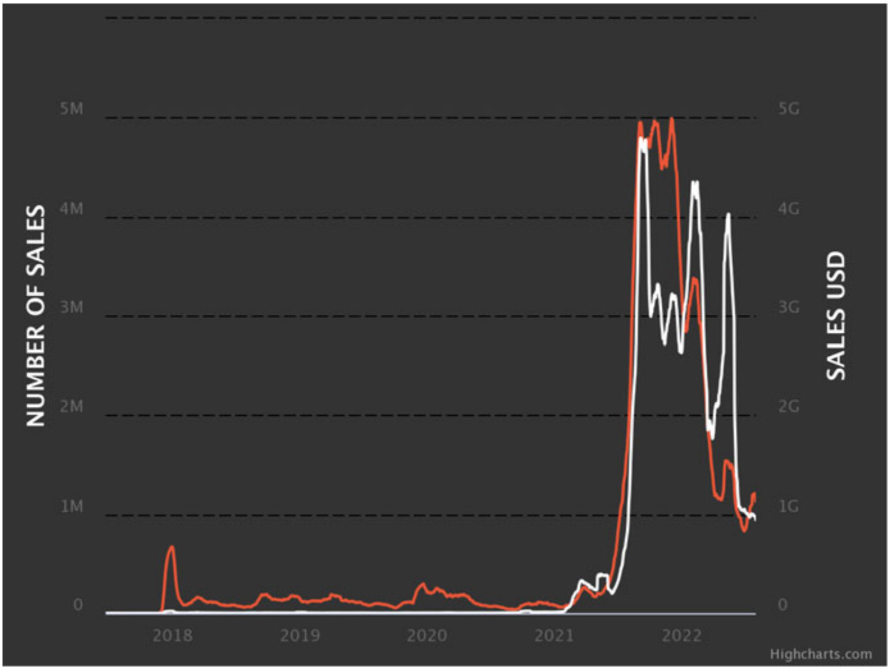
This topic has been commented persistently. In the middle of 2022 with the decline all cryptocurrency's value, NFTs have experienced an interesting deterioration in their prices as well. The relationship between Ethereum and NFTs is most intriguing, due to the fact that Ethereum holds the leadership as transaction coin in the non-fungible token marketplace.

According to Table 4 shows the all-time performance of Ethereum [6]. In the early year of 2020, one ETH had the value of 131.24 USD, one year after the same coin was valued at 726.59 USD, and in the beginning of November 2021 Ethereum reached its maximal value at 4593.47 USD for one coin, a gain of around 3358 times and 524 times regarding 2020 and 2021, respectively. However, Ethereum has faced difficult times, specially between the beginning of April and the mid of June 2022, where the value of one ETH declined to a new record drop to 1016.31 USD.

During this same period of time, considering the devastated decline of Ethereum, the NFT market had a similar phenomenon. Table 5 displays the number of sales in red and the sales in USD in white [16].

Table 4 Ethereum's price behavior from 2016 to 2022

Table 5 Number of sales and sales in USD of NFTs monthly until July 2022



During the same time period, the number in sales of NFT had a mayor drop of 83.3%, while the sales in USD had a 73.8%. In other words, the sales had a 2.97 billion USD drop in this time frame. Its performance is shockingly graphic in Table 5.

In spite of the last drops, significant opportunities still exist to take advantage in such situations.

Wavelet coherence analysis has shown that investors are more interested to NFTs after increments in Bitcoin and Ethereum, which supported the fact that the highest price increase in Bitcoin and Ethereum, recorded in 2021, explains the enthusiasm and boom of NFTs in the same period. Another inquiry comes to light when talking about NFTs and cryptocurrency. Are NFT prices steered by cryptocurrencies? Research shows that NFT and cryptocurrency pricing appear far from each other in terms of volatility transmission, as said before in investment portfolios application. Moreover, wavelet coherence analysis proposes a co-movement between both digital assets, meaning that there is concern and relevance at deepen researching the connection between NFTs and cryptocurrency pricing. Though the pricing relation between NFTs and cryptocurrency may be too blurry to visualize, the pricing within NFTs may be a step closer to identify a clearer path of pricing NFTs and cryptocurrencies [9].

As NFTs collections and classifications can vary from collectibles and art to virtual real estate, the attempt of understanding the pricing of these unique digital assets is vital.

The first attempt to comprehend the pricing of NFTs was by scouting the pricing of virtual real estate in the largest blockchain virtual world with the largest growth rate, called Decentraland, also known as LAND. The findings in this investigation were two. The first one is the inefficiency in pricing, and the second one is an accelerated rise of value, as new growing markets may be driven by manipulation or fraud behavior. As it happened with cryptocurrencies, when its excitement began. Though this conclusions may be insufficient, due to the small number of data registers available in that moment, future studies are limitless as new NFT collection rise with more data [8].

3.4 The Legal Challenges for NTFs

Nowadays, non-fungible tokens are targets for fraud and crime. Lots of people have been victims of malpractices, and this may actually be a response for a new emerging technology. This chapter is focused on portraying the benefits in intellectual property and patents and the risks and disadvantages of NFTs.

3.4.1 Intellectual Property and Patents

The main role of this section is to study the conditions to generate intellectual property assets, specifically patents focused on NFTs. The proposed framework of ideas provides building blocks and guidance so that companies have the necessary tools to use NFT for real-life problems. The challenges for patents created from NFT and their next steps are also stated.

If an NFT is visualized outside their technological applications, we recognize that the solving applications of real-world problems lack. However, NFTs have the ability to develop an acceleration in the process of registering a copyright or a patent, due to its inherently properties marked next:

(a) Ownership plays a crucial role since the owner of the NFTs can easily verify the possession of their asset.
(b) Transferable: the users can freely transfer NFTs and their respective property.
(c) Transparency: the use of blockchain is transparent as it is a decentralized technology and has all transactions available to everyone.
(d) Fraud prevention: during the sale of an asset, fraud is recurrent. By using some NFT, the buyer has the ability to ensure the legitimacy of the asset.
(e) Immutability: any transaction within the blockchain, such as transactions, metadata, token id, is recorded and impossible to change or alter.

Not only do NFTs grant incorruptible security to intellectual property and patents, hence the excitement of companies, governments, or inventors, but they also create a great way to "tokenize" their patents, since the inventor is able to retain royalties in each transaction. NFTs support the intellectual property market by embedding automatic royalty collection methods within inventors' works, providing them with financial benefits each time their innovation is licensed. For example, each inventor's patent would be minted as an NFT, and these NFTs would be combined to form a commercial portfolio of intellectual property and accumulated as a composite NFT. Each investor would automatically get their fair share of royalties as long as license revenue is generated without tracking it.

Currently, there are many ways based on legal and economic literature to solve problems that appear simple. The problem is most of these proposals result in millionaire financial investments for patent and intellectual protection agents. An NFT-based IP and patent system would support the easy problems and the hard ones as well. Well, in this way, the current bottlenecks in patent processes are reduced by making these processes more efficient, fast and convenient, open to everyone.

However, the blockchain-based patent service has not been implemented anywhere in the world. One of the challenges is the problems in the mining part of the blockchain since the miners must have titles from the accepted organizations. Misunderstandings about the proprietary rights of NFTs should also be considered. It is currently unclear when a person purchases an NFT exactly what rights are granted to them, that is, if they have property rights or also moral rights [2].

3.4.2 The Digital Crime in NFTs

The value proposition of NFTs is actually a crime prevention proposition, in consideration of their properties. However, the constant rise of new technologies, news of millionaire profits, social status together with the human need to participate in a technological competition, has proven to be a powerful cocktail in the NFT proposal. Consequently, they have taken advantage and found a way to carry out theft, fraud and other legal errors within the idea of incorruptibility of NFTs and blockchain.

In 2021, the "Evolved Apes" fraud was registered, an NFT project, in which the creator simply disappeared with 2.7 million USD [15]. An example of this NFT collection can be seen in Fig. 6 [17]. Also the case of the Squid token, a type of cryptocurrency, whose value fell sharply after going from being worth 1 cent to 2,800 USD in a week. From there, the value fell and the official Twitter account disappeared, as did its website and Medium account. It is estimated that the developers of the token kept 3.3 million USD [15]. The exploration of the criminal and social life of NFTs indicates that on the premise NFTs can represent a way to promote crime prevention. However, a more realistic view of this idea is that this prevention is only just being developed. A closer connection between seller and buyer between the token would achieve the premise of NFTs.

Fig. 6 EvolvedApe #3757

Without prompt action on the part of the NFT value proposition, it would seem that these assets generate more crime than they actually promote [14].

4 Discussions Around NFTs

NFT's currently proposal has an attractive value proposition perfectly timed during a digital era. However, as presented in the last four sections, the challenges and problems derived from this topic must be considered. This chapter has the purpose to clarify the advantages and disadvantages given previously around NFTs. This is in order to better understand their current situation topicality.

From a legal perspective, NFT's technology, such as smart contract in the blockchain, strongly supports its proposition of crime prevention and registration of copyright and patents. When executed flawlessly, any NFT suppose a perfect prove of ownership, transparency, and immutability. However, this is debated due to the huge evidence in crime and fraud allegations regarding dishonest transactions.

Seen NFTs as a business opportunity is uncertain, speculative, and risky. Leaving the theft section apart, NFTs hold proportional co-relations and share movements with cryptocurrency, creating a bubble easy to forget in the short term, but imminent in the long term, making it a possible ticking bomb.

Their applications such as crowdfunding enrich financial backing for entrepreneurs, but such options offer NFTs without real applications. There are millions of NFTs running through every corner of the Internet, but still the great majority has not been able to implemented successfully as a solution for a real-life problem.

The reality of the NFTs is difficult to understand. NFTs are essentially a tool to propose value in a digital matter, which fits perfectly with the current era, the digital age. Their identity in the eyes of people still makes no sense, an NFT of a digital painting for one million USD is incomprehensible when anyone with Internet access can just screenshot it for free. However, one should not lose sight of what NFTs really are. It is probable for NFTs to pass from a stage of experimentation, which is the one we are currently in, to a stage of generating real value by solving real-world problems.

5 Conclusions

Much has been said in this literature review, and it has been studied and learned the opportunities presented by NFTs in a digital world, the benefits of integrating NFTs in a portfolio, the correlation between non-fungible tokens and cryptocurrencies, and the legal challenges for NTFs. However, the clear contribution of this manuscript can be summarized in the following points: (a) Provide and share previous fundamental knowledge around NFTs. (b) Identify and develop current areas of discussion. (c) Identify and highlight future NFT research and applications into relevant investigation areas. (d) Substantiate the relevance of NFTs in human lives.

Notwithstanding, the need to go further and explore new branches of this topic is imminent. During the investigation, different questions were raised for the following articles.

As said previously, the technology of NFTs has a wide spectrum of applications and can become useful for our life. Naturally problems did arise from it, but if research can find solutions for IP and legitimacy for ownership, lower the crime rates around NFTs among others. Then, the NFT situation can be improved drastically. There is still a plenty of information in terms of NFTs waiting to be read and applied. But more importantly, there is still a plenty of advancements and investigations waiting to be discovered.

Important research opportunity around NFTs is displayed in the following areas: (a) Pricing of NFTs in regards of cryptocurrency. (b) Warnings in NFTs markets and projects. (c) Strategy of crime prevention in NFTs transactions. (d) Integrate NFT technology in today's lives.

References

1. Ali, M., Bagui, S.: Introduction to NFTs: the future of digital collectibles. Int. J. Adv. Comput. Sci. Appl. **12**(10), 50–56 (2021)
2. Bamakan, S.M.H., Nezhadsistani, N., Bodaghi, O., Qu, Q.: Patents and intellectual property assets as non-fungible tokens; key technologies and challenges. Sci. Rep. **12**(1), 1–13 (2022)

3. Basu, S., Basu, K., Austin, T.H.: Crowdfunding non-fungible tokens on the blockchain. In: Silicon Valley Cybersecurity Conference, pp. 109–125. Springer, Berlin (2021)
4. Castro, A.M.: Respuestas rápidas para los financieros. Pearson Educación, London (2002)
5. Chalmers, D., Fisch, C., Matthews, R., Quinn, W., Recker, J.: Beyond the bubble: will NFTs and digital proof of ownership empower creative industry entrepreneurs? J. Business Venturing Insights **17**, e00309 (2022)
6. Crypto.com: Ethereum price (2022). https://crypto.com/price/ethereum
7. Dailey, N.: NFTs ballooned to a $41 billion market in 2021 and are catching up to the total size of the global fine art market (2022). https://markets.businessinsider.com/news/currencies/nft-market-41-billion-nearing-fine-art-market-size-2022-1
8. Dowling, M.: Fertile land: pricing non-fungible tokens. Finance Res. Lett. **44**, 102096 (2022)
9. Dowling, M.: Is non-fungible token pricing driven by cryptocurrencies? Finance Res. Lett. **44**, 102097 (2022)
10. Jan van Eck, N., Waltman, L.: VOSViewer manual (2022). https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.18.pdf
11. Hore, K.: What is an NFT? – Forbes advisor (2022). https://cryptomyfuture.com/what-is-an-nft-forbes-advisor/
12. Ko, H., Son, B., Lee, Y., Jang, H., Lee, J.: The economic value of NFT: Evidence from a portfolio analysis using mean–variance framework. Finance Res. Lett. **47**, 102784 (2022)
13. Leggatt, J.: What is an NFT? non-fungible token explained (2022). https://www.forbes.com/advisor/au/investing/cryptocurrency/nft-non-fungible-token/
14. Mackenzie, S., Bērziņa, D.: NFTs: digital things and their criminal lives. Crime. Media. Culture. **18**, 17416590211039797 (2021)
15. Micucci, M.: Estafas más comunes con nft y cómo protegerse (2022). https://www.welivesecurity.com/la-es/2022/03/17/estafas-mas-comunes-con-nft-y-como-protegerse/
16. NonFungible.com: Number of sales · sales USD · all · monthly (2022). https://nonfungible.com/market-tracker
17. OpenSea.com: Evolved Apes Inc (2021). https://opensea.io/assets/ethereum/0x233a65b06ef83ccf2fd58832086dd782f9da1642/3757
18. Page, M.J. et al.: The Prisma 2020 statement: an updated guideline for reporting systematic reviews (2021). https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-021-01626-4
19. Sensorium.com: 10 most expensive NFTs ever sold (2022). https://sensoriumxr.com/articles/most-expensive-nft-sales

Part IV
Health 4.0 and Pervasive Health

A Camera-Based Remote Sensor for Physical Therapy in Parkinson's Disease

Jorge L. Rojas-Arce, José Antonio Marmolejo-Saucedo, and Luis Jimenez-Angeles

Abstract The present chapter shows a camera-based remote sensor for monitoring postural disorders. In recent years, several systems have been developed to capture human motion in real time using common RGB cameras. We analyze the performance of the sensor for the detection of basic body parts under a different exercise routine clinically proposed for Parkinson's disease patients. In this study, we evaluate the ability of the OpenPose algorithm to detect specific keypoints of anatomical landmarks and estimate rotation, rate, and amplitude for head, shoulders, back, and pelvis. Based on our findings, we can say that OpenPose is a sufficiently robust algorithm that is capable of detecting slight motion parameters of a specific rehabilitation therapy, which could facilitate their use in a home environment specially in follow-up and management of the motor rehabilitation therapy for Parkinson's disease.

Keywords Remote sensor · Physical therapy · Parkinson's disease

1 Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative diseases in adults, resulting from the loss of dopaminergic cells in the substantia nigra of the midbrain and the presence of cytoplasmic inclusions known as Lewy bodies [3, 18]. In Mexico, a prevalence has been estimated between 40 and 50 cases per 100,000 people, with a mean age of onset of 56.7 ± 13.4 years old, and

J. L. Rojas-Arce · L. Jimenez-Angeles (✉)
Department of Biomedical Systems, Engineering Faculty, Universidad Nacional Auonoma de Mexico, Mexico City, Mexico
e-mail: jorge.rojas.arce@comunidad.unam.mx; luis.jimenez@comunidad.unam.mx

J. A. Marmolejo-Saucedo
Facultad de Ingenieria, Universidad Nacional Autonoma de Mexico, Ciudad de Mexico, Mexico
e-mail: jose.marmolejo@fi.unam.edu

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_12

the mean duration of the disease until death is 6.9–14.3 years and increases the possibility of mortality by 50% [8, 17]. Treatment of PD has focused mainly on pharmacological management; however, it is also widely known that physical therapy is a fundamental part of preserving the functionality and independence of the patient according to the degree of evolution of the condition. Rehabilitation therapy focuses on cognitive movement and exercise strategies to optimize the patient's independence, safety, and well-being, thereby improving quality of life. A conventional physical therapy program that combines strength, resistance, and flexibility exercises, as well as balance training, is effective in reducing the degree of motor disability [5, 7, 10]. Telerehabilitation, or remote physical therapy, is one of the most common types of complex distance medicine that is applied in practice [19]. During recent years, a large number of motion capture (MoCap) systems detecting the pose of a human using a "markerless" approach have emerged [6], and these systems work without the necessity of placement of any markers on the human body [4]. MoCap system approaches reduce the technical and financial requirements and complexity of arrangement [16], and therefore it can be found in the context of distance medicine, not only in specialized clinics but also in the home environment [1]. Considering the application of distance medicine in the home environment, the most promising systems seem to be systems for the evaluation of body movements from commonly used standard video (RGB) records [2]. These systems have great potential for use mainly due to the reduced financial costs of purchasing these systems. These systems have reached such technical levels that they could be used in specific cases as alternatives to costly systems in clinics. However, these systems must also use special camera data processing software [15]. The most commonly used software tools for pose detection are OpenPose [12], Mask R-CNN [11], Google's Media Pipe [11], Alpha-Pose [9]; all of them available as open source. All these software tools use neural networks trained on annotated images [13]. The datasets contain general static images of people in undefined positions, according to which NN learns to recognize anatomical points on the body. Existing benchmarks compare the speed and accuracy of detection using the above algorithms based on NN [13]. A shortcoming that limits wider use of low-cost MoCap systems and the mentioned software is the absence of evaluation of the validity of the data provided by these systems. This raises doubts about the use of telemedicine where it is necessary to know relatively accurately the information about the movement of specific anatomical points that physiotherapists need to monitor and modulate the rehabilitation intervention [14]. Thus, the main aim of this work is to determine the feasibility of a camera-based remote sensor to be used for home rehabilitation in Parkinson's disease subjects. In a control group of 15 volunteers, we focus on the evaluation of head and shoulder movement in a specific therapy for cervical spine and pelvis and back for a lumbar spine therapy. Thus, our study aims to evaluate the feasibility of a remote camera system and thus applicable to home telerehabilitation of PD.

2 Methodology

2.1 Design

The advantage of the camera-based approach is the ability to detect motion from any regular RGB video, and however specific considerations (like space, background, clothes, luminosity, field of view, image resolution, etc.) must be considered in a video capture protocol before the video acquisition. The design of our study is shown in Fig. 1.

2.2 Video Capture Protocol

For the acquisition of the evaluation videos, the following recommendations were made to the volunteers:

– Select a large place and remove most of the obstacles to prevent falls during the evolution of the exercises.
– Wear clothes that allow you to stand out from the rest of the objects in the background, that is, select a color contrary to the background, for example: if the background is white, wear black clothes.

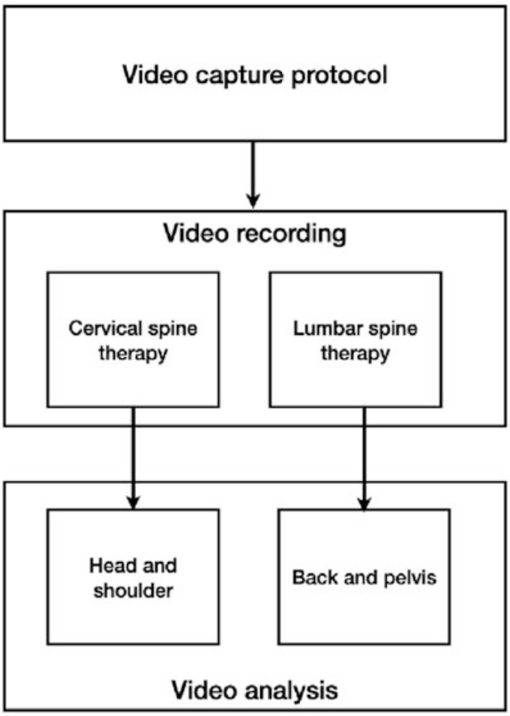


Fig. 1 Schematic representation of the study

- The lighting must be directed to the front of the person, in such a way that it illuminates their face and does not create a shadow. The shadow is generated when the light source is behind the person in the video or the video capture device.
- There should be no other person in the area, and large photos or portraits of people (60 cm × 70 cm) should be avoided.
- Use a support for the video capture device. If this is not possible, keep the device static for as long as possible and avoid sudden movements.
- The floor must be parallel to the horizontal axis of the video capture image.
- In the video capture, use a resolution of 480p, 720p, or 1080p. The format of the capture space is indistinct (16:9,4:3,11:9, etc.) as long as it complies with the capture type, and the capture types are indicated later.

Each video was manually checked to confirm that the entire trainee's body was in view throughout the recording. At the same time, manual categorization into specific groups was done by a single rater. Border cases were excluded from the analysis.

2.3 Physical Therapy

A physical therapy was designed by medical specialist in rehabilitation in order to evaluate specific angles related with the preservation of joint mobility and gross motor coordination. These angles were quantified after video recording by a small camera embedded into a micro-controller board (OpenMV Cam H7 plus, OpenMV) with capabilities to run a neural network and identify 17 points of the body (shoulders, hips, elbows, wrists, knees, and ankles). Table 1 shows the target evaluated while the subject performs a specific therapy.

2.4 Video Analysis

In the context of performing rehabilitation therapy, the body segments that are part of the appendicular skeleton are very often measured [26]. These segments allow for translational movement of the body targets through repeated cyclic movements as were described in Sect. 2.3. Video recordings of 15 normal control volunteers were analyzed, and angles from keypoints detected by OpenPose algorithm were computed to infer the rotation, rate, and amplitude of movement of head, shoulder, back, and pelvis. Table 2 shows a schematic representation of the computed angles while a subject performs a specific therapy.

Motion parameters (rotation, rate, and amplitude) of the head, shoulders, back, and pelvis were computed for 15 control volunteers after performing the therapies for cervical and lumbar spine, respectively, and were defined as follows:

- Rotation of the head was the angular change of a horizontal imaginary plane formed by keypoints detected for mouth. Maximum negative rotation values were

Table 1 Relationship between the specific target evaluated and therapy

Target	Therapy
Cervical spine	– Left lateral tilt of only the head
	– Right lateral tilt of only the head
Lumbar spine	– Hands on the waist and lateral flexion of the trunk
	– Arm contra-lateral to flexion extended to the sides of the trunk above the head
Shoulder mobility	– Raising the hand above the head with the left elbow extended – Raising the hand above the head with the right elbow extended
Combined	– Sit down and get up from a chair, without arm support
	– Hand to contralateral shoulder
	– Hand to contralateral ear, ipsilateral
	– Hand to shoulder and contralateral elbow
	– Right arm lift and left arm separation
	– Crossed hands in front of the neck
	– Three steps forward, three steps back, alternate arms, last with arm elevation

computed when the head had a maximum tilt to the left, and maximum positives values were computed for maximum tilt to the right.

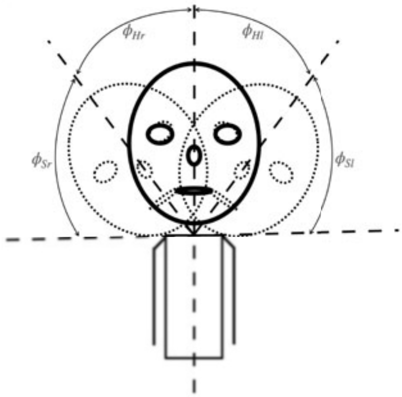
- Rotation of shoulders was the angular change of an imaginary horizontal plane formed by detected keypoints for left and right shoulders.
- Rotation of back was the angular change of the trunk (a perpendicular imaginary plane that starts in the midpoint between the femur greater trochanters and ends in the midpoint between shoulders).
- Rotation of pelvis was the angular change of an imaginary horizontal plane of left and right trochanters.
- In all cases, the rate was defined as the number of cycles per minute computed from rotation signal.
- In all cases, the amplitude is the absolute difference between consecutive minimum and maximum values of the rotation signal.

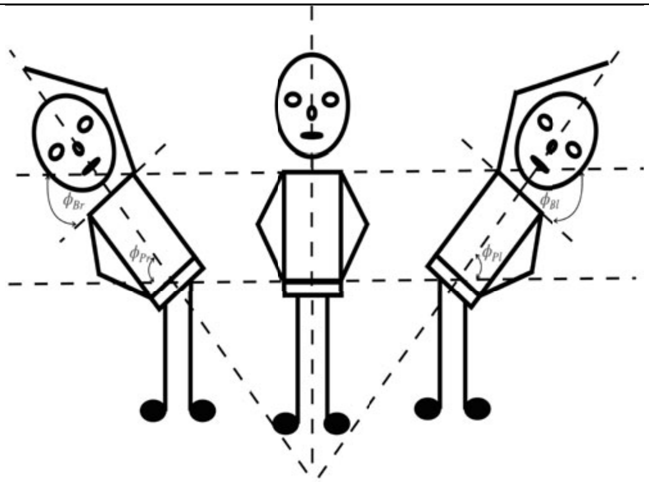
3 Results and Discussions

In this work, we analyzed videos from 15 control volunteers in order to identify the sensibility of OpenPose software to infer the motion parameters of a proposed specific motor rehabilitation therapy for Parkinson's disease. Each video contains at least 10 repetitions of each exercise, and we present the preliminary findings of a camera-based remote sensor for physical therapy in Parkinson's disease.

Table 3 depicts the keypoints computed by OpenPose software for a control volunteer during a specific therapy. Angles for head, shoulder, back, and pelvis are

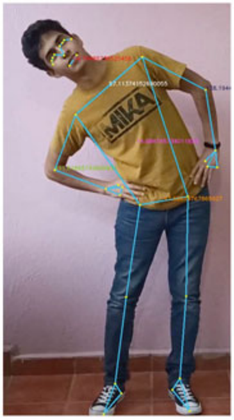
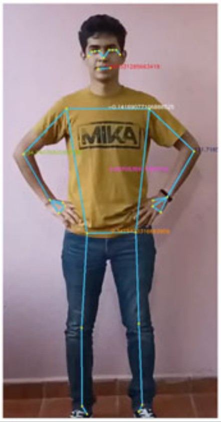
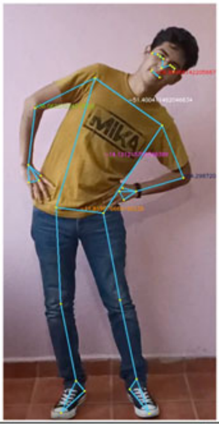
Table 2 Identification and nomenclature of the angles used in continuous monitoring

<table>
<tr><th>Computed angles for cervical spine</th></tr>
<tr><td>

ϕ_{Hr} = Angle of the head rotation during right lateral tilt.
ϕ_{Sr} = Angle of the shoulder rotation during right lateral tilt.
ϕ_{Hl} = Angle of the head rotation during left lateral tilt.
ϕ_{Sl} = Angle of the shoulder rotation during left lateral tilt.</td></tr>
<tr><th>Computed angles for lumbar spine</th></tr>
<tr><td>

ϕ_{Br} = Angle of the back rotation during right lateral tilt.
ϕ_{Pr} = Angle of the pelvis rotation during right lateral tilt.
ϕ_{Bl} = Angle of the back rotation during left lateral tilt.
ϕ_{Pl} = Angle of the pelvis rotation during left lateral tilt.</td></tr>
</table>

Table 3 Keypoints and angles detected during the physical therapy for a control volunteer

Target	Keypoints and angles		
Cervical spine			
Lumbar spine			
			

displayed overlaid on the image (see Table 2 for reference and nomenclature). As can be observed, OpenPose software has a high sensitivity to identify and follow 11 keypoints of face and 17 keypoints of trunk and lower and upper extremities during the proposed therapies. To achieve these results, it is mandatory to follow the video capture protocol as was described in Sect. 2.2.

In the first row of Table 3, corresponding to cervical spine target therapy, we can observe in the central column the reference (0° of head rotation) and the maximum rotation in the first and third columns where a slight difference between left and right keypoints for shoulders can be also observed, despite the indication for this therapy is that the subject only rotates the head. This observation explains the signal for shoulders rotation in Fig. 2.

In the second and third rows of Table 3, corresponding to lumbar spine target therapy, we can observe in the central column the reference (0° of rotation for pelvis and back) and the maximum rotation in the first and third columns. Supposed plans for shoulders, back, and pelvis are more evident and how it change according to the therapy.

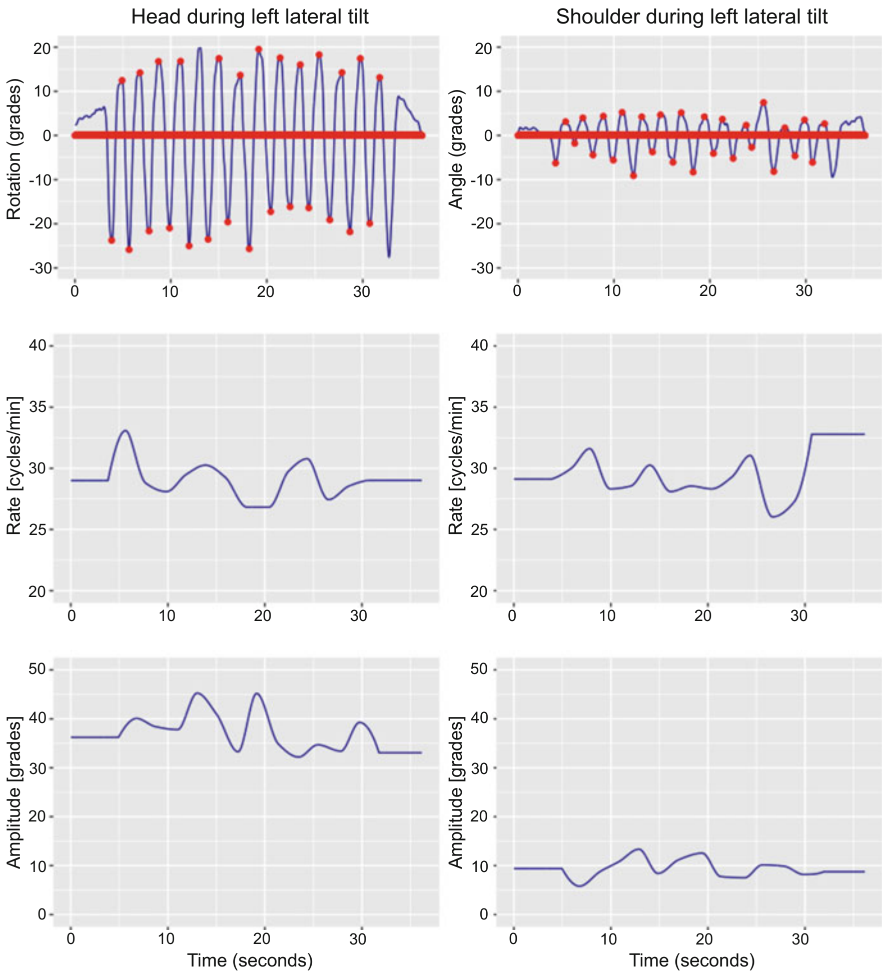


Fig. 2 Rotation, rate, and amplitude for head and shoulder while a normal volunteer exercises his/her cervical spine target

Figure 2 depicts rotation, rate, and amplitude for head and shoulders while a control volunteer exercises their cervical spine target. As was explained previously, although the therapy indicates that the subject only turns his head, we can observe that shoulders have a similar pattern for rate although a marked difference in amplitude. These results confirm the high sensitivity of OpenPose software to detect and follow slight changes in the motion parameters. Special care must be taken when it is required to quantify the speed and amplitude of the movement at the beginning or at the end of the exercise, since, being of lower amplitude, there are limitations to identify the maximum and minimum peaks, so these parameters can be overestimated.

Figure 3 depicts rotation, rate, and amplitude for pelvis and back while a control volunteer exercises his/her lumbar spine target. As can be observed in the second and third rows of Table 2, pelvis and back form the bases of a trapezoid that rotates in a similar rate and amplitude as was expected for this kind of therapy.

Table 4 summarizes the motion parameters for head and shoulders for 15 control volunteers after performing the cervical spine therapy. Parameters for head motion are similar when it rotates to left or right. All subjects have similar values for maximal rotation (range) for left or right tilt, centered around zero degrees (median value). Control volunteers had very similar values for head rate (45.97 cycles/min vs 48.49 cycles/min) for left and right tilt, respectively, and similar values for amplitude rotation (33.06° vs 38.21°), respectively.

A similar behavior was quantified for shoulders, with a less amplitude of rotation compared with head. These quantitative results confirm the high sensitivity of OpenPose software to detect and follow slight changes in the motion parameters.

Table 5 summarizes the motion parameters of back and pelvis for 15 control volunteers after perform the lumbar spine therapy. Parameters for back motion are similar when it rotates to left or right. All subjects have similar values for maximal rotation (range) for left or right tilt, centered around zero degrees (median value). Control volunteers had very similar values for back rate (22.66 cycles/min vs 17.58 cycles/min) for left and right tilt, respectively, and similar values for amplitude rotation (33.11° vs 37.53°), respectively.

A similar behavior was quantified for pelvis, with similar rate but a less amplitude of rotation compared with back.

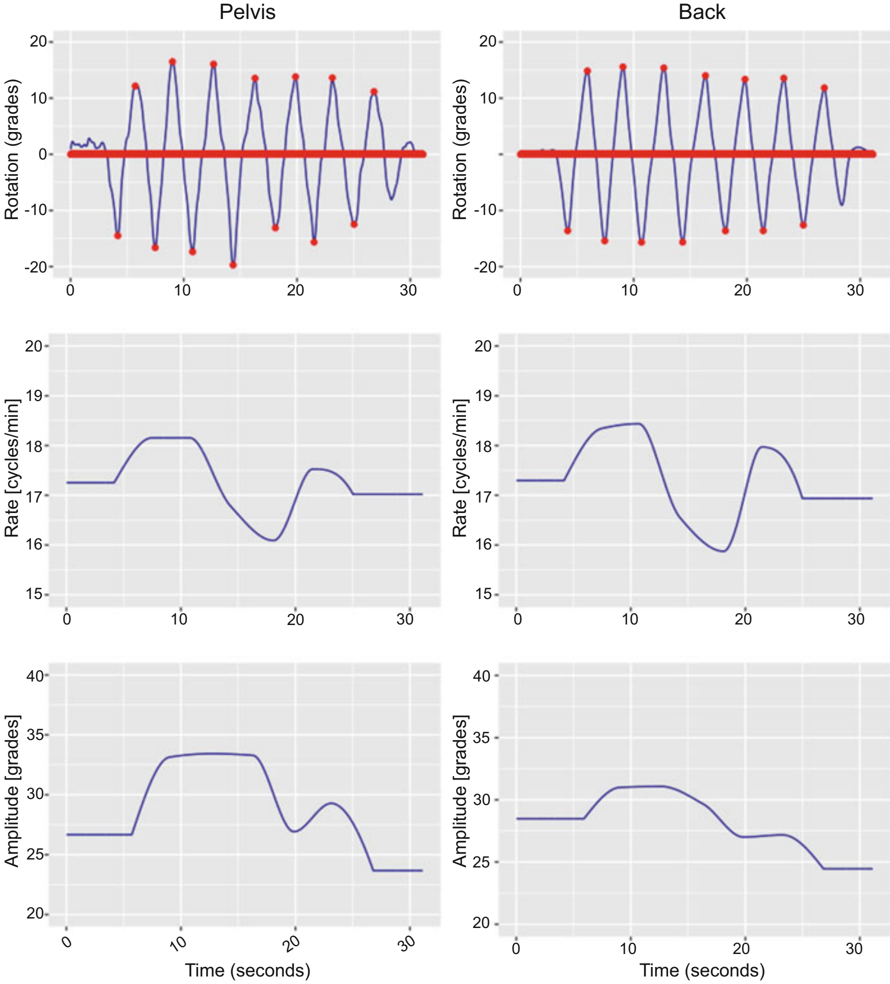


Fig. 3 Rotation, rate, and amplitude for pelvis and back while a control volunteer exercises his/her lumbar spine target

Table 4 Head and shoulder parameters of movement during cervical spine therapy

		Left tilt	Right tilt
Head			
Rotation (grades)	Range	−44.12 to 29.85	−54.02 to 35.25
	1st quartile	−11.44	−11.70
	Median	0.52	1.02
	3rd quartile	9.68	10.77
Rate (cycles/min)	Range	22.69–90.77	23.29–95.68
	1st quartile	30.21	40.52
	Median	45.97	48.49
	3rd quartile	53.08	57.98
Amplitude (grades)	Range	20.62–58.20	17.07–63.08
	1st quartile	28.23	28.18
	Median	33.06	38.21
	3rd quartile	43.79	45.92
Shoulder			
Rotation (grades)	Range	−18.44 to 12.015	−15.04 to 15.36
	1st quartile	1.50	−2.18
	Median	0.05	0.04
	3rd quartile	0.60	2.24
Rate (cycles/min)	Range	17.02–136.15	22.99–95.68
	1st quartile	29.13	40.36
	Median	44.40	47.23
	3rd quartile	51.13	59.39
Amplitude (grades)	Range	0.85–22.55	0.63–25.49
	1st quartile	3.05	4.47
	Median	5.18	8.21
	3rd quartile	11.89	12.85

Table 5 Back and pelvis parameters of movement during lumbar spine therapy

		Hands on the waist and lateral flexion of the trunk	Arm contra-lateral to flexion extended to the sides of trunk
Back			
Rotation (grades)	Range	−56.58 to 50.02	−50.91 to 54.65
	1st quartile	−6.60	−6.53
	Median	−0.14	−0.05
	3rd quartile	6.84	5.86
Rate (cycles/min)	Range	13.46–41.16	8.46–31.05
	1st quartile	17.53	14.87
	Median	22.66	17.58
	3rd quartile	25.07	19.84
Amplitude (grades)	Range	12.87–93.12	21.56–101.04
	1st quartile	24.45	29.11
	Median	33.11	37.53
	3rd quartile	60.01	65.70
Pelvis			
Rotation (grades)	Range	−27.05 to 38.77	−24.69 to 27.64
	1st quartile	−3.99	−2.50
	Median	0.01	0.24
	3rd quartile	3.99	2.52
Rate (cycles/min)	Range	13.83–41.16	8.36–86.34
	1st quartile	17.38	17.46
	Median	22.84	19.66
	3rd quartile	25.29	26.17
Amplitude (grades)	Range	5.82–51.43	1.74–48.36
	1st quartile	13.11	7.63
	Median	22.23	12.14
	3rd quartile	26.66	16.94

4 Conclusion

There have been many recent publications describing the possibility of using a camera-based system for home rehabilitation. In this chapter, we propose a rehabilitation therapy for cervical and lumbar spine and identify keypoints using OpenPose software to estimate motion parameters (rotation, rate, and amplitude of movement) for head, shoulders, back, and pelvis in 15 control volunteers videos of people performing rehabilitation exercises in front of a camera. Results shown that, in this control population, motion parameters are similar in amplitude and rate regardless of whether the exercise therapy is to the right or to the left. Based on our findings, we can say that OpenPose is a sufficiently robust algorithm that is capable of detecting slight motion parameters of a specific rehabilitation therapy, which could facilitate their use in a home environment specially in follow-up and management of the motor rehabilitation therapy for Parkinson's disease.

Acknowledgment Supported by UNAM-DGAPA-PAPIIT TA100921.

References

1. Adolf, J., Dolezal, J., Kutilek, P., Hourova, M., Hejda, J., Milerska, I., Lhotska, L.: Automatic telerehabilitation system in a home environment using computer vision. Studies Health Technol. Informat. **273**, 142–148 (2020). https://doi.org/10.3233/shti200629. http://europepmc.org/abstract/MED/33087604.
2. Airò Farulla, G., Pianu, D., Cempini, M., Cortese, M., Russo, L.O., Indaco, M., Nerino, R., Chimienti, A., Oddo, C.M., Vitiello, N.: Vision-based pose estimation for robot-mediated hand telerehabilitation. Sensors **16**(2) (2016). https://doi.org/10.3390/s16020208. https://www.mdpi.com/1424-8220/16/2/208
3. Bonnet, A.M., Houeto, J.L.: Pathophysiology of Parkinson's disease. Biomed. Pharmacoth. **53**(3), 117–121 (4 1999). https://doi.org/10.1016/S0753-3322(99)80076-6
4. Cotton, R.J.: Kinematic tracking of rehabilitation patients with markerless pose estimation fused with wearable inertial sensors. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), pp. 508–514 (2020). https://doi.org/10.1109/FG47880.2020.00092
5. Crizzle, A.M., Newhouse, I.J.: Is physical exercise beneficial for persons with Parkinson's disease? Clin. J. Sport Med. **16**(5) (2006), https://journals.lww.com/cjsportsmed/Fulltext/2006/09000/Is_Physical_Exercise_Beneficial_for_Persons_with.9.aspx
6. Desmarais, Y., Mottet, D., Slangen, P., Montesinos, P.: A review of 3D human pose estimation algorithms for markerless motion capture. Comput. Vision Image Understand. **212**, 103275 (2021). https://doi.org/10.1016/j.cviu.2021.103275. https://www.sciencedirect.com/science/article/pii/S1077314221001193
7. Dibble, L.E., Addison, O., Papa, E.: The effects of exercise on balance in persons with Parkinson's disease: a systematic review across the disability spectrum. J. Neurol. Phys. Therapy **33**(1) (2009). https://journals.lww.com/jnpt/Fulltext/2009/03000/The_Effects_of_Exercise_on_Balance_in_Persons_with.3.aspx
8. Estrada-Bellmann, I., Camara-Lemarroy, C.R., Calderon-Hernandez, H.J., Rocha-Anaya, J.J., Villareal-Velazquez, H.J.: Non-motor symptoms and quality of life in patients with Parkinson's disease in Northeastern Mexico. Acta Neurol. Belgica **116**(2), 157–161 (2016). https://doi.org/10.1007/s13760-015-0544-7
9. Fang, H.S., Xie, S., Tai, Y.W., Lu, C.: RMPE: Regional multi-person pose estimation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2334–2343 (2017)
10. Gobbi, L.T.B., Oliveira-Ferreira, M.D.T., Caetano, M.J.D., Lirani-Silva, E., Barbieri, F.A., Stella, F., Gobbi, S.: Exercise programs improve mobility and balance in people with Parkinson's disease. Parkinsonism Relat. Disord. **15**, S49–S52 (2009). https://doi.org/https://doi.org/10.1016/S1353-8020(09)70780-1. https://www.sciencedirect.com/science/article/pii/S1353802009707801
11. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. IEEE Trans. Pattern Analy. Mach. Intell. **42**(2), 386–397 (2020). https://doi.org/10.1109/TPAMI.2018.2844175
12. Hidalgo, G., Cao, Z., Simon, T., Wei, S.E., Joo, H., Sheikh, Y.: OpenPose library (2017). CMU-Perceptual-Computing-Lab. https://github.com/CMU-Perceptual-Computing-Lab
13. Kidziński, U., Yang, B., Hicks, J.L., Rajagopal, A., Delp, S.L., Schwartz, M.H.: Deep neural networks enable quantitative movement analysis using single-camera videos. Nat. Commun. **11**(1), 4054 (2020). https://doi.org/10.1038/s41467-020-17807-z

14. Marotta, N., Demeco, A., Moggio, L., Ammendolia, A.: Why is telerehabilitation necessary? A pre-post COVID-19 comparative study of ICF activity and participation. J. Enabling Technol. **15**(2), 117–121 (1 2021). https://doi.org/10.1108/JET-11-2020-0047
15. Ota, M., Tateuchi, H., Hashiguchi, T., Kato, T., Ogino, Y., Yamagata, M., Ichihashi, N.: Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. Gait Posture **80**, 62–67 (2020). https://doi.org/https://doi.org/10.1016/j.gaitpost.2020.05.027. https://www.sciencedirect.com/science/article/pii/S0966636220301776
16. Peretti, A., Amenta, F., Tayebati, S.K., Nittari, G., Mahdi, S.S.: Telerehabilitation: review of the state-of-the-art and areas of application. JMIR Rehabil. Assist. Technol. **4**(2), e7 (2017). https://doi.org/10.2196/rehab.7511. http://rehab.jmir.org/2017/2/e7/. http://www.ncbi.nlm.nih.gov/pubmed/28733271
17. Ringman, J.M., Diaz-Olavarrieta, C., Rodriguez, Y., Fairbanks, L., Cummings, J.L.: The prevalence and correlates of neuropsychiatric symptoms in a population with Parkinson's disease in Mexico. Cogn. Behav. Neurol. **15**(2) (2002). https://journals.lww.com/cogbehavneurol/Fulltext/2002/06000/The_Prevalence_and_Correlates_of_Neuropsychiatric.5.aspx
18. Rinne, J.O.: Nigral degeneration in Parkinson's disease. Movement Disorders **8**(S1), S31–S35 (1993). https://doi.org/10.1002/mds.870080507. https://movementdisorders.onlinelibrary.wiley.com/doi/abs/10.1002/mds.870080507
19. Suso-Martí, L., La Touche, R., Herranz-Gómez, A., Angulo-Díaz-Parreño, S., Paris-Alemany, A., Cuenca-Martínez, F.: Effectiveness of telerehabilitation in physical therapist practice: an umbrella and mapping review with meta–meta-analysis. Phys. Therapy **101**(5), pzab075 (2021). https://doi.org/10.1093/ptj/pzab075

Health 4.0, Prevention, and Health Promotion in Companies: A Systematic Literature Review

Sergio Arturo Domínguez-Miranda and Román Rodríguez-Aguilar

Abstract Noncommunicable diseases are growing worldwide and their impact within organizations affects the productivity of companies. The accelerated pace of life, sedentary lifestyle, eating habits, and lack of self-regulation have deteriorated workers' health conditions. The Health 4.0 paradigm can help in health prevention and promotion thanks to the use of smart devices and digital tools adaptable to users and companies. Trials from 11 bibliographic databases were consulted and out of a total of 742 articles, 86 were selected that met the selection criteria. There is scientific evidence that supports the use of smart devices in companies focusing on weight control, physical activity, sleep control, and glycemic index to impact the treatment and prevention of noncommunicable diseases such as diabetes, overweight, work stress, cardiovascular diseases, and in lifestyle. Using wearables or smartphones, incentive programs or assistance with specialists have been considered by some researchers; elements such as privacy and information security are essential in the implementation, as well as methods that can maintain the use of these prevention and health promotion programs. More research is necessary regarding the use of smart devices such as the permanence of health initiatives in companies, cost-effectiveness, and real-time analysis, and focus on various pathological conditions for success in prevention and health promotion strategies.

Keywords Health 4.0 · Wearables · mHealth · Pervasive health · Noncommunicable diseases · Physical activity · Workplace · Corporate culture

S. A. Domínguez-Miranda (✉) · R. Rodríguez-Aguilar
Facultad de Ciencias Económicas y Empresariales, Universidad Panamericana, Mexico City, Mexico
e-mail: 0246533@up.edu.mx

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_13

1 Introduction

Noncommunicable diseases (NCDs) are the leading causes of the global public health crisis of the present and the future, and they are treatable by medical interventions, health promotion, and prevention measures. The impact of this health issue can be significantly reduced by better social conditions, a higher standard of eating, maintaining a healthy weight, quitting smoking, and increasing physical activity (PA).

Even though the problem is increasing, one solution seems to be highly applicable in health care, with a relevant impact on the transformation of medicine, preventively, predictively, and participatively. It is called Health 4.0, which is a global concept oriented toward enhancing the health care experience leading to successfully improving the quality, flexibility, productivity, and dependability of health care services through the Internet of Health Things, wearables, the health Cloud, big data analytics, machine learning, blockchain, and smart algorithms, which are all integrated and used to support people and specialists to increase life expectancy [1]. Part of the development and unification of Health 4.0 has led to the construction of a great diversity of biomedical and portable elements that detect and show vital signs, temperature, pulse, among others, which reflect relevant statistics to maintain a healthy life.

Although the prevention and promotion of health depends on how people face the problem based on the culture of each of them, the increase in NCDs, sedentary activities, incorrect diet, genetic factors among others, had caused many companies, aware of the negative results of not taking care of employees, to seek diverse strategies to act together. It could be that strong steps in situational improvement can be taken through the artifacts and solutions of Health 4.0.

However, it is difficult to find in various companies' policies or schemes focused on worker health, which are usually understood as work initiatives or desires to make improvements without having a regulatory framework that can help to understand the effectiveness of the company's investments focused on the prevention of NCDs. Additionally, although there is literature in which the effectiveness of Health 4.0 in individuals is demonstrated, the use of this concept within companies is very disaggregated and studies focused on the benefit to workers are limited.

This article shows the most up-to-date information in the field of Health 4.0 concepts and the main elements to analyze such as the use of smart devices, digital solutions, and mobile health care (mHealth) whose use is focused on the labor field. The chapter is structured as follows. In the first section we observe the methodology used for the systematic literature review; in the second section the main findings of the 86 selected articles are shown derived from the methodology shown regarding Health 4.0, the generalities of digital health, as well as the operation of wearables and their use in NCDs, in addition to the diversity of studies analyzed that demonstrate the effectiveness of using smart devices in the labor field; in a third section, the discussions of the articles as well as future lines of research are presented.

2 Background

2.1 Noncommunicable Diseases

Noncommunicable diseases are one of the most serious causes of the increase in the number of deaths worldwide and account for approximately 70% of all deaths [2]. NCDs are defined as chronic diseases that include heart disease, stroke, cancer, chronic respiratory diseases, and diabetes related to lifestyle developed through unhealthy behaviors. This can lead to consequences for people's disability, reduction in quality of life, decrease in productivity in society, and an increased burden of health care costs [3].

In recent years, there has been a need for a growing and sustainable health system that manages not only critical care primarily in emergencies but also outpatient care, especially those with chronic illnesses. In addition, economically speaking, it costs enormous amounts of money for the medical care of patients, which is why agencies and governments are looking for a way to reduce their impact; as the global spending on chronic diseases has increased, immediate action is now required [4]. Owing to the large population base, only conventional medical institutions can currently offer systematic health testing services for chronic diseases; however, numerous researchers and organizations are looking for new ways to intervene because the current medical resources fall far short of meeting people's medical needs [5]. Therefore, there is an increasing demand for ongoing patient health monitoring to allay the anxiety of solving this growing demand.

As the trend in the aging of the population increases and the pace of life of workers accelerates, the incidence of chronic diseases and the proportion of people who are less healthy also increase year after year, given the prevalence of diseases. In addition to social and economic considerations, dietary habits are linked to chronic problems, such as a hypercaloric diet, sugary drinks, and saturated fats, that lead to high levels of triglycerides and glycemia, abdominal obesity, excessive alcohol consumption, smoking, and a sedentary lifestyle, which are key risk factors that increase the risk of acquiring the metabolic syndrome that includes diabetes, obesity, and hypertension. These problems lead to serious risk factors for heart disease and stroke and, at the same time, significantly reduce people's quality of life [6]. Thus, the greatest contribution to this situation is linked mostly to unhealthy lifestyles. In order to improve public health, PA levels must be increased [7] because metabolic syndrome is less common in those who are regularly active, less sedentary, and have superior cardiopulmonary functions [8]. In Table 1 a summary of information on the main NCDs, which illustrates the holistic problem in society, is shown.

Therefore, identifying innovative and attractive intervention strategies to promote health and focus on risk factor reduction in life-threatening diseases and chronic comorbidities through various tactics that bring key information from individuals to health specialists is of paramount importance.

Table 1 Noncommunicable diseases: relevant facts

Disease	Relevant information	Key statistics
Diabetes	Chronic and complex disease that requires continuous medical attention. A reduction in the impact of this disease that goes beyond glycemic control is required [9]	In 2017 [10] it was estimated that, worldwide, approximately 425 million adults, aged between 20 and 79, were living with diabetes. By 2045, this number is expected to rise to 629 million
Obesity	One of the biggest public health concerns of the twenty-first century is the global health issue, which is also the reason for the rise in morbidity and death [11]	Body weight loss (5–10%) and BMI reductions lower the health risks linked to chronic diseases [12]. Promotion of PA and change of lifestyle is crucial [13]
Hypertension	One of the key risk factors for all-cause mortality and the main factor in cardiovascular morbidity, death, and disability. It requires timely prescriptions and lifestyle measures [14]	High blood pressure affects 1 billion people, and it is projected that by 2025, its prevalence will be higher than 1.5 billion people in the world [15]
Cardiovascular diseases	The standard vital indicator for detecting changes in heart cycles is the heart rate [16]	An elevated risk of coronary artery disease and heart failure is associated with a high resting heart rate [17]. Most of these conditions' risk factors, including high blood pressure, an abnormal lipid profile, smoking, and physical inactivity, should be avoided [5]
Stroke	Ranks as one of the leading causes of death and disability worldwide. Several limitations as accessibility to clinical examinations may be relevant owing to geographical or resource restrictions [18]	80% of strokes can be prevented by attending to the risk factors [19] such as high blood pressure

2.2 *Prevention and Promotion*

The accelerated pace of life, the rapid population expansion, and the rising cost of health care present significant challenges to the existing state of health. This calls for faster, more inexpensive care options, especially as the world's older population rises. The importance of many chronic and acute diseases is rising, and the medical sector is undergoing a sea change as a result of the demand for real-time diagnosis and monitoring of long-term health conditions. Some correlations were made with respect to smart health in patients and deduced that weight gain in patients is due to an increase in calorie consumption, few or no minutes of PA, a reduction in the number of steps and floors, little exercise, little or no water intake, and a decrease in calories burned, and is likely to lead to an increase in heart rate [20].

The health model, therefore, focuses on both health promotion and disease prevention, and lifestyle habits have a great impact on the quality of human health. An effective framework focused on health prevention and promotion helps us to define the most appropriate forms of care for the health problems we face and can be used to visualize the kinds of mechanisms and strategies that are needed to support and encourage employees in their quest to live fuller, healthier lives.

In this sense, there are various strategies that can help to prevent and promote health, such as PA, nutrition, preventive check-ups, and constant monitoring of health indicators, which contribute to prolonging life and improving its quality, through physiological, psychological, and social benefits. Regular PA, including any bodily movement that requires energy, and a balanced diet, can reduce the risk of many NCDs and disorders, such as hypertension, coronary heart disease, stroke, diabetes, breast and colon cancer, and diabetes. On the other hand, maintaining reviews on a regular basis can prevent problems and major expenses, as the treatment of diseases at an early stage significantly reduces major interventions.

Several strategies have been used within the companies, including increasing knowledge, greater dissemination in different visual points, activation campaigns, approaches with specialized doctors, behavioral and social policies. Goal setting, social support, reinforcement through self-reward, and self-control have been identified to be important examples of behavioral and social techniques employed in prevention and promotion, according to some researchers [13].

In addition, looking for health monitoring has been a challenge to address as innovative methods are required to continuously measure relevant indicators that mark alerts or advances in health. In this regard, high-tech wearable technology offers a simple and appealing method of self-monitoring several parameters. Similarly, improvements in smartphone applications make it easy to combine many behavioral change strategies to encourage healthy lifestyle choices [21].

There is an important characteristic that has been found in health care in the framework of health and refers to how these elements are of great importance to contemplate in the future of health; those premises are reflected based on the principle of 7P [22]: Personalized, Persuasive, Predictive, Participatory, Preventive, Perpetual, Programmable. The expectation is that the smart health system could provide unique treatments and techniques, be able to predict disease with the participation of different stakeholders, prevent issues with patients, and be a continuous system.

How to enable continuous monitoring while the patient is out of the hospital or clinic and engaging in regular life activities in their home surroundings presents a significant difficulty. Also, how can we facilitate small, low-cost devices for continuous monitoring of health, mental, and activity status 24 h a day and anywhere?

2.3 Health 4.0

Although health and its actors can be understood as a complex and difficult system to solve, it is likely that we can find a solution in the creation of new technologies and innovative processes that result in differentiated elements in favor of health and collaborators.

Health 4.0 is a strategic concept derived from Industry 4.0 and its central focus lies in the progressive virtualization of the health system in favor of more personalized attention, including the digitalization of health, which is the convergence of the digital and genomic revolution, with health, health systems, our way of life and a new conception of society [23], is citizen-centered, and uses sophisticated analytics to extract knowledge from the data obtained.

Recent advances in microtechnology, data processing, wireless communication, and battery capacity have led to the proliferation of non-invasive wearable devices that integrate with the user's smartphone and can be used to measure multiple health-related signals in a free-environment life [24]. The implementation of technologies such as virtual reality, artificial intelligence, robots, and many more, allows work activity to be safer and healthier because through this an early and constant detection of possible risks can be carried out, in addition to also allowing their management through intelligent security technologies and virtual engineering. Through the Health 4.0 concept, real-time monitoring of patients is possible by connecting different sensors to measure body temperature, blood pressure, and cardiac response so that the analysis of the data can reach a device and issue an alert when any parameter is outside acceptable limits [25].

It may be considered that a new era of people-driven health has arrived, with tremendous future benefits in the prevention, diagnosis, and management of NCDs. As society advances and people's standard of living rises, people pay greater attention to quality of life. Technological innovations continue to be deeply rooted in everyday life and consumers are beginning to use consumer-grade software and hardware devices to monitor their health such as wearable devices and health mobile applications that come to light as relevant elements within the concept of Health 4.0 that help to improve health systems through prevention and promotion.

3 Methodology

The present study consists of a systematic review of the literature. Studies were collected using computer-based searches in a set of specialized databases considering as research products: book chapters, conference proceedings, papers, and theses.

A set of keywords was defined according to the objective of the research considering the variations in both Spanish and English, as shown in Table 2, to make a situational diagnosis in the work environment and its correlation with performance. The search was prioritized in these four levels, which had a focus

Table 2 Keywords used in the different search criteria

Level	Sorter	Terminology
1	Wearables in Health	Population Health Management Prevention in Health Health 4.0 Wearables mHealth
2	Approach	Diabetes Cardiovascular Disease control Chronical disease
3	Enterprises	Health investment Health and productivity Health policies in companies Environment, health, and safety
4	Economic evaluation	Economics of occupational health Cost-effectiveness preventive policies Economic evaluation of prevention strategies

on the use of portable mobile devices based on the framework of Health 4.0, considering the different words alluding to that term such as sensors, wearables, or mobile monitors. The second qualifying level would be focused on the use of these devices, and as the research is focused on the prevention and control of NCDs, the terms of prevention and control of chronic diabetes and cardiovascular diseases were sought. The first two levels of the classification were prioritized as being the most relevant, as all articles that talk about the use of portable technological tools in the control and prevention of diseases were evaluated in a first phase.

Then, in a next classification level, which was used to classify but not to eliminate possibilities of analysis of articles, constructs focused in companies were contemplated, reviewing the articles in which these technological elements have been contemplated within the business or organizational initiatives, as well as the qualifying level of economic evaluation, to review which of the articles found discussion of economic evaluation or cost-benefit analysis or effectiveness (Table 2).

In the first phase, where a search was carried out with the keywords using Boolean operators "And", "Or," and "Not". To ensure that we obtain the greatest diversity of results, this search was carried out on four occasions on different days with 16 combinations used identifying 21,300 options; a second filtering was carried out with greater Boolean indicators, selection of articles, books, and abstracts of conferences from 2010 to 2022, with more relevant documents being offered by the search algorithm of the selected platforms, and the total was reduced to 2577 articles.

Within the first filtering phase, words were removed, communicable diseases and orthopedics (as well as their variations), because they were not considered necessary or they did not generate input to this review, as the focus is on NCDs and mostly

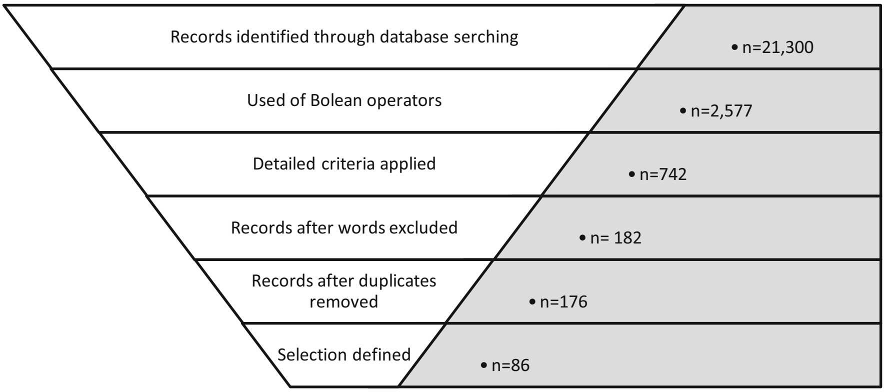


Fig. 1 The study selection process

on metabolic and cardiovascular syndrome, both in English and in Spanish. It was found that the use of the word company, industry or company is not necessarily useful when showing information about mobile device development companies, so to evaluate whether it is part of the state-of-the-art approach, verifications were made in the following phases. However, one of the tools used was to eliminate those articles focused on adolescents, children, or older adults for not being within the work focus group. The decision was also made to eliminate those articles focused on oncological issues, as the intervention in these conditions is different from that of cardiovascular diseases or diabetes control, then a filtering a total of 742 articles.

In the second phase, the names of the articles and documents found were read. In this phase, those articles whose approach was based on pain control, which spoke of disabilities, mental health, which was focused on ergonomics, which was a general document focused only on digital health, which was dedicated purely to clinical diagnosis or within hospitals, which was dedicated to research and development, were eliminated. that he had a focus on self-tracking or that his approach was on accident prevention. This time the options were reduced to 182 items. Repeated articles were deleted, resulting in 176 articles.

Along the third phase, the abstracts of the articles were scrutinized, being analyzed its objective, methodology and results, being selected those that made sense with the object of our investigation, reaching a filtering of 86 articles. Figure 1 shows the study selection process.

4 Results

Out of the 86 selected articles, we found that 1 was a book chapter, 3 were conference proceedings, and 82 corresponded to articles that showed us that the main research focus in the concept of Health 4.0 is referred to in research articles.

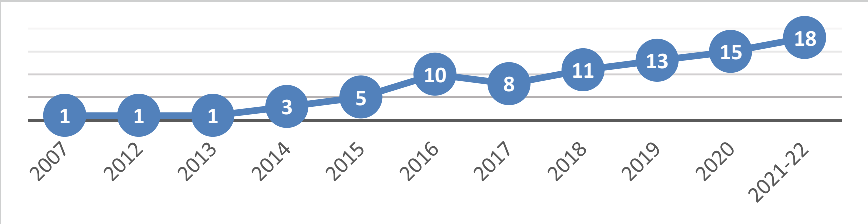


Fig. 2 Selected articles per year – Health 4.0

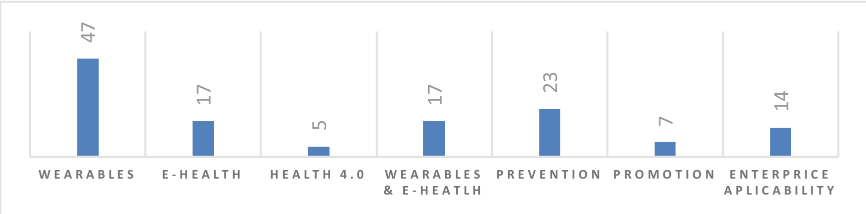


Fig. 3 Topics related in selected articles

Thirty-six percent (31 articles) focused on literature reviews, 26.7% (23 articles) had a qualitative approach, both documentary and exploratory, 33.7% (29 articles) had a quantitative approach, and 3.6% (3) had a mixed (qualitative and quantitative) approach.

Throughout the years of writing the articles, it was possible to verify the constant growth in the approach to the concept of Health 4.0 wearables and mHealth, which is a new concept and that, just as it has shown growth in technological progress and its applicability reflected in Fig. 2. It is clarified that the year 2022 was added to 2021 as, being at the beginning of this year when the literature search was carried out, the graph could suffer alterations by not contemplating the entire year running.

The literature selected has been cited an average of 59.47 times over the period established, of which 23 articles have less than five citations, it is argued that some of them are from 2022 (four articles) so they are newly published. Also they are in languages other than English or Spanish (6 articles) and in other cases they are either books or doctoral theses (4 articles), which also supports the argument that most researchers are focused on research articles. Seventeen articles have been referred to more than 100 times, of which 3 stand out, with more than 300 references.

Another relevant finding is the focus of the articles, both concepts, analyses, studies, or propositional elements: 5 of them discuss Health 4.0, 17 discuss e-Health, 47 wearables, 17 a mixture of wearables with e-Health, 23 prevention, 7 health promotion, and 14 showed a specific focus on application within companies, as shown in Fig. 3.

4.1 mHealth and Wearables

The technological advances that have occurred in recent decades have contributed to the development of increasingly smaller and more agile processors, whereas wireless sensor network technologies with low battery consumption become viable, having a wide range of devices that encompass applications for entertainment, health, education, assistive technologies, among others [26]. As a result, the development of new technical gadgets has been included in the monitoring and treatment of diseases.

Among the various applications of these devices, patient monitoring has been the subject of research around health. According to the World Health Organization [27], digital technologies are becoming an important resource for health services delivery and the public. Despite the potentially wide applicability of digital health strategies and solutions to address the diversity of patients' and populations' needs, governments have found it challenging to assess, scale up, and integrate such solutions.

Smart device-based lifestyle therapies are gaining popularity because they can be powerful weapons against obesity and help to get around the drawbacks of existing approaches. Additionally, they can be useful in accomplishing total lifestyle changes, which have been proven to be quite successful in reducing obesity. Edward O. Thorp introduced the idea of wearable technology in the 1960s [28]. Since then, it has received considerable attention from researchers around the world.

Numerous sectors have benefited from the expanding field of mHealth, including the self-management of chronic diseases, behavior change assistance, and health promotion [29]. Short message service (SMS), text messaging, smartphone apps, and wearable technologies are among the modalities frequently used in mHealth interventions [30]. A subset of digital health, often known as eHealth, which also encompasses telemedicine, customized medicine, and health information technology, is mHealth [31]. Many consumer technology companies have entered the health sector, including Apple, Google, Xiaomi, and Fitbit. Thousands of mobile health or PA apps are available on the Apple App Store and Google Play, but only a small number of them have received FDA approval [6].

mHealth offers real-time monitoring and detection of changes in health status in addition to assisting with the adoption and maintenance of a healthy lifestyle, offering quick diagnoses of health issues, and facilitating the implementation of interventions ranging from encouraging patient self-care to remote health care service delivery and, together with the Internet of Things (IoT), can integrate and store data, carry out intelligent identifications, along with patient data tracking and knowledge sharing [17].

The ability to monitor staff and patients, as well as locate and identify the status of health care equipment and assets, is one of the key changes in health care due to the IoT. These changes lead to improved employee productivity, resource use, efficiency gains, and cost savings, all of which are connected to an integrated system that is useful in the management of disease states [32]. Medically relevant irregularities can be carefully identified, which can lead to knowledge that can be vital for handling a global health crisis.

As part of the mHealth specialization, we discover smart wearable devices, which are electronic components used to monitor, record PA, sedentary behavior, vital signs, such as heart rate, temperature, blood pressure, and blood oxygen saturation, consist of systems that provide direct information on the level of PA, number of steps per day, and activity trackers that frequently include goal setting. As of now, they are frequently employed to quantify PA, mHealth applications, and fitness monitoring [33].

These smart, responsive, cost-effective, digitally based wearable fitness tracking technologies are an innovative platform for promoting health. Commercial smart wearable devices, such as smartwatches, sports bracelets and smart bracelets, rings, specialized clothing, gloves, glasses, headphones, or belts to name a few, all have high processing power and numerous sophisticated sensors [34, 35]. With the aid of numerous technologies for identification, detection, connection, Cloud services, and storage, health applications can also be used directly on the body to perceive, record, analyze, control, and intervene to preserve health. They can even be used to treat diseases [3]. These tools have been acknowledged to have gained popularity as early behavioral and psychological interventions to encourage people to adopt a healthy lifestyle and enhance their quality of life.

Smart wearables offer a wide range of uses in health care, from neurocognitive disorders such as Parkinson's disease, Alzheimer's disease, and other psychiatric diseases, to chronic conditions including cardiovascular disease, hypertension, and muscle disorders [36], as well as many other benefits, such as reducing the number of patients at risk, reducing the risk of readmission, epidemic monitoring, mobile home health care, environmental assisted living, and monitoring in patients with chronic diseases or the effect of medications on the physiological state of patients [37]. There are also advantages in drug trials, treatment efficacy, and chronic disease management.

Apple formally unveiled the Apple Watch in April 2015. By 2025, there will be more than 264 health-related portable sensor applications worldwide, with a Compound Annual Growth Rate of about 38.8% on average [38], with the development of the wristwatch predicted to increase at a particularly rapid rate.

Wearable devices are likely to play a role in reducing the costs of health checks and unwanted burden on health care programs that are already overburdened globally, as well as illuminating the needs in businesses as cost-effective strategies. Given the clinical importance of wearable devices and the associated physiological parameters being measured, companies may be more than interested in the benefits provided by this technology.

One promising way of promoting PA, healthy diets, continuous check-ups, and better health in companies is to implement interventions with the use of mHealth tools [38]. Modern information and communication technologies such as the Internet, cell phones, and other wireless devices are primarily used to offer these interventions.

The use of different portable technological elements such as technological sensors, medical chips, wireless communication, power management, visualization, and information feedback is necessary for the operation and implementation of these devices [39]. They are made up of a transducer and a target receiver. Once the element being examined has been identified, the receiver sends a notification, alarm, or measurement in response. The receiver's answer is subsequently transformed into useable signals by the transducer. Users can do a price analysis in the Cloud or send heart rate, temperature, and other physiological indicators to the relevant body for monitoring without leaving their homes thanks to wireless transmission. In order to remotely monitor a patient's health and analyze the data and provide feedback practically in real time, medical professionals connect to the data server through a secure Internet connection. Alternatively, they can use artificial intelligence to process the data and provide alerts, reminders, warnings, or additional notifications to the user [3]. This offers a low-cost, highly efficient method of managing population health.

These small, portable devices have their own receiver, a signal processor, and run-on batteries, enabling them to function as a "microcomputer" and connect all the processes, from information gathering and processing to communication and power supply. They can also connect to other smart devices via Bluetooth, infrared, radio frequency, or near-field communication technology, including 24 h a day, 7 days a week connectivity using GPS technology [30, 40]. Together, these connections have facilitated the creation of wearable systems that enable previously impractical remote and long-term patient monitoring in homes and communities. This skill is anticipated to make a significant contribution to cutting medical and health care expenditure, as well as being used in businesses to benefit employees' productivity.

Different physics theories are applied to wearable technology. The most common activity tracking technique used in today's portable gadgets, the triaxial accelerometer, monitors linear acceleration along three separate planes. These gadgets take advantage of Newton's second law, which allows the deflection of mass with a particular degree of acceleration to be electronically detected. The benefits of these accelerometers over the piezoresistive kind, which can be influenced by its temperature sensitivity, are low power consumption and quick response.

As part of the design of these smart elements, wearable monitoring systems or sensors have been integrated to capture a variety of data sets, including gait patterns, tremor episodes, activity levels, movement, and physiological parameters with biocompatible materials and nanomaterials; such physical elements are based on the skin, in tattoos, portable devices based on textiles or biofluids [38, 39]. Compared with conventional devices, these flexible electronic devices are not only inexpensive, but also energy efficient, allowing uninterrupted data acquisition for a long period.

Commercial wearable devices can diagnose simple and frequent arrhythmias such as atrial fibrillation or photoplethysmography (PPG) by employing electrocardiography (ECG), which measures heart rate and rhythm by placing a contralateral finger on the crown (negative electrode on the side of the watch). This emits a continuous pulse of photons through the skin, a photodetector analyzes the varying intensity of the photons reflected by the tissue, and a tachogram records the information by classifying heart rate using algorithms and computing the duration between beats [41, 42]. However, owing to the limitations of PPG technology, such as inappropriate skin contact, skin color, moisture, and even tattoos, these single-shunt ECGs are frequently insufficient for the accurate identification of more complex arrhythmias and other diseases, such as myocardial infarction [43].

Other types of motion detection forms, physiological and biochemical characteristics can employ physical principles such as cardiotachometers that measure heart rate, bioimpedance sensors that measure resistance to blood flow and skin tissue, and electromyography sensors that measure changes in muscle status through hand movements [44]. From a design point of view, measurements of physiological and biochemical signals from bodies can be more complicated than tracking physical movement, as they require the collection, processing, and analysis of data, mainly in relation to body fluids.

Other types of measuring devices use the electrochemical principle of glucose oxidase [45]. Devices based on this idea use a minimally invasive needle sensor that is typically placed into the arm, abdomen, or subcutaneous tissue and monitors an electric current signal produced by the glucose oxidase reaction. A calibration technique that is typically done twice a day converts this signal, which is proportional to the concentration of glucose present in the interstitial fluid, into a glucose concentration.

There is also smart clothing, which functions in a similar manner to other wearable technology in that it monitors the wearer's health. With a variety of sensors and functionalities, smart clothing includes a variety of products but typically shirts, socks, yoga pants, shoes, helmets, and hats. It has the potential to be extremely useful for firefighters, workers on construction sites, and for transportation.

These tools have demonstrated good accuracy in measuring the user's physiological parameters, which has led to their application in several academic investigations. Examples of wearable medical monitoring devices include the Apple® Watch® 2, Samsung Gear S3®, Xiaomi Mi®, and Huawei Talk Band B2®. These devices have expanded functionality that is frequently linked to communication and seem to offer a promising way of addressing issues in serious diseases such as long-term glycemic control in type 2 diabetes [46–48].

Finally, the most cutting-edge wearables can be implanted into the user's body or consumed by users. Examples of these technologies include sensors that can be ingested in a patient's medication to confirm in real time whether the medication is being ingested and technology that enables a wireless sensor to be implanted in the skin to monitor changes in blood sugar levels for patients who require constant monitoring, such as those with diabetes [6].

These medical devices used to measure key health indicators, detect relevant diseases, and for chronic disease management are reflected in Table 3.

Wearable systems backed by advanced information technology and sophisticated sensors could continuously monitor human physiology and consequently accelerate treatments.

Table 3 Wearables and mHealth – clinical usage

Key health indicators	Disease detection	Chronic disease management
ECG Heart rate Blood pressure Blood oxygen saturation Body temperature Posture P	1. Hypertension 2. Atrial fibrillation 3. Cholesterol levels 4. Ventricular dysfunctions	1. Cardiovascular: Heart failure Arrhythmias Coronary artery disease QT interval measurements Hyperkalemia Cardiac rehabilitation Peripheral vascular disease, 2. Lung diseases: Obstructive pulmonary disease Bronchial asthma Obstructive sleep apnea Hypopnea syndrome 3. Metabolic diseases: Diabetes Hyperglycemia Hypertension 4. Other disorders Alzheimer's Malignant gliomas

4.2 Evidence for Functionality in the Integration of Wearables for Prevention and Health Promotion

Several studies have combined capnography, stroke volume, pain, level of consciousness, and urine production to accurately determine associated physiological changes in patients as well as the mobile monitoring system, which is a solution that allows mobile monitoring of the patient through mobile phones and biometric devices [16]. Some research has been done analyzing the use of hybrid solutions (mobile and multimedia), which is preferred by 50.4% over the Cloud-based framework to monitor NCDs, followed by the framework with mobile elements with 29.5%, and finally with 20% Cloud-based multimedia elements [3]. Researchers also discovered that in the health support criterion is more relevant, with 41.8% for the implementation of digital health.

Researchers conducted a randomized study in the working population using an awareness and knowledge program for the prevention of coronary heart disease. The findings show that after the mHealth intervention, participants in the intervention group further improved their knowledge of coronary heart disease, including that a family history of heart disease is a risk factor and that controlling the level of glucose in the blood helps to reduce the cardiac risk [49]. These results imply that specific mHealth applications can be utilized as efficient means of enhancing current services and programs and educating the general public about illness prevention.

Markov models have been used for the identification, measurement, and monitoring of sleep to design a customized model to improve cycles in people through algorithm-based recommendations [50]. An example of the analysis of these smart devices was elaborating on measures related to sleep, contemplating the total time of sleep and stage of sleep efficiency [51].

Belt wearables provide users with a daily activity score according to the employee's habits and making comparisons between daily activities and activities of the previous days. Smart device-based interventions in these populations significantly reduced systolic blood pressure, waist circumference, and low-density lipoprotein cholesterol concentration [33, 52].

Apple's Health Kit enables the integration of smart medical devices from many vendors that have been bought for app and data integration. The Internet can be used to send information to medical professionals so that they can complete the monitoring of the user's health and the initial diagnosis of the ailment. This is done through the integrated processing of the obtained signal data [53].

Wearables can improve quality of life, they effectively monitor PA and rest, and with respect to smartphones, there is a diversity of measurements in PA that show relevance in their usefulness [24]. The result of meditation and mindfulness notably helps the parasympathetic system to have greater activity, which translates into a better quality of life.

Continuous glucose monitoring sensors, non-invasive or minimally invasive wearable devices that measure glucose levels by utilizing various physical principles, have revolutionized diabetes monitoring [54]. Research reported that people with type 1 or type 2 diabetes who used SMS messages to send self-controlled blood glucose values and receive self-monitoring information revealed a 0.5% decrease in HbA1c. Another study on mobile apps for diabetes suggested an overall efficacy in reducing HbA1c, with an average decrease of 0.44% in the intervention [55, 56]. Others implemented a digital therapeutic diabetes prevention program and found it to be effective in preventing type 2 diabetes through significant body weight reduction and increased PA [11]. Other studies still used strategies for insulin management such as education, self-management, and prevention through smart devices. Most interventions demonstrated clinically and statistically significant efficacy in reducing HbA1c values related to intervention and control [57].

Although some insurance firms are advocating the use of portable gadgets to promote healthy eating habits and increase employee well-being, the logistics sector has started using ProGloves, portable gloves that can scan barcodes to make it easier to do work that does not require the use of hands [49].

As a part of the concept of Health 4.0 artificial intelligence has taken on relevance, also because of the efficiency in the diagnostics and fast results. Researchers have been make great efforts to train computers using the deep learning mechanism with the IoT and based on wearable sensors for prediction and symptom analysis in the health care sector [58].

Literature reviews show that smart wearable devices have had effects on exercise, stress management, self-realization, and comprehensive, physical, and mental quality of life, and interventions in PA based on wearable health technology are effective in increasing and improving healthy habits in these populations, given their accessibility, cost-effectiveness, and motivating characteristics [59, 60].

Using smart devices for monitoring and better management of diseases found that wearable devices can help in the control of diabetes, as well as in the prevention of complications associated with this condition and the effectiveness, feasibility, and acceptability of mHealth technology, including wearable activity monitors and smartphone apps for promoting PA and reducing sedentary behavior in workplaces. A significant increase in PA was reported with a significant reduction in sedentary time [61, 62].

Not only analytic and quantitative scholars have evaluated the concepts in Health 4.0; other relevant areas of science had also taken on relevance in recent years and complement in depth the different concepts evaluated. This is qualitative research. In addition, researchers have been involved in evaluating the feelings, fears, and motivations using this new wave of technologies [63]. For example, understanding how the IoT has been supportive for users and health associates, expressing the feelings and distress from truck drivers with the usage of wearables during work, although they were positive about living a better lifestyle, or implying the constant usage of a digital tool, is hard work and constant reinforcement from medical experts is needed; however, the benefits have been felt [64, 65].

4.3 mHealth and Wearables Within the Companies

In the sense of boosting motivation and fostering personal growth, work should be a source of health, pride, and enjoyment [66]. Employees who are happy and healthy perform better and stick with the company longer. However, if a person is consistently under a heavy burden for an extended length of time and is unable to recover, the labor may have long-lasting bad effects and even serious ill effects. Owing to the transition from manual labor to mostly sedentary tasks, many job descriptions have changed. Working long hours over extended periods of time is linked to chronic stress illness, chronic heart disease, and sleeping difficulties [66].

Sedentary behavior and physical inactivity provide health concerns to workers and place a financial burden on employers because they cause frequent business losses owing to employee illness and bad health. But as people who work with this illness have a lower work performance, lower wages, and experience job loss owing to the decrease in work capability related to the condition [5], this issue also affects employees. Many businesses are integrating smart wearable technology into their corporate wellness programs to improve employee health and well-being and save insurance costs [67].

Organizations are introducing more and more portable devices in the hope of improving performance. Wearables provide new and unique opportunities to engage employees with their work and organizational environment and the performance-related feedback provided by these devices helps both employees and managers to navigate the work environment more effectively [68].

With these things in mind, the workplace is one of the optimal places for comprehensive mobile-based lifestyle interventions. The number of employees will be constant during the intervention period, and resources provided by the employer (cafeteria and gym), couriers, and periodic medical check-up services can be incorporated into the interventions. The characteristics of mobile health apps (e.g., real-time lifestyle tracking and participant interactivity) can be easily integrated into workplace interventions [69].

Five hundred wearable technology adopters who stated that they were presently using, testing, or planning to implement wearable technology in the enterprise were surveyed by Salesforce Research. Seventy-nine percent of users concur that wearable technology is crucial to the future success of their organization, and 76% say that since using portable devices in the workplace, business performance has improved [70]. Additionally, 13% mention that they are already using smart devices for the measurement of vital biomedical elements, 28% are planning to do so, and in the same way 7% are already using them for follow-up of clinical trials. Therefore, it is crucial to comprehend how technology can assist organizations in achieving the benefits of having healthy employees with the potential to increase productivity, efficiency, connectivity, health, and well-being given the impact of a happy and healthy workforce on organizations and public health [71].

The use of wearable devices has the potential to engage employees through user engagement functions by collecting quantified self-data, such as weight, diet, exercise routines or sleep patterns and skin conductance of heart rate and blood pressure. Wearables can transform the work experience by eradicating some of the limitations and improving the capabilities and competencies of employees by informing employees of useful elements about their body, temperature, blood pressure, oxygen saturation, among others [47]. Monitoring employee physiological responses, such as heart rate variability or respiration, can serve as an indicator of employees' stress and/or fatigue responses. These smart devices can alert employees to take breaks from work when their performance deteriorates to optimize productivity.

Finally, there are correlations between health apps and three main factors. First, health apps can potentially influence employee retention and relationships. Second, health apps have the underlying potential to assist in the prevention of widespread disease, the early detection of disease, the treatment of disease, and the provision of targeted offers to customers. Third, there are the possibilities of big data, such as detecting trends on the collective of employees and providing specific initiatives based on knowledge [72].

As part of the findings of the systematic review of the literature, 14 articles were found, which are reflected in Table 4, and were considered a relevant result in the analysis of the information as they manifest certain interventions carried out within companies based on the concept of Health 4.0, where elements such as wearables, smartphones, and specialized applications were used to address various problems such as sedentary lifestyle, obesity, metabolic syndrome, cardiovascular disease, or stress.

Numerous staff programs that result in a sizable decrease in body weight and a rise in PA are successful in preventing type 2 diabetes [11]. Waist circumference considerably decreased during all time periods, and the mean weight decreased

Table 4 Health 4.0 programs in firms

No.	Year	Initiative	Variables	Methods	Tools used	References
1	2019	Diabetes Prevention Program	Absenteeism Weight loss	Interactive computing, remote monitoring, monitoring tools, in-process tracking for 1 year	Wearables	[11]
2	2019	Weight Management Program	Obesity Metabolic profile	Integrated, customized mobile technology, called Health-On, optimized for workplaces	Smartphone	[17]
3	2021	Occupational Stress Reduction Program	Sleep quality Cholesterol	Multiple linear regression models to determine the association between exposures over 3 months	App and smartphone	[67]
4	2019	Obesity Care Program	Obesity Physical activation	Mixed model analysis, for 12 weeks	Belt wearable	[51]
5	2016	Wellness and Exercise Program	Cardiovascular events PA	Measurement of heart rate, breathing, temperature and ECG. A decision model based on cost-effective calculation	Wearable	[78]
6	2016	PA trackers through incentives	Incentives/donations PA	Computer-generated assignments, impact of incentives vs steps generated over 12 months	Wearable	[76]
7	2020	eMotivate4Change program	Metabolic syndrome Lifestyle	Entertainment and motivating factors were measured for 4 months	Applications and wearable	[6]

8	2020	Program to reduce sedentary lifestyle	Sedentary Lifestyle	Preliminary registration, workshop, boards, examination of usability, goals set, warnings and reminders were generated when sedentarism occurred	Digital application	[74]
9	2021	Wearable devices and glycemic metrics	Glycemic profile	Analysis of stored wearable data for 250 days compared with the glycemic profile	Wearable	[73]
10	2015	Customized real-time monitoring program	Environmental temperature PA	Measurement of environmental and body sensors and continuous monitoring of PA performance in occupational heat stress	Wearable	[77]
11	2016	Internet and mobile technologies program	PA Social networks	Interviews	Wearables	[75]
12	2017	Coronary Heart Disease Prevention Program	Coronary heart disease Stress Lifestyle	Advice by cardiologists, use of applications for training, lifestyle improvement program for 4 weeks	Applications and wearable	[8]
13	2017	Corporate Wellness Program	Step counting Lifestyle Incentives	Daily step counting, incentive analysis, continuous surveys	Wearable	[69]
14	2013	Stress Minimization Program	PA Quality of sleep	Multinomial logistic regression models were built using audio, PA, and communication data gathered during the working day and heart rate variability collected throughout the night	Chest-type wearable, smartphone, audio recording system	[68]

by 5.8%. The median weight also decreased, from 81.3 kg before the intervention to 76.6 kg afterward [17]. Waist circumference decreased consistently and equally [58]. Stress levels were inversely and strongly correlated with low-density lipoprotein cholesterol [67]. The classification of the importance of the wearable sensor in estimating HbA1c was skin temperature (33%), electrothermal activity (28%), accelerometry (25%), and heart rate (14%). Studies demonstrate the feasibility of using non-invasive wearable devices to estimate glucose and HbA1c variability for glycemic control [73]. There were significant effects on BMI and cholesterol of group and time and a better overall level of knowledge, and better behavior regarding blood cholesterol control improves the level of health [6, 8]. The findings imply that particular apps were well received and that users thought favorably of their features.

Prolonged use of technology has a positive impact on employee well-being, and support from clinical experts in helping and explaining the clinical pathway to employees has been crucial [69, 74]. In addition, social media containment could be beneficial for increased users' adherence and engagement with wearables, but long-term incentives may need to be implemented to avoid decreased PA [75, 76]. The wearables provided more awareness than motivation in PA with goal setting and progress tracking and it allows workers to be notified in a timely manner so that they can make decisions based on objective information [75, 77]. However, it is relevant to consider the cost-effectiveness of the program to evaluate the correct investment in the company [78]. Finally, privacy concerns need to be addressed before the general adoption of wearable technology is considered by businesses [51].

4.4 Challenges to Be Addressed

Because they gather, store, and transmit a variety of health- and fitness-related data, wearable health care devices are acknowledged to be a great support tool for an individual's health and can have a substantial impact on people who are health conscious. People's intent to use a gadget or program will be positively influenced if they believe that its services or features are beneficial [6]. In an evaluation, 69.5% of respondents were willing to adopt portable devices, although 77.8% of them were concerned about issues related to economic benefits, data privacy, and, to a lesser extent, technological precision [79]. Additionally, because 5G technology can increase geographic coverage, the spread of remote medical services is anticipated to pick up steam [80].

Despite the compelling benefits of wearables there is still a lot of room for improvement in smart devices owing to their limitations, including stability, sensitivity, privacy, power source, and limited applications in certain diseases such as neurological disorders; in addition, they can be detrimental to the performance of the organization unless several issues are addressed for success in implementing the use of these elements as an improvement strategy. Barriers to the adoption of wearables, quality of detection, connectivity, data processing, integration, and commitment and interaction with the patient have been detected. Thus, these areas of concern were divided into four categories: technological, security, acceptance, and permanence.

In the technology category, one of the main limitations hindering the usefulness of portable devices is a continuous power supply. The batteries used in portable devices have limited space, as they are required to align with the design of portable devices. Other studies have noted issues with a lack of visible information, difficult data entry, and biased feedback [81]. In addition, most wearable devices available on the market have a limited scope, track few variables and are unable to accurately evaluate several health status indicators, including heart rate variability, diet, and mood [82].

The data collected from the sensors in most wearable devices can be distorted and noisy owing to the body's constant movement and skin hairs, which result in minimal adhesion between the skin and the wearable device. Most diagnostic techniques use samples such as blood, urine, and saliva, with only a limited amount of integration with these samples [38].

Another category can be included in the issues of security and data privacy; because these devices contain protected health information, uncompromising privacy is required. To safeguard user data in this regard, secure communication protocols are crucial [83]. The protection of personal information and the assurance of data security in communication channels are of the utmost importance. The acceptance of wearable health monitoring technology could be hampered by worries about trust and privacy [38]. These worries may have a negative impact on how much value and acceptance users give to wearable health monitoring equipment. Health information technology must abide by the Health Insurance Portability and Accountability Act, as well as patient demands, care expectations, and the evolving health care landscape, in order to prevent this [84].

It is recommended as a prerequisite for its adoption to contemplate different regulations with the various interested parties such as data ownership rights, data rights, intellectual property, copyright in data, and even what is contained in various contracts [85]. These devices should be regulated through comprehensive assessment frameworks and appropriate regulatory supervision policies to ensure safety and efficacy.

The third category we call adoption challenges. Incentives may be a valuable tactic for boosting the maintenance and efficacy of behavior modification treatments, particularly PA therapies that rely on wearable devices [7]. This is according to economic knowledge. Using rewards that are directly related to PA is one way of improving the efficacy of wearable technologies. According to economic theory, people increase their PA in an effort to accomplish a PA objective and receive incentives that do not always require direct cash payments.

Other ways of promoting acceptance have been with text messaging, apps, and several components of the mHealth programs that have provided some success with weight management [86]. In addition, the real-time delivery of output data offers the possibility of continuous adjustment of interventions according to the changes a person makes after each message they receive from the program.

These incentives, which are linked to healthy behaviors, such as assigning rewards to charity, as is frequently done on walks, are also consistent with the theory of reasoned action, which holds that people will engage in a behavior if they positively evaluate it and if they believe it is crucial for others to think that they should engage in it [7]. Third parties who could be more ready to cover the expenditure owing to the favorable publicity generated by fundraising for charities and the frequent eligibility for tax deductions may also like this tactic.

Finally, a category we call the problem of permanence, as there are presently several applications that assist in managing users' health. Users may report a poor experience after downloading the desired app, as they did not find the information to be useful or relevant, which makes it challenging to utilize the apps over time [81]. Research revealed that more than half of US consumers who owned a modern activity tracker no longer used it after 18 months, and one third quit within 6 months [87]. Encouraging users to share data publicly is one controllable aspect that can increase device engagement for longer periods of time. Wearable adoption seems to be transient, and self-efficacy may have contributed to adherence to PA.

Potentially, the' self-managing, motivating, and goal-setting features of these commercially available gadgets could also assist patients with chronic conditions in adhering to long-term PA more successfully. By raising awareness among patients and physicians and encouraging them to take the necessary precautions to manage such diseases, the use of health monitoring through smart devices will have a positive impact on reducing the abandonment of programs and precautions related to health management at the individual and collective levels [88]. On the other hand, more investment and developments in digital tools are needed to break down the barriers of adoption, mainly regarding cost, sustainability, and integration with health care systems [63].

To the extent that wearable devices overcome these limitations, they hold great promise for expanding the clinical repertoire of patient-specific measurements and are considered an important tool for the future of precision health. Their exploitation of IoT technology, especially with the advent of 5G technology, and interaction with electronic medical records on smartphones, project telemedicine toward new horizons and challenges, such as better management of workers and the reduction of relative direct and indirect costs [89].

Therefore, it is crucial to involve a variety of researchers (such as those who specialize in exercise science, public health, bioinformatics, statistics, and methodology) and stakeholders (such as those who work with patients, doctors, electronic health records, and wearable technology companies) to comprehend the problems with mHealth technology and collaborate in finding solutions [90].

5 Discussion

mHealth technology is an excellent tool for utilizing in corporate wellness programs aimed at building a happy and healthy workforce because technology in Health 4.0 is promoted as a way for individuals to better their health and well-being. Additionally, it might make social interactions amongst coworkers easier. Managers frequently use incentives to encourage people to meet PA goals by promoting competence across organizational units. By coming together to decide on winning strategies and encourage one another during games and competitions, employees who participate in such events have the chance to take advantage of the conditions for social cohesion.

Health 4.0 improves organizational performance by raising competences and bolstering organizational culture. However, the benefits of using wearable technology to boost organizational performance come at the expense of subjecting workers to ongoing surveillance, which poses significant ethical questions.

Integrating mHealth devices into the workplace presents a number of difficulties related to more general features of the workplace environment linked to both the immediate effects of wearable devices on employees and particular operational aspects related to the data collected by wearable devices.

mHealth can raise concerns about privacy risks and ethical issues if used in a mass environment such as a workplace, as has been commented already [91]. Similarly, self-tracking through wearable devices in the workplace can have political and social justice implications because employees must participate in the imposed tracking. Also, corporate wellness programs have historically been affected by the lack of employee participation and the difficulty of tracking employees' progress.

There are different approaches to seeking and improving health policies within companies. Some of them have ventured to seek tangible benefits for workers using Health 4.0 tools, including weight control [11, 58, 67], PA [68, 69, 74, 76–78], sleep control [51], reduction of a sedentary lifestyle [76], and glycemic index [74] for and impact on NCDs such as diabetes [6, 9, 11, 51], on weight [28], work stress [51, 68, 73], cardiovascular diseases [8, 78], and lifestyle [6, 69, 74, 77] through the use of bracelet-type wearables [11, 69, 74, 76–78], those worn on a belt [58, 68], or those on clothing [73], as well as the use of smartphones with applications [6, 8, 67, 68, 74], and encourage its use through an incentive program [76], increased knowledge [6, 8], or with the assistance of professionals as a medical guide [6] or containment through the use of social networks [77], but only [78] with a cost-effective approach, although with results not favorable to the use of portable devices. However, it should be noted that, during the time that has elapsed, the technology of the devices is already on an economy of scale and the prices of such as devices are below the effective cost limit shown in the article, meaning that, probably today we can get better results. In addition, few are addressing the issues concerning security and data privacy, which are understood to be relevant to successfully implementing these health programs.

This article shows the details of the operation of Health 4.0 and its various existing tools, mainly the mHealth vector, revealing both quantitative and qualitative information on how mHealth has worked in health care, the elements that can be measured, and the results evaluated in various diseases. Moreover, various articles were found showing how certain mHealth tools have been implemented within companies and how they have benefited from the measurement of indicators. We can also observe certain complications and challenges in the implementation of Health 4.0 in the work environment.

Although these studies already set a precedent in the use of the technological elements contained in the Health 4.0 concept, we can still observe that no article or study has been found in which the impact of both prevention and promotion initiatives can be demonstrated in the drivers of key labor indicators such as productivity. It was found that there is a lack of focus on diseases such as metabolic syndrome, some cardiovascular diseases, and even mental illnesses, with solutions based on wearables, smartphones, use of a medical and nutritional guide with continuous advice from expert doctors, support groups for the continuation of the program, and everything based on a series of rewards that the company provides. In the same way, it is important to contemplate, through a savings analysis, either for medical premiums paid to health care institutions or a reduction of costs associated with absenteeism, the benefits of said programs. Cost-effectiveness analyses are necessary to better extend implementation opportunities through effective resource utilization testing, initiative validation, as well as technology utilization and policy focus within companies. Moreover, research on technology to accelerate diagnostics through real-time data analysis and on the usage of more elements from the IoT such as the health Cloud or artificial intelligence is needed. Finally, the security and data privacy approaches need to be considered in every step of the programs.

It is important to continue researching more about the operation of smart devices and to improve the conditions of workers, the work environment, and the benefits that can impact the organization.

References

1. Lu, L., Zhang, J., Xie, Y., Gao, F., Xu, S., Wu, X., Ye, Z.: Wearable health devices in health care: Narrative systematic review. JMIR Mhealth Uhealth. **8**(11) (2020). https://doi.org/10.2196/18907
2. WHO, World Health Organization.: mHealth: New horizons for health through mobile technologies: Second global survey on eHealth (2011). https://apps.who.int/iris/handle/10665/44607. Last accessed 2022/09/18
3. Al-Qurishi, M., Al-Rakhami, M., Al-Qershi, F., Hassan, M.M., Alamri, A., Khan, H.U., Xiang, Y.: A framework for cloud-based healthcare services to monitor noncommunicable diseases patient. Int. J. Distrib. Sensor Netw. **11**(3), 985629 (2015). https://doi.org/10.1155/2015/985629
4. Baig, M.M., GholamHosseini, H., Moqeem, A.A., Mirza, F., Lindén, M.: A systematic review of wearable patient monitoring systems–current challenges and opportunities for clinical adoption. J. Med. Syst. **41**(7), 1–9 (2017). https://doi.org/10.1007/s10916-017-0760-1

5. Zhang, H., Jiang, Y., Nguyen, H.D., Poo, D.C.C., Wang, W.: The effect of a smartphone-based coronary heart disease prevention (SBCHDP) programme on awareness and knowledge of CHD, stress, and cardiac-related lifestyle behaviours among the working population in Singapore: A pilot randomised controlled trial. Health Qual. Life Outcomes. **15**(1), 1–13 (2017). https://doi.org/10.1186/s12955-017-0623-y
6. Lee, J.S., Kang, M.A., Lee, S.K.: Effects of the e-Motivate4Change program on metabolic syndrome in young adults using health apps and wearable devices: Quasi-experimental study. J. Med. Internet Res. **22**(7), e17031 (2020). https://doi.org/10.2196/17031
7. Finkelstein, E.A., Sahasranaman, A., John, G., Haaland, B.A., Bilger, M., Sloan, R.A., Khaing, E.E., Evenson, K.R.: Design and baseline characteristics of participants in the TRial of Economic Incentives to Promote PA (TRIPPA): A randomized controlled trial of a six-month pedometer program with financial incentives. Contemp. Clin. Trials. **41**, 238–247 (2015). https://doi.org/10.1016/j.cct.2015.01.020
8. Zhang, Y., Cui, J., Ma, K., Chen, H., Zhang, J.: A wristband device for detecting human pulse and motion based on the Internet of Things. Measurement. **163**, 108036 (2020). https://doi.org/10.1016/j.measurement.2020.108036
9. WHO, World Health Organization.: Diabetes (2021). https://www.who.int/news-room/fact-sheets/detail/diabetes. Last accessed 2022/09/18
10. IDF: Diabetes Atlas, 8th edn. International Diabetes Federation, Executive Office, Brussels (2017) https://diabetesatlas.org/upload/resources/previous/files/8/IDF_DA_8e-EN-final.pdf, Last accessed 2022/09/18
11. Alwashmi, M., Mugford, G., Abu-Ashour, W., Nuccio, M.: A digital diabetes prevention program (transform) for adults with prediabetes: Secondary analysis. JMIR Diab. **4**(3) (2019). https://doi.org/10.2196/13904
12. Garvey, W., Arathuzik, G.: Lifestyle therapy for diabetes mellitus. In: Mechanick, J., Kushner, R. (eds.) Lifestyle Medicine. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-24687-1_22
13. Wong, S.S., Yu, M., Loprinzi, P.D., Hongu, N.: Smart applications to track and record physical activity: Implications for obesity treatment. Smart Homecare Technol. TeleHealth. **2**, 77 (2014) https://doi.org/10.2147/SHTT.S41484
14. Mancia, G., Fagard, R., Narkiewicz, K., Redán, J., Zanchetti, A., Böhm, M., et al.: 2013 practice guidelines for the management of arterial hypertension of the European Society of Hypertension (ESH) and the European Society of Cardiology (ESC): ESH/ESC Task Force for the Management of Arterial Hypertension. J. Hypertens. **31**(10), 1925–1938 (2013). https://doi.org/10.3109/08037051.2013.817814
15. Parati, G., Torlasco, C., Omboni, S., Pellegrini, D.: Smartphone applications for hypertension management: A potential game-changer that needs more control. Curr. Hypertens. Rep. **19**(6), 1–9 (2017). https://doi.org/10.1007/s11906-017-0743-0
16. Sunny, J.S., Patro, C.P.K., Karnani, K., Pingle, S.C., Lin, F., Anekoji, M., et al.: Anomaly detection framework for wearables data: A perspective review on data concepts, data analysis algorithms and prospects. Sensors. **22**(3), 756 (2022). https://doi.org/10.3390/s22030756
17. de la Torre Díez, I., Garcia-Zapirain, B., Méndez-Zorrilla, A., López-Coronado, M.: Monitoring and follow-up of chronic heart failure: A literature review of eHealth applications and systems. J. Med. Syst. **40**(7), 1–9 (2016). https://doi.org/10.1007/s10916-016-0537-y
18. Chen, Y.H., Sawan, M.: Trends and challenges of wearable multimodal technologies for stroke risk prediction. Sensors. **21**(2), 460 (2021). https://doi.org/10.3390/s21020460
19. Benjamin, E.J., Muntner, P., Alonso, A., Bittencourt, M.S., Callaway, C.W., Carson, A.P., et al.: Heart disease and stroke statistics—2019 update: A report from the American Heart Association. Circulation. **139**(10), e56–e528 (2019). https://doi.org/10.1161/CIR.0000000000000659
20. Machorro-Cano, I., Alor-Hernández, G., Paredes-Valverde, M.A., Ramos-Deonati, U., Sánchez-Cervantes, J.L., Rodríguez-Mazahua, L.: PISIoT: A machine learning and IoT-based smart health platform for overweight and obesity control. Appl. Sci. **9**(15) (2019) http://dx.doi.org/3037.10.3390/app9153037

21. Gal, R., May, A.M., van Overmeeren, E.J., Simons, M., Monninkhof, E.M.: The effect of PA interventions comprising wearables and smartphone applications on PA: A systematic review and meta-analysis. Sports Med. Open. **4**(1), 1–15 (2018). https://doi.org/10.1186/s40798-018-0157-9
22. Obaidat, M., Rana, S., Maitra, M., Giri, D., Dutta, S.: Chapter 19: Biometric security and Internet of Things (IoT). In: Biometric-Based Physical and Cybersecurity Systems, pp. 477–509 (2018). https://doi.org/10.1007/978-3-319-98734-7_19
23. Sonnier, P.: The Fourth Wave: Digital Health. Independently Published (2017)
24. Rodriguez, V.H.: Aplicación de sensores vestibles y teléfonos inteligentes en el bienestar personal: Cuantificación de la actividad física y control de la práctica de mindfulness. Tesis Doctoral. Universidad de Zaragoza (2021)
25. Zhu, H., Wu, C.K., Koo, C.H., Tsang, Y.T., Liu, Y., Chi, H.R., Tsang, K.F.: Smart healthcare in the era of internet-of-things. IEEE Consum. Electron. Mag. **8**(5), 26–30 (2019). https://doi.org/10.1109/MCE.2019.2923929
26. García Lupi Vergara, L., García Pereira, A., Holdorf Lopez, M.: Estado da arte em wearables para saúde (2014). https://repositorio.uca.edu.ar/handle/123456789/7964
27. WHO, World Health Organization.: mHealth. Use of appropriate digital technologies for public health: report by Director-General. 71st World Health Assembly provisional agenda item, 12, A71 (2018). Last accessed 2022/09/18. https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_20-en.pdf
28. Amft, O., Lukowicz, P.: From backpacks to smartphones: Past, present, and future of wearable computers. IEEE Pervas. Comput. **8**(3), 8–13 (2009). https://doi.org/10.1109/MPRV.2009.44
29. Balbim, G.M., Marques, I.G., Marquez, D.X., Patel, D., Sharp, L.K., Kitsiou, S., Nyenhuis, S.M.: Using Fitbit as an mHealth intervention tool to promote PA: Potential challenges and solutions. JMIR Mhealth Uhealth. **9**(3) (2021). https://doi.org/10.2196/25289
30. Muralidharan, S., Ranjani, H., Anjana, R.M., Allender, S., Mohan, V.: Mobile health technology in the prevention and management of type 2 diabetes. Indian J. Endocrinol. Metab. **21**(2), 334 (2017)
31. Guk, K., Han, G., Lim, J., Jeong, K., Kang, T., Lim, E.K., Jung, J.: Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials. **9**(6), 813 (2019). https://doi.org/10.3390/nano9060813
32. Shah, M.: Big Data and the Internet of Things. Big Data Analysis: New Algorithms for a New Society, pp. 207–237. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26989-4_9
33. Franssen, W., Franssen, G.H., Spaas, J., Solmi, F., Eijnde, B.O.: Can consumer wearable activity tracker-based interventions improve PA and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomized controlled trials. Int. J. Behav. Nutr. PA. **17**(1), 1–20 (2020). https://doi.org/10.1186/s12966-020-00955-2
34. Villarreal, V., Berbey-Alvarez, A.: Evaluation of mHealth applications related to cardiovascular diseases: A systematic review. Acta Informática Medica. **28**(2), 130 (2020) https://dx.doi.org/10.5455%2Faim.2020.28.130-137
35. Bayoumy, K., Gaber, M., Elshafeey, A., Mhaimeed, O., Dineen, E.H., Marvel, F.A., Martin, S.S., Muse, E.D., Turakhia, M.P., Tarakji, K.G., Elshazly, M.B.: Smart wearable devices in cardiovascular care: Where we are and how to move forward. Nat. Rev. Cardiol. **18**(8), 581–599 (2021). https://doi.org/10.1038/s41569-021-00522-7
36. Iqbal, S., Mahgoub, I., Du, E., Leavitt, M.A., Asghar, W.: Advances in healthcare wearable devices. npj Flex. Electron. **5**(1), 1–14 (2021). https://doi.org/10.1038/s41528-021-00107-x
37. Binyamin, S.S., Hoque, M.: Understanding the drivers of wearable health monitoring technology: An extension of the unified theory of acceptance and use of technology. Sustainability. **12**(22), 9605 (2020). https://doi.org/10.3390/su12229605
38. Lucivero, F., Jongsma, K.R.: A mobile revolution for healthcare? Setting the agenda for bioethics. J. Med. Ethics. **44**(10), 685–689 (2018) https://doi.org/10.1136/medethics-2017-104741

39. Müller, A.M., Alley, S., Schoeppe, S., Vandelanotte, C.: The effectiveness of e-& mHealth interventions to promote PA and healthy diets in developing countries: A systematic review. Int. J. Behav. Nutr. PA. **13**(1), 1–14 (2016). https://doi.org/10.1186/s12966-016-0434-2
40. Kristoffersson, A., Lindén, M.: A systematic review of wearable sensors for monitoring PA. Sensors. **22**(2), 573 (2022). https://doi.org/10.3390/s22020573
41. Yang, C.C., Hsu, Y.L.: A review of accelerometry-based wearable motion detectors for PA monitoring. Sensors. **10**(8), 7772–7788 (2010). https://doi.org/10.3390/s100807772
42. Cobos Gil, M.Á.: Standard and precordial leads obtained with an Apple Watch. Ann. Intern. Med. **173**(3), 249–250 (2020). https://doi.org/10.7326/L20-0588
43. Kamišalić, A., Fister, I., Turkanović, M., Karakatič, S.: Sensors and functionalities of non-invasive wrist-wearable devices: A review. Sensors. **18**(6), 1714 (2018). https://doi.org/10.3390/s18061714
44. Dagher, L., Shi, H., Zhao, Y., Marrouche, N.F.: Wearables in cardiology: Here to stay. Heart Rhythm. **17**(5), 889–895 (2020). https://doi.org/10.1016/j.hrthm.2020.02.023
45. Chan, M., Estève, D., Fourniols, J.Y., Escriba, C., Campo, E.: Smart wearable systems: Current status and future challenges. Artif. Intell. Med. **56**(3), 137–156 (2012). https://doi.org/10.1016/j.artmed.2012.09.003
46. Wang, J.: Electrochemical glucose biosensors. Chem. Rev. **108**(2), 814–825 (2008). https://doi.org/10.1021/cr068123a
47. Xie, J., Wen, D., Liang, L., Jia, Y., Gao, L., Lei, J.: Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: Comparative study. JMIR Mhealth Uhealth. **6**(4), e9754 (2018). https://doi.org/10.2196/mhealth.9754
48. Luo, J., Zhang, K., Xu, Y., Tao, Y., Zhang, Q.: Effectiveness of wearable device-based intervention on glycemic control in patients with type 2 diabetes: A system review and meta-analysis. J. Med. Syst. **46**(1), 1–15 (2022). https://doi.org/10.1007/s10916-021-01797-6
49. Mardonova, M., Choi, Y.: Review of wearable device technology and its applications to the mining industry. Energies. **11**(3), 547 (2018). https://doi.org/10.3390/en11030547
50. Villarreal, V., Hervás, R., Bravo, J.: A systematic review for mobile monitoring solutions in m-health. J. Med. Syst. **40**(9), 1–12 (2016). https://doi.org/10.1007/s10916-016-0559-5
51. Meguro, K., Svensson, T., Chung, U.I., Svensson, A.K.: Associations of work-related stress and total sleep time with cholesterol levels in an occupational cohort of Japanese office workers. J. Occup. Health. **63**(1), e12275 (2021). https://doi.org/10.1002/1348-9585.12275
52. Lee, M., Shin, J.: Change in waist circumference with continuous use of a smart belt: An observational study. JMIR Mhealth Uhealth. **7**(5), e10737 (2019). https://doi.org/10.2196/10737
53. Zhang, G.: A wearable device for health management detection of multiple physiological parameters based on ZigBee wireless networks. Measurement. **165**, 108168 (2020). https://doi.org/10.1016/j.measurement.2020.108168
54. Cappon, G., Acciaroli, G., Vettoretti, M., Facchinetti, A., Sparacino, G.: Wearable continuous glucose monitoring sensors: A revolution in diabetes treatment. Electronics. **6**(3), 65 (2017). https://doi.org/10.3390/electronics6030065
55. Liang, X., Wang, Q., Yang, X., Cao, J., Chen, J., Mo, X., Huang, J., Wang, L., Gu, D.: Effect of mobile phone intervention for diabetes on glycemic control: A meta-analysis. Diabet. Med. **28**(4), 455–463 (2011). https://doi.org/10.1111/j.1464-5491.2010.03180.x
56. Bonoto, B.C., de Araújo, V.E., Godói, I.P., de Lemos, L.L.P., Godman, B., Bennie, M., Diniz, L.M., Junior, A.A.G.: Efficacy of mobile apps to support the care of patients with diabetes mellitus: A systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth. **5**(3), e6309 (2017). https://doi.org/10.2196/mhealth.6309
57. Shan, R., Sarkar, S., Martin, S.S.: Digital health technology and mobile devices for the management of diabetes mellitus: State of the art. Diabetologia. **62**(6), 877–887 (2019). https://doi.org/10.1007/s00125-019-4864-7

58. Muthu, B., Sivaparthipan, C.B., Manogaran, G., Sundarasekar, R., Kadry, S., Shanthini, A., Dasel, A.: IOT based wearable sensor for diseases prediction and symptom analysis in healthcare sector. Peer-to-Peer Netw. Appl. **13**(6), 2123–2134 (2020). https://doi.org/10.1007/s12083-019-00823-2
59. Yen, H.Y.: Smart wearable devices as a psychological intervention for healthy lifestyle and quality of life: A randomized controlled trial. Qual. Life Res. **30**(3), 791–802 (2021). https://doi.org/10.1007/s11136-020-02680-6
60. McDonough, D.J., Su, X., Gao, Z.: Health wearable devices for weight and BMI reduction in individuals with overweight/obesity and chronic comorbidities: Systematic review and network meta-analysis. Br. J. Sports Med. **55**(16), 917–925 (2021) https://doi.org/10.1136/bjsports-2020-103594
61. Makroum, M.A., Adda, M., Bouzouane, A., Ibrahim, H.: Machine learning and smart devices for diabetes management: Systematic review. Sensors. **22**(5), 1843 (2022). https://doi.org/10.3390/s22051843
62. Buckingham, S.A., Williams, A.J., Morrissey, K., Price, L., Harrison, J.: Mobile health interventions to promote PA and reduce sedentary behavior in the workplace: A systematic review. Dig. Health. **5** (2019). https://doi.org/10.1177/2055207619839883
63. Silveira, F.D., Rodeghiero Neto, I., Santos, B.M.D., Gasparetto, R.M.D.O., Machado, F.M., Rodrigues, P.C.C., Amaral, F.G.: Industry 4.0 perspectives in the health sector in Brazil. Ind. J. Manag. Prod. **1**(12), 1–14 (2021). https://doi.org/10.14807/ijmp.v12i1.1289
64. Greenfield, R., Busink, E., Wong, C.P., Riboli-Sasco, E., Greenfield, G., Majeed, A., Car, J., Wark, P.A.: Truck drivers' perceptions on wearable devices and health promotion: A qualitative study. BMC Public Health. **16**(1), 1–10 (2016). https://doi.org/10.1186/s12889-016-3323-3
65. Torbjørnsen, A., Ribu, L., Rønnevig, M., Grøttland, A., Helseth, S.: Users' acceptability of a mobile application for persons with type 2 diabetes: A qualitative study. BMC Health Serv. Res. **19**(1), 1–14 (2019). https://doi.org/10.1186/s12913-019-4486-2
66. Muaremi, A., Arnrich, B., Tröster, G.: Towards measuring stress with smartphones and wearable devices during workday and sleep. BioNanoScience. **3**(2), 172–183 (2013). https://doi.org/10.1007/s12668-013-0089-2
67. Giddens, L., Leidner, D., Gonzalez, E.: The role of Fitbits in corporate wellness programs: Does step count matter? In: Proceedings of the 50th Hawaii International Conference on System Sciences (2017). http://hdl.handle.net/10125/41596
68. Maltseva, K.: Wearables in the workplace: The brave new world of employee engagement. Bus. Horiz. **63**(4), 493–505 (2020). https://doi.org/10.1016/j.bushor.2020.03.007
69. Han, M., Cho, B., Kwon, H., Son, K., Lee, H., Lee, J., Park, J.: A mobile-based comprehensive weight reduction program for the workplace (Health-On): Development and pilot study. JMIR Mhealth Uhealth. **7**(11), e11158 (2019). https://doi.org/10.2196/11158
70. Salesforce Research.: Putting wearables to work: special report 2 (2015). https://a.sfdcstatic.com/content/dam/www/ocms/assets/pdf/misc/StateOfWearablesReport.pdf. Last accessed 2022/09/18
71. Khakurel, J., Melkas, H., Porras, J.: Tapping into the wearable device revolution in the work environment: A systematic review. Inf. Technol. People. **31**(3), 791–818 (2018). https://doi.org/10.1108/ITP-03-2017-0076
72. Munayat, G.: Impact of health applications on the Indian health insurance companies. Thesis (2018). https://urn.fi/URN:NBN:fi:amk-201805168582
73. Bent, B., Cho, P.J., Wittmann, A., Thacker, C., Muppidi, S., Snyder, M., Matthew, J.C., Feinglos, M., Dunn, J.P.: Non-invasive wearables for remote monitoring of HbA1c and glucose variability: Proof of concept. BMJ Open Diabetes Res. Care. **9**(1) (2021) https://doi.org/10.1136/bmjdrc-2020-002027
74. Stephenson, A., Garcia-Constantino, M., McDonough, S.M., Murphy, M.H., Nugent, C.D., Mair, J.L.: Iterative four-phase development of a theory-based digital behavior change intervention to reduce occupational sedentary behavior. Dig. Health. **6**, 2055207620913410 (2020). https://doi.org/10.1177/2055207620913410

75. Chang, R.C.S., Lu, H.P., Yang, P., Luarn, P.: Reciprocal reinforcement between wearable activity trackers and social network services in influencing PA behaviors. JMIR Mhealth Uhealth. **4**(3) (2016). https://doi.org/10.2196/mhealth.5637
76. Finkelstein, E.A., Haaland, B.A., Bilger, M., Sahasranaman, A., Sloan, R.A., Nang, E.E.K., Evenson, K.R.: Effectiveness of activity trackers with and without incentives to increase PA (TRIPPA): A randomized controlled trial. Lancet Diab. Endocrinol. **4**(12), 983–995 (2016) https://doi.org/10.1016/S2213-8587(16)30284-4
77. Pancardo, P., Acosta, F.D., Hernández-Nolasco, J.A., Wister, M.A., López-de-Ipiña, D.: Real-time personalized monitoring to estimate occupational heat stress in ambient assisted working. Sensors. **15**(7), 16956–16980 (2015) https://doi.org/10.3390/s150716956
78. Patterson, P.D., Smith, K.J., Hostler, D.: Cost-effectiveness of workplace wellness to prevent cardiovascular events among US firefighters. BMC Cardiovasc. Disord. **16**(1), 1–7 (2016). https://doi.org/10.1186/s12872-016-0414-0
79. Soliño-Fernandez, D., Ding, A., Bayro-Kaiser, E., Ding, E.L.: Willingness to adopt wearable devices with behavioral and economic incentives by health insurance wellness programs: Results of a US cross-sectional survey with multiple consumer health vignettes. BMC Public Health. **19**(1), 1–8 (2019). https://doi.org/10.1186/s12889-019-7920-9
80. Baldassarre, A., Mucci, N., Padovan, M., Pellitteri, A., Viscera, S., Lecca, L.I., Galea, R.P., Arcangeli, G.: The role of electrocardiography in occupational medicine, from Einthoven's invention to the digital era of wearable devices. Int. J. Environ. Res. Public Health. **17**(14), 4975 (2020). https://doi.org/10.3390/ijerph17144975
81. Wang, J.B., Cadmus-Bertram, L.A., Natarajan, L., White, M.M., Madanat, H., Nichols, J.F., Ayala, G., Pierce, J.: Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase PA in overweight and obese adults: A randomized controlled trial. Telemed. J. E Health. **21**(10), 782–792 (2015). https://doi.org/10.1089/tmj.2014.0176
82. Smuck, M., Odonkor, C.A., Wilt, J.K., Schmidt, N., Swiernik, M.A.: The emerging clinical role of wearables: Factors for successful implementation in healthcare. NPJ Dig. Med. **4**(1), 1–8 (2021). https://doi.org/10.1038/s41746-021-00418-3
83. Stefanicka-Wojtas, D., Kurpas, D.: eHealth and mHealth in chronic diseases—Identification of barriers, existing solutions, and promoters based on a survey of EU stakeholders involved in Regions4PerMed (H2020). J. Pers. Med. **12**(3), 467 (2022). https://doi.org/10.3390/jpm12030467
84. Ahmed, S.A., Moore, M.B.J.: Health information technology implementation challenges and responsive solutions for health systems. Int. J. Mod. Res. Eng. Technol. **2**(6), 1–8 (2017)
85. Kauffman, M.E., Soares, M.N.: New technologies and data ownership: Wearables and the erosion of personality rights. Revista Direitos Sociais e Políticas Públicas (UNIFAFIBE). **6**(1), 512–538 (2018) https://doi.org/10.25245/rdspp.v6i1.444
86. Allman-Farinelli, M., Chen, J.: mHealth technologies in the management of obesity: A narrative review. Smart Homecare Technol. TeleHealth. **4**, 53–59 (2017)
87. McAuley, E., Blissmer, B.: Self-efficacy determinants and consequences of physical activity. Exerc. Sport Sci. Rev. **28**(2), 85–88 (2000) https://www.researchgate.net/publication/12417418
88. Clim, A., Răzvan, D.Z.: Game theory in designing mHealth apps for monitoring hypertension. Manag. Market. **14**(2), 220–231 (2019). https://doi.org/10.2478/mmcks-2019-0015
89. Engbers, L., Sattelmair, J.: Monitoring and Evaluation of Worksite Health Promotion Programs: Current State of Knowledge and Implications for Practice. World Health Organization (WHO) (2008)
90. Phillips, S.M., Cadmus-Bertram, L., Rosenberg, D., Buman, M.P., Lynch, B.M.: Wearable technology and PA in chronic disease: Opportunities and challenges. Am. J. Prev. Med. **54**(1), 144–150 (2018). https://doi.org/10.1016/j.amepre.2017.08.015
91. Ray, P.P.: A survey on Internet of Things architectures. J. King Saud Univ. Comput. Inf. Sci. **30**(3), 291–319 (2018). https://doi.org/10.1016/j.jksuci.2016.10.003

Catastrophic Health Spending by COVID-19 in the Mexican Insurance Sector

Ulises Domínguez-Gutiérrez and Román Rodríguez-Aguilar

Abstract The COVID-19 pandemic that the world has been suffering for 3 years has generated major impacts worldwide, both in public health systems and in the private insurance industry. The high costs of care derived from cases with complications have likewise generated a great impact on the private insurance industry. In the case of Mexico, the mortality rates observed are among the first places, in addition to generating a great impact on private insurance. This work deals with the measurement of the impact of catastrophic expenses derived from COVID-19 in an insurance company; using a set of machine learning models, the key variables in the estimation of patients with potential catastrophic expenses were determined. The results show that the estimated classification model has a positive performance in addition to allowing the identification of the main risk factors of the insured as well as their potentially catastrophic impact on insurance companies.

Keywords Machine learning · COVID-19 · Private insurance · Catastrophic health spending · México

1 Background

Nowadays, there is a SARS-CoV-2 spread around the world. The virus is responsible for causing COVID-19, and most people who experience symptoms (around 80%) get better without the necessity of hospital care. Around 15% experience serious illness and need oxygen as part of the treatment, while 5% achieve a critical status and need intensive care. In some cases, death can be possible [1]. Until November

U. Domínguez-Gutiérrez
Facultad de Ingeniería, Universidad Anáhuac México, Naucalpan, Mexico

R. Rodríguez-Aguilar (✉)
Facultad de Ciencias Económicas y Empresariales, Universidad Panamericana, Mexico City, Mexico
e-mail: rrodrigueza@up.edu.mx

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_14

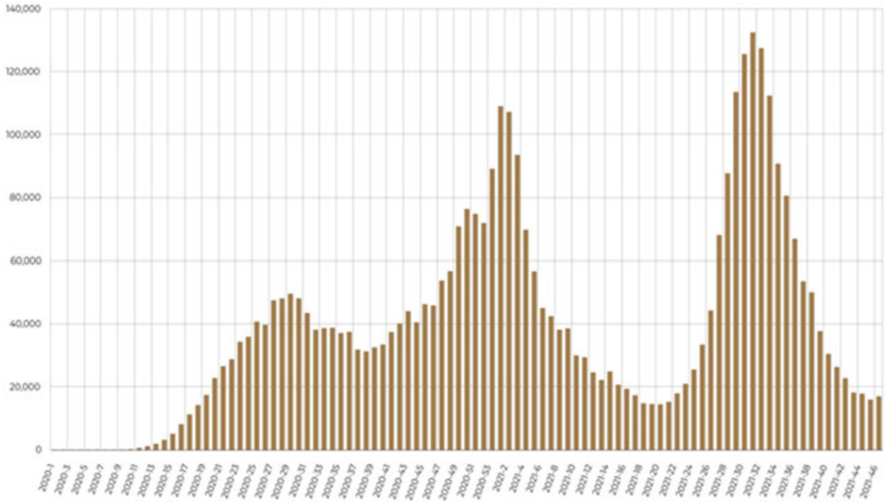


Fig. 1 Evolution of confirmed cases of COVID-19 in Mexico according to the epidemiological week

16, 2021, there were 3,847,243 cumulative COVID cases in Mexico (see Fig. 1). From these total cases, 1.23% were persons with medical insurance, and there were 44,358 cover cases with an average cost of medical care of 506,981 Mexican pesos. The most expensive case was for 52,541,000 Mexican pesos [2].

Since the beginning of the pandemic, the medical insurance industry was in favor of supporting insureds who had COVID. On March 2020, the agency in charge of monitoring the insurance sector, the CNSF (for its acronym in Spanish), and the entity in charge of promoting the development of the insurance industry, the AMIS (for its acronym in Spanish), established an agreement for implementing regulatory facilities to support insureds to affront the illness. With the agreement, the most important temporary change in insurance terms and conditions was to eliminate the illness excluded due to pandemics [3]. After 19 months, the pandemic situation has been the most expensive event in Mexican private insurance history [2].

Derived from the above, it is necessary to have reliable and timely information on the potential impact of new patients diagnosed with COVID-19. Machine learning is a current tool that has been taking importance in many fields of science. It has motivated researchers to use techniques to help with the diagnosis and prediction of mortality rates around critical COVID-19 cases. By identifying probable critical cases, it is possible to reduce the mortality rate and the probability of complications and observe a decrease in the financial impact caused by medical care.

There have been some developments in machine learning with efficient algorithms for the detection and diagnosis of many illnesses, including COVID-19. Recent research with decision trees, logistic regression, random forest, and neural networks are some examples that have been used to predict mortality rate on COVID-19 cases. These algorithms were trained with the data from patients from

146 countries and gave good results. The deep neural network was the one with the best results regarding evaluation metrics. It achieved an accuracy of 0.970, precision and specificity of 1, sensitivity of 0.970, and F1 score of 0.985 [4].

In addition, machine learning models have been developed to predict the health, behavior, and performance of treatment to calculate the cost of possible therapies to the patients. The health of people who take medications is affected by a variety of factors. These factors can include conditions, genetics, and lifestyle behaviors, such as smoking, weight changes, and medication adherence, among others. A recent study uses these factors to predict the cost of the treatment to achieve a healthier lifestyle with the main beneficiaries being the patients, healthcare providers, and insurance companies because they would have the treatment and cost program for a period according to the characteristics of patients [5]. For this reason, using machine learning in the insurance sector estimating efficiently the cost of COVID-19 could generate positive results that impact the profit of the business, as well as help the attention of the insured.

The economic impact of COVID-19 on private insurance generated a greater impact on the field of medical and life insurance, showing high sinistrality because of expenses around medical care and deaths. So far, now the number of claims is starting to be stable because of available vaccines for the Mexican population; nevertheless, new COVID-19 variants and post-COVID-19 conditions are responsible for the still risk of occurrence and claims for COVID-19 in the insurance industry. The attention to this kind of claim will continue for an undefined time.

The objective of this study is to develop machine learning models, using medical expense insurance data to classify catastrophic claims and predict the cost that medical care will have due. The database used for this case contains historical information belonging to a Mexican insurance company. This data considers all claims for COVID-19 from March 2020 to March 2022. The considered data includes historical sinistrality and risk factors like age and gender, among others. The analysis also considers chronic conditions insureds have before and after buying the insurance, conditions that can result in a higher probability of having COVID with greater affections, like diabetes and hypertension, among others. The results will permit to identify and manage catastrophic claims, making it possible to have to carry on medical care insureds, reducing sinistrality, and therefore promoting better company results, which were affected since 2020. The work is organized as follows: Section 2 corresponds to the applied methodology, the third section deals with the main results of the modeling, and finally the last section presents the conclusions.

2 Methods

The main goal of the study is to generate machine learning model that allow classifying and estimating the cost of catastrophic claims due to COVID-19 based on the insured's risk factors, their morbidities, and historical claims. The foregoing will allow the monitoring and management of cases to care for the insured, the reserve, and the accident rate generated by those claims, therefore improving the company's results.

2.1 Algorithms Implemented

As mentioned above, machine learning models effective in detecting hidden connections have been developed to predict the cost of various diseases. These algorithms use mathematical analysis to identify patterns that cannot be detected by traditional exploration, because the relationships are too complex or because there is too much data. Next, the conceptual aspects of the models used are briefly explained.

Ridge Regression
Ridge regression was proposed by Hoerl and Kennard in 1970. This regression penalizes the sum of the squared coefficients. This penalty is known as L2 norm and has the effect of reducing the value of all the coefficients of the model without reaching zero. The degree of penalty is controlled by the hyperparameter λ. When $\lambda = 0$, the penalty is null, and the result is equivalent to that of a linear model by ordinary least squares. As λ increases, the greater the penalty and the smaller the value of the predictors. The Ridge estimator is defined in Eq. 1:

$$\widehat{\beta}_R = \arg \min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \left(\beta_j\right)^2 \tag{1}$$

The Ridge estimator is a linear transformation of the least squares estimator of the residuals, whose bias increases with increasing λ, but at the same time its variance decreases [6].

Lasso Regression
The Lasso linear model estimation method was proposed by Tibshirani. It consists of the estimation by ordinary least squares, restricting the absolute sum of the regression coefficients [7]. The Lasso estimator is defined in Eq. 2:

$$\widehat{\beta_L} = \arg \min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \left|\beta_j\right| \tag{2}$$

Lasso penalizes the sum of the absolute value of the regression coefficients. This penalty is known as L1 norm and has the effect of forcing the coefficients of the predictors to approach zero. Given that a predictor with a zero-regression coefficient does not influence the model, Lasso manages to exclude the least relevant predictors, considering only the predictors that influence the response variable in the model. Like Ridge, the degree of penalty is controlled by the hyperparameter λ. When $\lambda = 0$, the result is equivalent to that of a linear model by ordinary least squares. As λ increases, the penalty is greater, and more predictors are excluded.

Elastic Networks

The elastic networks were proposed by Zou, and Hastie tries to combine the advantages of Ridge and Lasso regression, that is, it combines the L1 and L2 penalties. The degree to which each of the penalties influences is controlled by the hyperparameter α. Its value is included in the interval [0,1]. When $\alpha = 0$, the Ridge is applied, and when $\alpha = 1$, Lasso is applied. The elastic estimator is defined in Eq. 3:

$$\mathrm{EN}\,(\beta) = \sum_{i=1}^{n}\left(\gamma_i - x_i^T\beta\right)^2 + \lambda_1\sum_{j=1}^{p}\left|\beta_j\right| + \lambda_2\sum_{j=1}^{p}\left(\beta_j\right)^2 \tag{3}$$

Due to Ridge regularization, the elastic network estimator can handle the correlations between predictors better than a Lasso, and due to L1 regularization, variables are reduced [8].

Random Forest

A random forest is an advanced implementation of a bootstrap aggregation algorithm with a tree model as the base model. In random forests, each tree in the ensemble is generated from a sample drawn with substitution from the training set. By splitting a node during tree construction, the chosen split is no longer the best split among all features. Instead, the split that is chosen is the best split among a random set of features. Due to this randomness, as a rule, the bias of the forest increases slightly (concerning the bias of a single non-random tree), but, due to averaging, its variance also decreases, usually more than compensating for the increase in bias, and for this reason, generates a better model in general [9].

Neural Networks

A neural network is a simplified model that emulates the way the human brain processes information. The basic units are neurons, which are generally organized in layers. The processing units are organized in layers. There are normally three parts to a neural network: an input layer, with units representing the input fields; one or more hidden layers; and an output layer, with a unit or units representing the destination field or fields. Units connect with variable connection strengths (or weights). The input data is presented in the first layer, and the values are propagated from each neuron to each neuron in the next layer. In the end, a result is sent from the output layer.

The network learns by examining the individual records, generating a prediction for each record, and adjusting the weights when it makes an incorrect prediction. This process is repeated many times, and the network continues to improve its predictions until one or more stopping criteria have been met [10].

Gradient Boosting Machine
Gradient boosting machine is a direct learning ensemble method. The guiding heuristic is that good predictive results can be obtained through increasingly refined approximations. GBM trains many regression trees incrementally, additively, and sequentially. It is a method that combines the efforts of multiple trees to create a better model, and with each tree that is added, the mean square error of the overall model is reduced [11].

Stacked Ensembles
The stacked ensemble method is a supervised ensemble machine learning algorithm that finds the optimal combination of a collection of prediction algorithms through a process called stacking. The stacked ensemble can be viewed as a more sophisticated form of cross-validation, as it combines models to estimate and then corrects the error of one individual model with another. Therefore, these ensemble methods use multiple learning algorithms to obtain a better predictive performance than could be obtained from any of the learning algorithms individually, since they are a means to estimate and correct the biases of the individual models [12].

2.2 Treatment of Unbalanced Data

In some classification problems, the problem of imbalance in the response variable can arise. This implies that there is an important bias in some of the categories to be predicted. Class imbalance is a disproportionate representation of one of the target classes. To solve the bias problem, it is necessary to apply additional sampling methods. Oversampling involves randomly selecting examples from the minority class, with replacement, and adding them to the training data set making it equal to the majority class. Under sampling involves randomly selecting examples of the majority class and removing them from the training data set by making them equal to the minority class. Sometimes the combination of both random sampling methods can result in improved overall performance compared to the independent methods.

2.3 Data Description

The database on which the study was developed is historical information that considers all care for COVID-19 from March 2020, the first month where expenses were generated for the said condition, until March 2022. This considers characteristics of the insured, from hiring and care (see Table 1). Likewise, the catastrophic claim indicator was constructed by identifying claims with a cost greater than 1 million pesos (90th percentile).

Table 1 shows the four categories into which the variables of the database are divided with a total of 13,018 observations to classify the catastrophic claims of the disease.

Table 1 Variables by category

Category	# Variables
Insured and comorbidities	27
Hiring	16
Attention	8

3 Results

The database considers historical information on COVID-19 claims from March 2020 to March 2022. Data also consider some characteristics of the insured, the medical services, and the conditions of the insurance. Additionally, a catastrophic claim index was calculated by identifying claims with a cost greater than a million pesos (90th percentile). After univariate and bivariate analysis, 27 variables were selected and 13,017 observations with balanced classes. In Table 2a summary of this data is displayed.

Table 3 displays the percentage of accuracy, sensitivity, and specificity reached by each model. Lasso model that considers 16 variables is the best with an accuracy of 83.6% to classify catastrophic COVID-19 claims.

Different algorithms such as regressions, elastic net, random forest, and deep neural network were applied and evaluated to select the final model. The Lasso regression model was used as a variable selection mechanism, so the random forest and neural network models were estimated based on the variables selected by the Lasso model. The selected model is the one that presented the best performance in the three calculated performance metrics. It is important to mention that all models were replicated considering an additional variable, the number of the wave at which the care was given. This variable was integrated into the analysis to identify if the type of virus of the vaccines could modify the results. However, it did not generate a representative change.

4 Conclusions

COVID-19 has generated a great impact on the insurance industry stability, and the future is still uncertain due to the new type of virus and associated conditions to the illness which can increase the risk of having health issues in the long term so the development of both models will allow the company to identify, measure, and manage the financial impact.

The proposed model classifies catastrophic cases, and this generates an opportunity for the insurance company to monitor the medical care of the insured in all stages of the condition guaranteeing health and service for the patient and, in addition, to have control of the claim cost. Robust estimates of potential catastrophic cases due to COVID-19 allow efficient management of the insurance company's resources. Follow-up and punctual healthcare for patients who develop severe complications from COVID-19, will allow the focus of the resources on

Table 2 Variables considered for modeling

Category	Variable	Type	Description
Insured – health conditions	Gender	Categorical	Insured's gender
	Age	Numerical	Age at the time of claim
	City	Categorical	City of residence
	Conditions_amount	Numerical	Amount spent on previous conditions
	Conditions_claims	Numerical	Number of previous condition claims
Plan conditions	Agent	Categorical	Agent
	Pay	Categorical	Payment method
	Branch	Numerical	Branch office
	Business	Categorical	Insured business
	Tabulator	Categorical	Medical tabulator
	Type_plan	Categorical	Plan type
	Plan	Categorical	Plan
	Level_hosp	Categorical	Hospital level
	Sa	Numerical	Sum assured
	Currency_sa	Categorical	Sum currency
	Insured	Categorical	Indicator of insured person or dependency
	Deductible	Numerical	Deductible
	Currency_ded	Categorical	Deductible currency
	Coinsurance	Numerical	Coinsurance
Service health	Modality	Categorical	Type of care
	Hospital	Categorical	Hospital
	City_hospital	Categorical	Hospital city
	Level_ant	Categorical	Hospital's level 1
	Level_act	Categorical	Hospital's level 2
	Doctor	Categorical	Doctor's name
	Procedure	Categorical	Procedure performed
Response variable	Amount	Categorical	Catastrophic claim indicator

Table 3 Evaluation measures

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ridge (balanced weights)	81.18	63.67	83.09
Lasso (balanced weights)	**83.60**	**80.47**	**83.94**
Elastic (balanced weights)	83.83	78.91	84.37
Random forest (both-sample)	86.52	57.42	89.69
Random forest (variables Lasso and both-sample)	84.95	60.16	87.65
Neural network (variables Lasso and both-sample)	94.62	62.89	98.08

severe cases, and therefore increase the probability of improved health status of the patients. The selected machine learning model to strengthen the company's leadership within the insurance market, improving the reserve calculation process for a new condition. In addition, the insurance company provides personalized attention and service taking care of client health and the company's profit.

The study was developed with the latest machine learning algorithms. However, their improvement will occur in the coming months with more information available. The proposed study is feasible to replicate in any insurer that has the data used, which will allow the proposed methodology to be feasible to implement in any interested insurer.

References

1. World Health Organization: Basic Information About COVID-19. WHO. Recover 2021 December 1, https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19 (n.d.)
2. Asociación Mexicana de Instituciones de Seguros: COVID Conference, November 18. AMIS. Recover 2021 December 1, from https://www.amisprensa.org/wp-content/uploads/2021/11/Conferencia-COVID-19-61-V2.pdf (2021, November 18)
3. CONDUSEF: Confused. Recover 2021 December 1, from https://www.condusef.gob.mx/?p=contenido&idc=1356&idcat=1 (s. f.)
4. Khan, I.U., Aljabri, M., Aljameel, S.S., Mariam Moataz, A.K., Alshamrani, F.M., Sara Mhd, B.C.: Computational intelligence-based model for mortality rate prediction in COVID-19 patients. Int. J. Environ. Res. Public Health. **18**(12), 6429 (2021). https://doi.org/10.3390/ijerph18126429
5. Kutzko, J.D.: Computer-implemented systems and methods for predicting the health and therapeutic behavior of individuals using artificial intelligence, smart contracts, and blockchain. Google Patents. Recuperado 1 de diciembre de 2021, de https://patents.google.com/patent/US20210241918A1/en?q=(G06N5%2f00)+(MODEL+ESTIMATION+COST)(COVID)&oq=(G06N5%2f00)+(MODEL+ESTIMATION+COST)(COVID) (2018, 18 mayo)
6. Hoerl, A.E., Kennard, R.W.: Ridge regression: biased estimation for nonorthogonal problems. Technometrics. **12**, 55–67 (1970)
7. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. **58**(1), 267–288 (1996)
8. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. R. Stat. Soc. Series B Stat. Methodol. **67**(2), 301–320 (2005). JSTOR, http://www.jstor.org/stable/3647580. Accessed 24 May 2022
9. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
10. IBM: El modelo de redes neuronales. Recuperado 19 de septiembre de 2021, de https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=networks-neural-model (s. f.-a)
11. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**, 1189–1232 (2001)
12. Wolpert, D.H.: Stacked generalization. Neural Netw. **5**, 241–259 (1992)

Modeling and Computer Simulation of Nanocomplexation for Cancer Therapy

Tatiana Romanova, Anna Grebinyk, Alexander Pankratov, Yuri Stoyan, Alina Nechyporenko, Yuriy Prylutskyy, Igor Grebennik, and Marcus Frohme

Abstract Smart systems and data-driven services have the potential to answer medical needs in nanomedicine to get faster to the clinic and manufacturing stage. As an example from galenics, we explored the molecular packing within nanoparticles composed of fullerene C_{60} (C_{60}) and the anthracycline antibiotic doxorubicin (Dox); the nanoparticle was previously proven to be a promising candidate for photodynamic and chemotherapeutic treatment of cancer. The Dox-C_{60} hybrid was evaluated to be around 135 nm and forms an aqueous monodisperse colloid solution. We consider a non-standard packing problem for the geometric design of the Dox-C_{60} nanocomplex. Each placement object was a disconnected set with a core sphere and two identical spherical components that were allowed moving only along the core sphere orbit at the given distance. Allowable distances between each pair of components, as well as between objects, were given. All disconnected objects could be freely moving within the given volume (cuboid). The packing

T. Romanova (✉) · A. Pankratov
A. Pidhornyi Institute for Mechanical Engineering Problems of the National Academy of Sciences of Ukraine, Kharkiv, Ukraine

Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

A. Grebinyk · M. Frohme
Technical University of Applied Sciences Wildau, Wildau, Germany

Y. Stoyan
A. Pidhornyi Institute for Mechanical Engineering Problems of the National Academy of Sciences of Ukraine, Kharkiv, Ukraine

A. Nechyporenko
Technical University of Applied Sciences Wildau, Wildau, Germany

Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

Y. Prylutskyy
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

I. Grebennik
Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_15

problem was aimed to maximize the number of disconnected objects that could be fully arranged inside the given volume, considering distance constraints. The problem was formulated as MIP (mixed integer problem). A solution strategy was proposed that combines multistart strategy, nonlinear optimization, and decomposition algorithm. Computational results for 2D and 3D objects were provided.

Our findings may contribute to solve molecular packaging problems which play a role in synthesis, upscaling, production as well as formulation, and medication of the complexed drug.

Keywords Geometric design · Packing · Phi-function technique · Doxorubicin · Nanocomplex

1 Introduction

Nineteen Mio new cases of cancer and ten Mio deaths worldwide in 2020 [1] rank this noncommunicable disease as the single most important barrier that is preventing an expected increase of life expectancy in the twenty-first century and as one of the most difficult global healthcare problems. Despite considerable advances in diagnosis and therapy, mortality rates have demonstrated no improvement due to the increased incidence rates over the last few decades [2, 3]. Chemotherapy is one of the most common type of treatment [4] that aims to slow down the growth of cancer cells that evolve to divide quickly [5, 6]. The first generation of anticancer chemotherapeutic agents made use of a single toxic drug such as the anthracycline antibiotic doxorubicin (Dox) [7]. A common feature of a large number of anticancer drugs, including Dox, is the ability to hinder DNA synthesis and therefore cell division [5]. However efficient clinical chemotherapy is challenged with a number of setbacks, including poor specificity, high systemic toxicity, and an induced drug resistance of the cancer cells [8, 9]. So, new approaches to improve efficiency and attenuate side effects in cancer chemotherapy are urgently needed.

The advent of nanomedicine and application of biocompatible, bioavailable, and nontoxic nanoparticles brought significant advances in cancer chemotherapy offering customizable and safer treatment options. Thus, nanomedicine formulations for drug delivery could pave a way to overcome limitations and improve treatment efficacy of the chemotherapeutic drugs. The combination of carbon nanoparticle C_{60} fullerene's (C_{60}) antioxidant potential [10–13] and its ability for drug delivery makes it very attractive for anticancer therapy. A pioneering attempt [14] showed a simple and fast method of C_{60} noncovalent complexation with Dox in water and later in physiological solution [15]. Molecular modeling, spectroscopy, atomic force microscopy, dynamic light, and small-angle X-ray scattering evidenced doxorubicin-C_{60} nanocomplex (Dox-C_{60}) formation [14, 15]. The proposed nanomedicine formulation was shown to have higher toxicity compared to the free drug against various human tumor cell lines in vitro [16–18]. In another approach, an antimicrobial effect and the applicability of noncovalent Dox-C_{60} for in vivo imaging was shown [19]. C_{60} nanomolar concentrations were found to be efficient for the enhancement of Dox's pro-oxidant and proapoptotic activities in vitro that pointed toward a further promising application of C_{60}-based nanocomplexes for optimization of drug's efficiency against cancer cells [20].

As currently more than 50 nanomedicines have been already approved for clinical use with more than 400 nanomedicine formulations being evaluated in clinical trials [21], the further support of nanomedicine development with in silico modeling has a potential to answer unmet medical needs and help the field get faster to the clinic and manufacturing stage. The development of nanomedicine formulations relies on diverse complex data from both in vitro and in vivo studies. Therefore, it requires new interdisciplinary strategies to overcome data complexity and heterogeneity. In this regard, in silico models are indispensable to explore the molecules within C_{60} nanocomplex with an anthracycline antibiotic Dox that can be achieved with the simulation of its geometric design. Over the past decade, in silico modeling has demonstrated a tremendous potential in cancer nanomedicine [22, 23]. It allows researchers not only to reduce costs of experimental work but also to increase research efficiency using computer-based modeling tools such as operations research (OR). In the context of the anticancer drug delivery, OR opens up new opportunities for smart nanocomplex design by solving the packing problems of molecules.

A lot of publications address the sphere packing problems (see, e.g., [24–30] and the reference therein) that is proposed to characterize the geometric design of the Dox-C_{60} nanocomplex. Sphere packing is an NP-hard problem (non-deterministic polynomial time) [31], and thus heuristic approaches are widely used to obtain good feasible solutions. Packing spherical objects as nonlinear optimization problems using the phi-function technique for modeling placement constraints are considered, e.g., in [28, 32–36].

One of the major challenges of achieving the transition of nanoparticle-based therapeutics use into clinical practice is the complete characterization of these materials composition that would enable scaling-up and smart manufacturing. For that we estimate the size distribution and solve the packing problem of C_{60} and Dox molecules within proposed aqueous monodisperse colloid solution of Dox-C_{60} nanocomplex.

2 Methods

2.1 C_{60} and Dox-C_{60} Complex Synthesis

The pristine C_{60} aqueous colloid solution was prepared by C_{60} transfer from toluene to water using continuous ultrasound sonication as described by [37, 38]. The obtained C_{60} water colloid solution had a final concentration of 150 μg/ml with 99% purity, stability, and homogeneity and an average nanoparticle's size of 100 nm.

Dox was dissolved in water at initial concentration in 18.5 and 2.8 mm, respectively. A nanocomplex with C_{60} was prepared according to the protocol [15]. Briefly, the Dox solutions were mixed with C_{60} colloid solution in 2:1 molar ratio of the components. The mixtures were treated in an ultrasonic disperser for 30 min, stirred magnetically for 24 h at room temperature, and centrifuged at 4000 g for 15 min with the use of centrifuge filters Amicon Ultra-0.5 3 K for sample purification. The stability (optical spectra and size distribution) of nanocomplexes was systematically controlled and shown to be practically unchanged after 6 months of storage in physiological saline solution.

2.2 Dynamic Light Scattering (DLS) Measurement

Short ultrasonication (30 s, 35 kHz) was applied to remove air bubbles. Size distribution of Dox-C_{60} nanocomplex aqueous colloid solution was evaluated with a Zetasizer Nano S equipped with a He-Ne 633 nm laser (Malvern Instruments, UK). Data were recorded at 37 °C in backscattering mode at a scattering angle of 173°. Dox-C_{60}, placed in disposable polystyrene cuvettes, was measured 15 times to establish average diameters and intensity distributions. The autocorrelation function of the scattered light intensity was analyzed by the Malvern Zetasizer Software (Malvern Instruments, UK) with the Smoluchowski approximation.

2.3 Mathematical Assumptions

Let a collection of disconnected objects T_i with three connected components $i \in I_n = \{1, 2, \ldots, n\}$ be given. Each object can be presented in the form $T_i = A_i \cup B_{i1} \cup B_{i2}$ (simplified model of the nanocomplex) (Fig. 1), where A_i, B_{i1}, B_{i2} are disjoint spheres. The sphere A_i is defined by its radius r (simplified model of C_{60} molecule), while two identical spheres B_{ik}, $k = 1, 2$, have the same radius $R > r$ (simplified models of Dox molecules). Let $v_i = (x_i, y_i, z_i)$ be a variable motion vector of the object T_i, that arranged at the center of the sphere A_i, $i \in I_n$. We denote centers of the components B_{ik} by $v_{ik} = (x_{ik}, y_{ik}, z_{ik})$, $k = 1, 2$, where

$$x_{ik} = x_i + \rho \sin\theta_{ik} \cos\varphi_{ik},\ y_{ik} = y_i + \rho \sin\theta_{ik} \sin\varphi_{ik}, \tag{1}$$

$$z_{ik} = z_i + \rho \cos\theta_{ik}, 0 \le \theta_{ik} \le \pi, 0 \le \varphi_{ik} < 2\pi,\ \rho = \Delta_{ab} + r + R,$$

$\Delta_{ab} > 0$ is a given Euclidean distance between components A_i and B_{ik}.

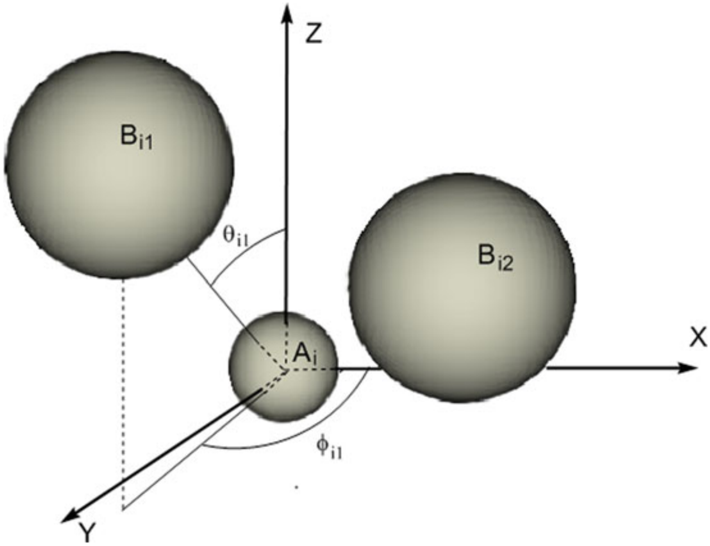


Fig. 1 Disconnected object $T_i = A_i \cup B_{i1} \cup B_{i2}$ – a simplified model of the Dox-C_{60} nanocomplex

In addition, the minimal allowable distance Δ_{aa} between spheres $A_i \subset T_i$ and $A_j \subset T_j$ as well as the minimal allowable distance Δ_{bb} between components B_{i1} and B_{i2} are claimed.

A cuboid $\Omega = \{(x, y, z)|0 \leq x \leq l, 0 \leq y \leq w, 0 \leq x \leq h\}$ is considered as a container.

The non-standard optimization packing problem is aimed to arrange the maximum number of disconnected objects T_i, $i \in I_n$, into a cuboid Ω providing the following geometric constraints:

- *Distance conditions*

$$\text{dist}\left(A_i\left(v_i\right), A_j\left(v_j\right)\right) \geq \Delta_{aa}, \text{for } i > j \in I_n, \tag{2}$$

$$\text{dist}\left(B_{i1}\left(v_i, \theta_{i1}, \varphi_{i1}\right), B_{i2}\left(v_i, \theta_{i2}, \varphi_{i2}\right)\right) \geq \Delta_{bb}, \text{for } i \in I_n, \tag{3}$$

$$\text{dist}\left(A_i\left(v_i\right), B_{ik}\left(v_i, \theta_{ik}, \varphi_{ik}\right)\right) = \Delta_{ab}, \text{for } k = 1, 2, i \in I_n, \tag{4}$$

$$\text{dist}\left(B_{ik}\left(v_i, \theta_{ik}, \varphi_{ik}\right), B_{jk}\left(v_j, \theta_{jk}, \varphi_{jk}\right)\right) \geq \Delta_{bb}, \text{for } k = 1, 2, i > j \in I_n; \tag{5}$$

- *Containment conditions*

$$T_i\left(t_i\right) \subset \Omega \iff \text{int}\, T_i\left(t_i\right) \cap \Omega^* = \varnothing, \text{for } i \in I_n, \tag{6}$$

where $t_i = (v_i, \theta_{i1}, \varphi_{i1}, \theta_{i2}, \varphi_{i2})$ is a vector of variables of the object T_i, $v_i = (x_i, y_i, z_i)$, $\Omega^* = R^3 \setminus \text{int}\, \Omega$.

For analytical descriptions of the conditions (2)–(5), we use the phi-function technique [28, 32–36].

2.4 Adjusted Phi-Functions for Distance Constraints

Let $T_i(t_i) = A_i(v_i) \cup B_{i1}(t_i) \cup B_{i2}(t_i)$ and $T_j(t_j) = A_j(v_j) \cup B_{j1}(t_j) \cup B_{j2}(t_j)$ be given, where $t_i = (v_i, \theta_{i1}, \varphi_{i1}, \theta_{i2}, \varphi_{i2})$, $t_j = (v_j, \theta_{j1}, \varphi_{j1}, \theta_{j2}, \varphi_{j2})$. In addition scaling parameter $0 \leq \lambda \leq 1$ for each object $T_i(t_i)$ is introduced.

The adjusted phi-function describing distance constraints (2)–(5) for connected components of objects $\lambda T_i(t_i)$ and $\lambda T_j(t_j)$ can be defined as follows:

$$\widehat{\Phi}_{ij}\left(v_i, v_j, \lambda\right) = \left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2 + \left(z_i - z_j\right)^2 - \lambda^2(2r + \Delta_{aa})^2$$

for $\lambda A_i(v_i)$ and $\lambda A_j(v_j)$;

$$\widehat{\Phi}^{i}_{12}(t_i, \lambda) = (x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2 + (z_{i1} - z_{i2})^2 - \lambda^2(2R + \Delta_{bb})^2$$

for $\lambda B_{i1}(v_i, \theta_{i1}, \varphi_{i1})$ and $\lambda B_{i2}(v_i, \theta_{i2}, \varphi_{i2})$;

$$\widehat{\Phi}^{i}_{k}(t_i, \lambda) = (x_i - x_{ik})^2 + (y_i - y_{ik})^2 + (z_i - z_{ik})^2 - \lambda^2\rho^2$$

for objects $\lambda A_i(v_i)$ and $\lambda B_{ik}(v_i, \theta_{ik}, \varphi_{ik})$;

$$\widehat{\Phi}^{kp}_{ij}\left(t_i, t_j, \lambda\right) = \left(x_{ik} - x_{jp}\right)^2 + \left(y_{ik} - y_{jp}\right)^2 + \left(z_{ik} - z_{jp}\right)^2 - \lambda^2(2R + \Delta_{bb})^2$$

for $\lambda B_{ik}(v_i, \theta_{ik}, \varphi_{ik})$ and $\lambda B_{jp}(v_j, \theta_{ip}, \varphi_{jp})$.

Taking into account (1), we have

$$\begin{aligned}
&\widehat{\Phi}^{kp}_{ij}\left(t_i, t_j, \lambda\right) = \left(x_{ik} - x_{jp}\right)^2 + \left(y_{ik} - y_{jp}\right)^2 + \left(z_{ik} - z_{jp}\right)^2 - \lambda^2(2R + \Delta_{bb})^2 = \\
&\left(x_i + \rho \sin\theta_{ik} \cos\varphi_{ik} - x_j - \rho \sin\theta_{jp} \cos\varphi_{jp}\right)^2 + \\
&\left(y_i + \rho \sin\theta_{ik} \sin\varphi_{ik} - y_j - \rho \sin\theta_{jp} \sin\varphi_{jp}\right)^2 + \\
&\left(z_i + \rho \cos\theta_{ik} - z_j - \rho \cos\theta_{jp}\right)^2 - \lambda^2(2R + \Delta_{bb})^2, \\
&\widehat{\Phi}^{i}_{12}(t_i, \lambda) = 2\rho^2 \cdot \Big(1 - \sin\theta_{i1} \sin\theta_{i2} \left(\cos\varphi_{i1} \cos\varphi_{i2} + \sin\varphi_{i1} \sin\varphi_{i2}\right) \\
&- \cos\theta_{i1} \cos\theta_{i2}\Big) - \lambda^2(2R + \Delta_{bb})^2.
\end{aligned}$$

In addition, the function $\widehat{\Phi}^{i}_{k}(t_i, \lambda)$ can be replaced with the adjusted phi-function

$$\widehat{\Phi}^{k}_{ij}\left(t_i, v_j, \lambda\right) = \left(x_{ik} - x_j\right)^2 + \left(y_{ik} - y_j\right)^2 + \left(z_{ik} - z_j\right)^2 - \lambda^2\, \rho^2 =$$
$$\left(x_i + \rho \sin\theta_{ik} \cos\varphi_{ik} - x_j\right)^2 + \left(y_i + \rho \sin\theta_{ik} \sin\varphi_{ik} - y_j\right)^2$$
$$+ \left(z_i + \rho \cos\theta_{ik} - z_j\right)^2 - \lambda^2\, \rho^2,$$

for objects $\lambda A_j(v_j)$ and $\lambda B_{ik}(v_i, \theta_{ik}, \varphi_{ik})$.

Phi-Function for Containing $T_i(t_i)$ into a Cuboid Ω

Phi-function for objects $T_i(t_i)$ and Ω^* describing containment constraints (6) can be defined as follows:

$$\Phi_i\left(t_i, \lambda\right) = \min\left\{\Phi_i\left(v_i, \lambda\right), \Phi_i^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right), k = 1, 2\right\},$$

where

$$\Phi_i\left(v_i, \lambda\right) = \min\left\{\phi_{si}\left(v_i, \lambda\right), s = 1, 2, \ldots, 6\right\},$$

is a phi-function for the component $\lambda A_i(v_i, \lambda)$ and the object Ω^*,

$$\phi_{1i}\left(v_i, \lambda\right) = x_i - \lambda r, \phi_{2i}\left(v_i, \lambda\right) = -x_i + \left(l - \lambda r\right),$$

$$\phi_{3i}\left(v_i, \lambda\right) = y_i - \lambda r, \phi_{4i}\left(v_i, \lambda\right) = -y_i + \left(w - \lambda r\right),$$

$$\phi_{5i}\left(v_i, \lambda\right) = z_i - \lambda r, \phi_{6i}\left(v_i, \lambda\right) = -z_i + \left(h - \lambda r\right);$$

$$\Phi_i^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right) = \min\left\{\phi_{si}^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right), s = 1, 2, \ldots, 6\right\},$$

is a phi-function for the component $\lambda B_{ik}(v_i, \theta_{ik}, \varphi_{ik}, \lambda)$ and the object Ω^*, $k = 1, 2$,

$$\phi_{1i}^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right) = x_{ik} - \lambda R = x_i + \rho \sin\theta_{ik} \cos\varphi_{ik} - \lambda R,$$

$$\phi_{2i}^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right) = -x_{ik} + \left(l - \lambda R\right) = -x_i - \rho \sin\theta_{ik} \cos\varphi_{ik} + \left(l - \lambda R\right),$$

$$\phi_{3i}^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right) = y_{ik} - \lambda R = y_i + \rho \sin\theta_{ik} \sin\varphi_{ik} - \lambda R,$$

$$\phi_{4i}^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right) = -y_{ik} + \left(w - \lambda R\right) = -y_i - \rho \sin\theta_{ik} \sin\varphi_{ik} + \left(w - \lambda R\right),$$

$$\phi_{5i}^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right) = z_{ik} - \lambda R = z_i + \rho \cos\theta_{ik} - \lambda R,$$

$$\phi_{6i}^k\left(v_i, \theta_{ik}, \varphi_{ik}, \lambda\right) = -z_{ik} + \left(h - \lambda R\right) = -z_i + \rho \cos\theta_{ik} + \left(h - \lambda R\right).$$

A mathematical model of the packing problem can be stated in the form of the following MIP (mixed integer problem):

$$\max f = \sum_{i=1}^{N} \tau_i \text{ s.t. } (t, \tau) \in W \subset \mathbb{R}^{7N} \times \Lambda^N \tag{7}$$

$$W = \Big\{ (t, \tau) \in R^{\varsigma} : \tau_i \cdot \tau_j \cdot \widehat{\Phi}_{ij}\left(v_i, v_j\right) \geq 0, i < j \in I_N, \tau_i \cdot \widehat{\Phi}^{i}_{12}(t_i) \geq 0, i \in I_N,$$
$$\tau_i \cdot \tau_j \cdot \widehat{\Phi}^{k}_{ij}\left(t_i, v_j\right) \geq 0, i, j \in I_N, k = 1, 2, \tau_i \cdot \tau_j \cdot \widehat{\Phi}^{kp}_{ij}\left(t_i, t_j\right) \geq 0, i < j \in I_N,$$
$$k = 1, 2, p = 1, 2 \Big\}, \tag{8}$$

$$\tau_i = \begin{cases} 1, \text{if } \ \Phi_i(t_i) \geq 0 \\ 0, \text{if } \ \Phi_i(t_i) < 0 \end{cases},$$

where $\varsigma = 8N$ is the dimension of the solution space, (t, τ) is the vector of all variables, $t = (t_1, \ldots, t_N) \in \mathbb{R}^{7N}$ is the vector of continuous variables, and $\tau = (\tau_1, \ldots, \tau_N) \in \Lambda^N$ is the vector of Boolean variables.

The inequalities $\widehat{\Phi}_{ij}\left(v_i, v_j,\right) \geq 0$, $\widehat{\Phi}^{i}_{12}(t_i) \geq 0$, $\widehat{\Phi}^{k}_{ij}\left(t_i, v_j\right) \geq 0$, $\widehat{\Phi}^{kp}_{ij}\left(t_i, t_j\right) \geq 0$, provide distance constraints (2), (3), (4), and (5), respectively, for $\lambda = 1$. The containing $T_i(t_i)$ into Ω defined in (6) is described by the inequality $\Phi_i(t_i) \geq 0$ for $\lambda = 1$.

2.5 Solution Strategy

The solution strategy for the packing problem (7) and (8) embraces the following main stages.

Stage 1. Define starting value of $N = n$, based on the fast heuristic algorithm introduced in [28].

Stage 2. Solve the following optimization problem:

$$\max_{t \in V \subset \mathbb{R}^{\sigma}} \lambda, \tag{9}$$

$$V = \Big\{ (t, \lambda) \in \mathbb{R}^{\sigma} : 0 \leq \lambda \leq 1, \widehat{\Phi}_{ij}\left(v_i, v_j, \lambda\right) \geq 0, i < j \in I_n, \tag{10}$$

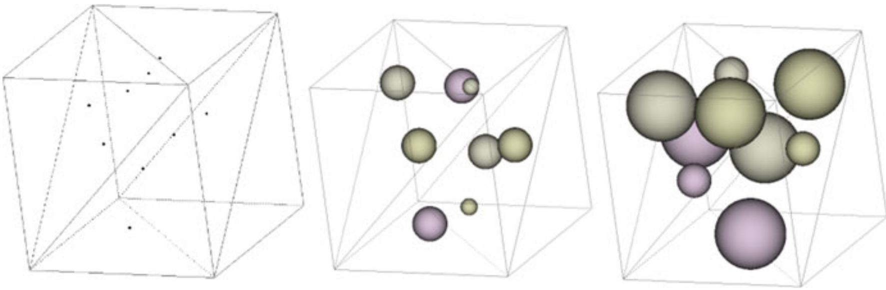


Fig. 2 Illustration to Step 3 – scaling three disconnected objects $T_i = A_i \cup B_{i1} \cup B_{i2}$, $i = 1, 2, 3$ to the full sizes

$$\overset{\frown i}{\Phi}_{12}(t_i, \lambda) \geq 0, i \in I_n, \overset{\frown k}{\Phi}_{ij}\left(t_i, v_j, \lambda\right) \geq 0, i = j \in I_n, k = 1, 2,$$

$$\overset{\frown kp}{\Phi}_{ij}\left(t_i, t_j, \lambda\right) \geq 0, i < j \in I_n, k = 1, 2, p = 1, 2, \Phi_i(t_i, \lambda) \geq 0, i \in I_n\Big\}.$$

In the problem (9) and (10), $\sigma = 1 + 7n$ is the solution space dimension, and $(t, \lambda) = (t_1, \ldots, t_n, \lambda)$ is the vector of variables.

If a solution of (9) and (10) is not found, then go to *Stage 4*; otherwise go to *Stage 3*.

Stage 3. Set $n = n + 1$ and go to *Stage 2*.

Stage 4. Take the obtained solution for $N = n - 1$ as a solution of the problem (7) and (8).

Let us consider an algorithm for searching for local maxima of problem (9) and (10).

This algorithm is based on the homothetic transformations of the objects $T_i(t_i)$, $i \in I_n$ and includes the following steps.

Step 1. Set the number of iterations K. Assume $s = 1$.

Step 2. Generate a feasible starting point $t^0 = \left(\lambda^0 = 0, t_1^0, \ldots, t_n^0\right)$ for the problem (9) and (10), where $t_i^0 = \left(v_i^0, \theta_{i1}^0, \varphi_{i1}^0, \theta_{i2}^0, \varphi_{i2}^0\right)$, $v_i^0 = \left(x_i^0, y_i^0, z_i^0\right) \in \Omega$, and $0 \leq \theta_{ik}^0 \leq \pi$, $0 \leq \varphi_{ik}^0 < 2\pi$, $i = 1, \ldots, n$, $k = 1, 2$. Here $x_i^0, y_i^0, z_i^0, \theta_{i1}^0, \varphi_{i1}^0, \theta_{i2}^0, \varphi_{i2}^0$, $i = 1, \ldots, n$, are randomly generated values in the given ranges.

Step 3. Scale the objects $\lambda T_i(t_i)$, $i \in I_n$, placed inside Ω, to the full size, searching for a point of the global maximum $t^* = \left(\lambda^*, t_1^*, \ldots, t_n^*\right)$ of the problem (9) and (10), starting from the feasible point t^0 (Fig. 2).

Step 4. If $\lambda^* = 1$ (solution is found), then our algorithm is terminated; otherwise go to *Step 5*.

Step 5. Set $s = s + 1$. If $s > K$, the algorithm is terminated (solution is not found for K iterations); otherwise go to *Step 2*.

An iterative decomposition procedure for searching for a local maximum of the problem (9) and (10) is used [39]. The optimization procedure reduces the problem (9) and (10) with $O(n^2)$ inequalities, to a sequence of nonlinear programming subproblems having a smaller number of nonlinear inequalities ($O(n)$).

The key idea of the algorithm is based on [32]. For each vector t_i, $i \in I_n$: (1) create fixed ε-cubes housing spherical objects that form disconnected objects, (2) allow moving each component within the corresponding ε-cube, (3) form a feasible subregion of the set (9) substituting $O(n^2)$ nonlinear inequalities for the pairs of components having non-overlapping ε-cubes by $O(n)$ inequalities, and (4) search for a local maximum on the subregion defined by $O(n)$ nonlinear inequalities. Use this local maximum as a starting point for the next iteration.

3 Results and Discussion

Experimental Results

In order to check the stability of Dox-C_{60} nanocomplex aqueous colloid solution, the size distribution was monitored with dynamic light scattering. The average of Dox-C_{60} nanocomplex was evaluated to be 135 nm (Fig. 3) that evidenced its monodisperse distribution and matched previous investigations [18] and evidenced storage stability during 6 months.

As an exemplar potent nanomedicine formulation, we have explored the packing of the molecules within C_{60} nanocomplex with an anthracycline antibiotic Dox in the molar ratio 1:2. C_{60} forms with Dox aqueous monodisperse colloid solution, with a size distribution of around 135 nm (Fig. 3). Previously the analysis of the energetics of Dox-C_{60} nanocomplexation pointed on its stabilization in aqueous

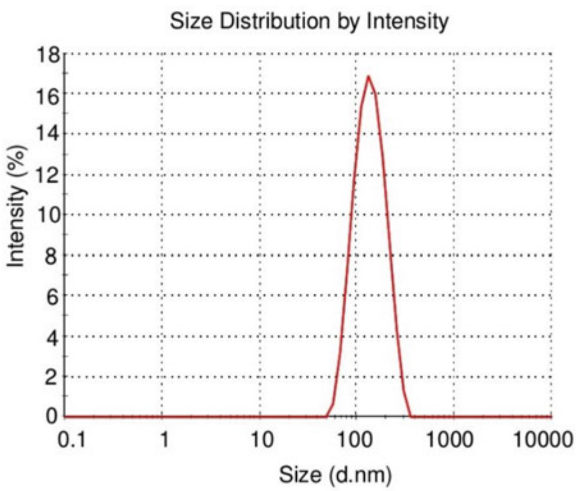


Fig. 3 Hydrodynamic size (diameter, nm) of Dox-C_{60} nanocomplex aqueous colloid solution. Intensity (%) – percentage of all scattered light intensity

Table 1 The geometric parameters of the Dox-C_{60} nanocomplexation

C_{60} size, diameter	0.72 nm
Dox size, diameter	1.5 nm
Dox-C_{60} distance	0.6 nm
C_{60}-C_{60} minimum distance	0.2 nm
Dox-Dox minimum distance	
3D structure of C_{60}	Sphere (simplified)
3D structure of Dox	

solution by intermolecular van der Waals interactions and hydrophobic interactions [14, 40]. The stacking-type binding of Dox was proposed as the main structural specificity of Dox-C_{60} nanocomplexation that occurred mainly within large C_{60} clusters with the distance of 0.6 nm between the aromatic planes of Dox and C_{60} molecules [14, 15, 40]. The central point in this model is the view of Dox-C_{60} nanocomplex as aggregates in which Dox acts as a layer between any two neighboring C_{60} molecules, pulling them together due to electrostatic forces [14].

4 Computational Results

We provide two instances to illustrate the work of our algorithm. In all the experiments, an AMD Athlon 64 X2 5200+ computer was used. For solving nonlinear programming problems, the open local optimization solver IPOPT was applied [41].

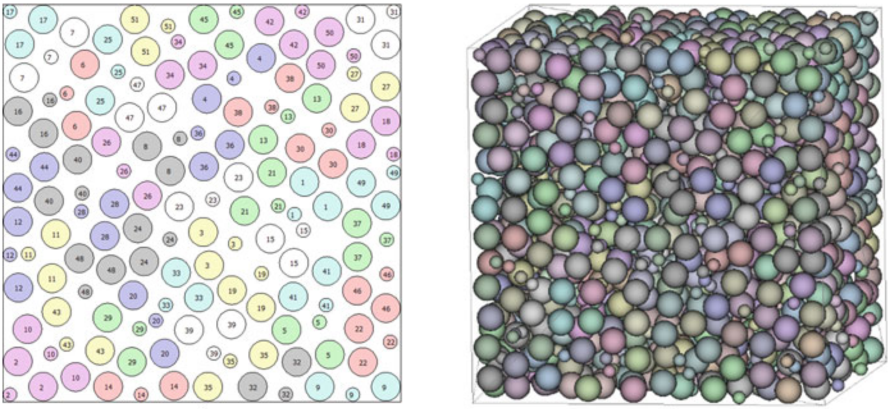
Required geometric parameters for the Dox-C_{60} nanocomplexation used in computational experiments are provided in Table 1.

Example 1 Test for 2D spheres: $r = 0.36$, $R = 0.75$, $l = w = 20$, $\Delta_{ab} = 0.6$, $\Delta_{aa} = \Delta_{bb} = 0.2$. The objective function value is $f^* = 68$, area $= 268.0181$ (the occupied area). The appropriate local optimal arrangement of $N = 68$ disconnected objects – Dox-C_{60} nanocomplexes, is shown in Fig. 4a.

Example 2 Test for 3D spheres: $r = 0.36$, $R = 0.75$, $l = w = h = 20\Delta_{ab} = 0.6$, $\Delta_{aa} = \Delta_{bb} = 0.2$. The objective function value is $f^* = 667$, volume $= 2487.7255$ (the occupied volume). The appropriate local optimal arrangement of $N = 667$ disconnected objects – C_{60}-Dox nanocomplexes, is shown in Fig. 4b.

5 Conclusions

As an exemplar potent nanomedicine formulation, we have explored the packing of the molecules within the C_{60} nanocomplex with the anthracycline antibiotic Dox that was previously proven to be efficient for the photodynamic and chemother-



a) b)

Fig. 4 Optimized packings: (**a**) 2D spheres and (**b**) 3D spheres

apeutic treatment of cancer cells. C_{60} forms with Dox aqueous monodisperse colloid solution (with a diameter of 135 nm). The non-standard packing problem of disconnected multi-spherical objects considering distance constraints was studied for a 2:1 Dox-C_{60} nanocomplex. For that, the objective was maximizing the number of objects that could be packed within the given cuboid. The mathematical model was constructed in the form of a mixed integer problem. A solution strategy was proposed and illustrated with examples (Fig. 4). This packing problem was motivated by applications in biomedicine, i.e., to develop an applicable nanomedicine formulation for a cancer therapy.

The proposed geometric design of the Dox-C_{60} formulation has a potential to provide a reliable tool for the smart systems and data-driven services in healthcare that aim to support the efficient and fast transition of the nanomedical research to the manufacturing and clinic stage.

6 Summary

The most attractive carbon nanomaterials include spherical fullerenes with C_{60} fullerene as the most prominent representative with unique physico-chemical properties, biological activity, and dual functionality as a nanocarrier for drug delivery and photosensitizer for photodynamic therapy. As currently more than 50 nanomedicines have been already approved for clinical use with more than 400 nanomedicine formulations being evaluated in clinical trials [21], the further support of nanomedicine development with smart system and data-driven services has a potential to answer unmet medical needs and help the field get faster to the clinic and manufacturing stage.

As an exemplar potent nanomedicine formulation, we have explored the packing of the molecules within C_{60} nanocomplex with an anthracycline antibiotic Dox in the molar ratio 1:2 that was previously proven to be the most efficient for the photodynamic and chemotherapeutic treatment of cancer cells [20]. C_{60} forms with Dox aqueous monodisperse colloid solution, which size distribution was evaluated to be around 135 nm (Fig. 3). Previously the analysis of the energetics of Dox-C_{60} nanocomplexation pointed on its stabilization in aqueous solution by intermolecular van der Waals interactions and hydrophobic interactions [14, 40]. The stacking-type binding of Dox was proposed as the main structural specificity of Dox-C_{60} nanocomplexation that occurred mainly within large C_{60} clusters with the distance of 0.6 nm between the aromatic planes of Dox and C_{60} molecules [14, 15, 40]. The central point in this model is the view of Dox-C_{60} nanocomplex as aggregates in which Dox acts as a layer between any two neighboring C_{60} molecules, pulling them together due to electrostatic forces [14].

A non-standard packing problem of disconnected multi-spherical objects considering distance constraints was studied for a Dox-C_{60} nanocomplex. For that, the objective was maximizing the number of objects that could be packed within a given cuboid. The mathematical model was constructed in the form of MIP and reduced to a sequence of NLP subproblems. A solution strategy was proposed and illustrated with examples (Fig. 4). This packing problem was motivated by applications in biomedicine (to develop an applicable nanomedicine formulation for a successful anticancer drug delivery). Developing mathematical tools for the packing problem with more complicated shaped components (polyhedra, ellipsoids) based on the results of [35, 42–44] and the experimental data of the nanocomplexes characterization is on the way. To cope with large-scale problems arising in different steps of the algorithmic approach Lagrangian heuristics [45], decomposition or aggregation techniques [46, 47] can also be applied.

Acknowledgments The work was supported by the Volkswagen Foundation under grant nos. 97 775 and 9C086 and the National Research Foundation of Ukraine under grant no. 02.2020/0128.

References

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. (2021). https://doi.org/10.3322/caac.21660
2. You, W., Henneberg, M.: Cancer incidence increasing globally: the role of relaxed natural selection. Evol. Appl. **11**, 140–152 (2017). https://doi.org/10.1111/eva.12523
3. Lin, L., Yan, L., Liu, Y., Yuan, F., Li, H., Ni, J.: Incidence and death in 29 cancer groups in 2017 and trend analysis from 1990 to 2017 from the Global Burden of Disease Study. J. Hematol. Oncol. **12**, 96 (2019). https://doi.org/10.1186/s13045-019-0783-9
4. Arruebo, M., Vilaboa, N., Sáez-Gutierrez, B., Lambea, J., Tres, A., Valladares, M., González-Fernández, Á.: Assessment of the evolution of cancer treatment therapies. Cancers (Basel). **3**, 3279–3330 (2011). https://doi.org/10.3390/cancers3033279

5. Shewach, D.S., Kuchta, R.D.: Introduction to cancer chemotherapeutics. Chem. Rev. **109**, 2859–2861 (2009). https://doi.org/10.1021/cr900208x
6. Chabner, B.A., Roberts, T.G.: Timeline: chemotherapy and the war on cancer. Nat. Rev. Cancer. **5**, 65–72 (2005). https://doi.org/10.1038/nrc1529
7. Cortés-Funes, H., Coronado, C.: Role of anthracyclines in the era of targeted therapy. Cardiovasc. Toxicol. **7**, 56–60 (2007). https://doi.org/10.1007/s12012-007-0015-3
8. Chen, G., Roy, I., Yang, C., Prasad, P.N.: Nanochemistry and nanomedicine for nanoparticle-based diagnostics and therapy. Chem. Rev. **116**, 2826–2885 (2016). https://doi.org/10.1021/acs.chemrev.5b00148
9. Shi, J., Kantoff, P.W., Wooster, R., Farokhzad, O.C.: Cancer nanomedicine: progress, challenges and opportunities. Nat. Rev. Cancer. **17**, 20–37 (2017). https://doi.org/10.1038/nrc.2016.108
10. Gharbi, N., Pressac, M., Hadchouel, M., Szwarc, H., Wilson, S.R., Moussa, F.: [60]fullerene is a powerful antioxidant in vivo with no acute or subacute toxicity. Nano Lett. **5**, 2578–2585 (2005). https://doi.org/10.1021/nl051866b
11. Delgado, J.L., Filippone, S., Giacalone, F., Herranz, M.A., Illescas, B., Pérez, E.M., Martín, N.: Buckyballs. Top. Curr. Chem. **350**, 1–64 (2014). https://doi.org/10.1007/128_2012_414
12. Castro, E., Hernandez Garcia, A., Zavala, G., Echegoyen, L.: Fullerenes in biology and medicine. J. Mater. Chem. B Mater. Biol. Med. **5**, 6523–6535 (2017). https://doi.org/10.1039/C7TB00855D
13. Gonchar, O.O., Maznychenko, A.V., Bulgakova, N.V., Vereshchaka, I.V., Tomiak, T., Ritter, U., Prylutskyy, Y.I., Mankovska, I.M., Kostyukov, A.I.: C60 fullerene prevents restraint stress-induced oxidative disorders in rat tissues: possible involvement of the Nrf2/ARE-antioxidant pathway. Oxidative Med. Cell. Longev. (2018). https://doi.org/10.1155/2018/2518676
14. Evstigneev, M.P., Buchelnikov, A.S., Voronin, D.P., Rubin, Y.V., Belous, L.F., Prylutskyy, Y.I., Ritter, U.: Complexation of C60 fullerene with aromatic drugs. ChemPhysChem. **14**, 568–578 (2013). https://doi.org/10.1002/cphc.201200938
15. Prylutskyy, Y.I., Evstigneev, M.P., Cherepanov, V.V., Kyzyma, O.A., Bulavin, L.A., Davidenko, N.A., Scharff, P.: Structural organization of C60 fullerene, doxorubicin, and their complex in physiological solution as promising antitumor agents. J. Nanopart. Res. **17**, 45 (2015). https://doi.org/10.1007/s11051-015-2867-y
16. Panchuk, R.R., Prylutska, S.V., Chumakl, V.V., Skorokhyd, N.R., Lehka, L.V., Evstigneev, M.P., Prylutskyy, Y.I., Berger, W., Heffeter, P., Scharff, P., Ritter, U., Stoika, R.S.: Application of C60 fullerene-doxorubicin complex for tumor cell treatment in vitro and in vivo. J. Biomed. Nanotechnol. **11**, 1139–1152 (2015)
17. Prylutska, S.V., Skivka, L.M., Didenko, G.V., Prylutskyy, Y.I., Evstigneev, M.P., Potebnya, G.P., Panchuk, R.R., Stoika, R.S., Ritter, U., Scharff, P.: Complex of C60 fullerene with doxorubicin as a promising agent in antitumor therapy. Nanoscale Res. Lett. **10** (2015). https://doi.org/10.1186/s11671-015-1206-7
18. Grebinyk, A., Prylutska, S., Grebinyk, S., Prylutskyy, Y., Ritter, U., Matyshevska, O., Dandekar, T., Frohme, M.: Complexation with C60 fullerene increases doxorubicin efficiency against leukemic cells in vitro. Nanoscale Res. Lett. **14**, 61 (2019). https://doi.org/10.1186/s11671-019-2894-1
19. Blazkova, I., Viet Nguyen, H., Kominkova, M., Konecna, R., Chudobova, D., Krejcova, L., Kopel, P., Hynek, D., Zitka, O., Beklova, M., Adam, V., Kizek, R.: Fullerene as a transporter for doxorubicin investigated by analytical methods and in vivo imaging. Electrophoresis. **35**, 1040–1049 (2014). https://doi.org/10.1002/elps.201300393
20. Grebinyk, A., Prylutska, S., Chepurna, O., Grebinyk, S., Prylutskyy, Y., Ritter, U., Ohulchanskyy, T.Y., Matyshevska, O., Dandekar, T., Frohme, M.: Synergy of chemo- and photodynamic therapies with C60 fullerene-doxorubicin nanocomplex. Nano. **9**, 1540 (2019). https://doi.org/10.3390/nano9111540
21. Germain, M., Caputo, F., Metcalfe, S., Tosi, G., Spring, K., Åslund, A.K.O., Pottier, A., Schiffelers, R., Ceccaldi, A., Schmid, R.: Delivering the power of nanomedicine to patients today. J. Control. Release. **326**, 164–171 (2020). https://doi.org/10.1016/j.jconrel.2020.07.007

22. Klein, S., Maggioni, S., Bucher, J., Mueller, D., Niklas, J., Shevchenko, V., Mauch, K., Heinzle, E., Noor, F.: In silico modeling for the prediction of dose and pathway-related adverse effects in humans from in vitro repeated-dose studies. Toxicol. Sci. **149**, 55–66 (2016). https://doi.org/10.1093/toxsci/kfv218
23. Esch, E.W., Bahinski, A., Huh, D.: Organs-on-chips at the frontiers of drug discovery. Nat. Rev. Drug Discov. **14**, 248–260 (2015). https://doi.org/10.1038/nrd4539
24. Ungson, Y., Burtseva, L., Garcia-Curiel, E.R., Valdez Salas, B., Flores-Rios, B.L., Werner, F., Petranovskii, V.: Filling of irregular channels with round cross-section: modeling aspects to study the properties of porous materials. Materials. **11**, 1901 (2018). https://doi.org/10.3390/ma11101901
25. Otaru, A.J., Kennedy, A.R.: The permeability of virtual macroporous structures generated by sphere packing models: comparison with analytical models. Scr. Mater. **124**, 30–33 (2016). https://doi.org/10.1016/j.scriptamat.2016.06.037
26. Hifi, M., Yousef, L.: A local search-based method for sphere packing problems. Eur. J. Oper. Res. **274**, 482–500 (2019). https://doi.org/10.1016/j.ejor.2018.10.016
27. Halkarni, S.S., Sridharan, A., Prabhu, S.V.: Experimental investigation on effect of random packing with uniform sized spheres inside concentric tube heat exchangers on heat transfer coefficient and using water as working medium. Int. J. Therm. Sci. **133**, 341–356 (2018). https://doi.org/10.1016/j.ijthermalsci.2018.05.023
28. Duriagina, Z., Lemishka, I., Litvinchev, I., Marmolejo, J.A., Pankratov, A., Romanova, T., Yaskov, G.: Optimized filling of a given cuboid with spherical powders for additive manufacturing. J. Oper. Res. Soc. China. **9**, 853–868 (2021). https://doi.org/10.1007/s40305-020-00314-9
29. Burtseva, L., Salas, B.V., Romero, R., Werner, F.: Recent advances on modelling of structures of multi-component mixtures using a sphere packing approach. Int. J. Nanotechnol. **13**, 44–59 (2016). https://doi.org/10.1504/IJNT.2016.074522
30. Blyuss, O., Koriashkina, L., Kiseleva, E., Molchanov, R.: Optimal placement of irradiation sources in the planning of radiotherapy: mathematical models and methods of solving. Comput. Math. Methods Med. **2015**, e142987 (2015). https://doi.org/10.1155/2015/142987
31. Chazelle, B., Edelsbrunner, H., Guibas, L.J.: The complexity of cutting complexes. Discrete Comput. Geom. **4**, 139–181 (1989). https://doi.org/10.1007/BF02187720
32. Romanova, T., Litvinchev, I., Pankratov, A.: Packing ellipsoids in an optimized cylinder. Eur. J. Oper. Res. **285**, 429–443 (2020). https://doi.org/10.1016/j.ejor.2020.01.051
33. Romanova, T., Stoyan, Y., Pankratov, A., Litvinchev, I., Avramov, K., Chernobryvko, M., Yanchevskyi, I., Mozgova, I., Bennell, J.: Optimal layout of ellipses and its application for additive manufacturing. Int. J. Prod. Res. **59**, 560–575 (2021). https://doi.org/10.1080/00207543.2019.1697836
34. Stoyan, Y.G., Romanova, T.E., Pankratov, O.V., Stetsyuk, P.I., Maximov, S.V.: Sparse balanced layout of ellipsoids. Cybern. Syst. Anal. **57**, 864–873 (2021). https://doi.org/10.1007/s10559-021-00412-3
35. Pankratov, A., Romanova, T., Litvinchev, I.: Packing ellipses in an optimized rectangular container. Wireless Netw. **26**, 4869–4879 (2020). https://doi.org/10.1007/s11276-018-1890-1
36. Stoyan, Y., Yaskov, G., Romanova, T., Litvinchev, I., Yakovlev, S., Cantú, J.M.V., Stoyan, Y., Yaskov, G., Romanova, T., Litvinchev, I., Yakovlev, S., Cantú, J.M.V.: Optimized packing multidimensional hyperspheres: a unified approach. MBE. **17**, 6601–6630 (2020). https://doi.org/10.3934/mbe.2020344
37. Ritter, U., Prylutskyy, Y.I., Evstigneev, M.P., Davidenko, N.A., Cherepanov, V.V., Senenko, A.I., Marchenko, O.A., Naumovets, A.G.: Structural features of highly stable reproducible C60 fullerene aqueous colloid solution probed by various techniques. Fullerenes, Nanotubes, Carbon Nanostruct. **23**, 530–534 (2015). https://doi.org/10.1080/1536383X.2013.870900
38. Turov, V.V., Chehun, V.F., Krupskaya, T.V., Barvinchenko, V.N., Chehun, S.V., Ugnivenko, A.P., Prylutskyy, Y.I., Scharff, P., Ritter, U.: Effect of small addition of C60 fullerenes on the hydrated properties of nanocomposites based on highly dispersed silica and DNA. Chem. Phys. Lett. **496**, 152–156 (2010). https://doi.org/10.1016/j.cplett.2010.07.001

39. Romanova, T., Stoyan, Y., Pankratov, A., Litvinchev, I., Marmolejo, J.A.: Decomposition algorithm for irregular placement problems. In: Vasant, P., Zelinka, I., Weber, G.-W. (eds.) Intelligent Computing and Optimization, pp. 214–221. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-33585-4_21
40. Prylutskyy, Y.I., Evstigneev, M.P., Pashkova, I.S., Wyrzykowski, D., Woziwodzka, A., Gołuński, G., Piosik, J., Cherepanov, V.V., Ritter, U.: Characterization of C60 fullerene complexation with antibiotic doxorubicin. Phys. Chem. Chem. Phys. **16**, 23164–23172 (2014). https://doi.org/10.1039/c4cp03367a
41. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. **106**, 25–57 (2006). https://doi.org/10.1007/s10107-004-0559-y
42. Stoyan, Y., Pankratov, A., Romanova, T., Fasano, G., Pintér, J.D., Stoian, Y.E., Chugay, A.: Optimized packings in space engineering applications: part I. In: Fasano, G., Pintér, J.D. (eds.) Modeling and Optimization in Space Engineering : State of the Art and New Challenges, pp. 395–437. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-10501-3_15
43. Romanova, T., Bennell, J., Stoyan, Y., Pankratov, A.: Packing of concave polyhedra with continuous rotations using nonlinear optimisation. Eur. J. Oper. Res. **268**, 37–53 (2018)
44. Romanova, T., Pankratov, A., Litvinchev, I., Strelnikova, E.: Modeling nanocomposites with ellipsoidal and conical inclusions by optimized packing. In: Marmolejo-Saucedo, J.A., Vasant, P., Litvinchev, I., Rodriguez-Aguilar, R., Martinez-Rios, F. (eds.) Computer Science and Health Engineering in Health Services, pp. 201–210. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-69839-3_14
45. Litvinchev, I.S.: Decomposition-aggregation method for convex programming problems. Optimization. **22**, 47–56 (1991). https://doi.org/10.1080/02331939108843642
46. Litvinchev, I.S., Rangel, S.: Localization of the optimal solution and a posteriori bounds for aggregation. Comput. Oper. Res. **26**, 967–988 (1999). https://doi.org/10.1016/S0305-0548(99)00027-1
47. Litvinchev, I., Mata, M., Rangel, S., Saucedo, J.: Lagrangian heuristic for a class of the generalized assignment problems. Comput. Math. Appl. **60**, 1115–1123 (2010). https://doi.org/10.1016/j.camwa.2010.03.070

Multiple Sensor Fusion for Stress Detection in the Hospital Environment

Muhammad Ali Fauzi and Bian Yang

Abstract Researchers have found that workers of all types have been affected by stress in recent years, with hospital staff being more likely to suffer from depression than others. Thus, stress detection for hospital staff is crucial as part of stress management. Smartwatch is one of the devices that is suitable for use in the hospital setting for stress detection tasks because of its usability for the workplace. Smartwatch also provides several built-in sensors that can be utilized to create a robust stress detection system. In this chapter, we implement and compare some multiple sensor fusion strategies for stress detection using Logistic Regression and four sensors including skin temperature, accelerometer, electrodermal activity, and blood volume pulse sensors. The multiple sensor fusion strategies used in this study are feature-level, decision-level, and score-level strategies. According to the experiment's findings, the accelerometer sensor model performed the best when compared to other sensor models, scoring 0.866 accuracy and 0.758 F_1-measure. The results also show that all models of the multiple sensor fusion approach outperformed all individual sensor models in terms of performance. The score-level fusion method produced the best results at a rate of 0.921 accuracy and 0.844 F_1-measure. Our code is available at https://github.com/cahkanor/WESAD-Multiple-Sensor-Fusion-Stress-Detection.

Keywords Stress detection · Sensor · Fusion · Smartwatch · Hospital · Logistic regression

1 Introduction

Stress has become a trending issue in our modern society recently. Many reports suggest that in recent years all workers are susceptible to stress with hospital

M. A. Fauzi (✉) · B. Yang
Norwegian University of Science and Technology (NTNU), Gjøvik, Norway
e-mail: muhammad.a.fauzi@ntnu.no; bian.yang@ntnu.no

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4_16

staff members being more likely to suffer from depression than others [12]. The high-stress level is attributed mainly to the heavy workloads, long working hours, intensive time constraints, lack of support, and intense interaction with patients and family [1, 20, 29]. Moreover, the introduction of COVID-19 in 2019 made this condition worse. Healthcare systems have seen a major resource overflow due to the fast spread of COVID-19 [7, 21]. As the consequence, the stress level of hospital staff members is also elevated [3, 26].

The negative impacts of work-related stress on employees' productivity and happiness, as well as their physical and mental health, absenteeism and its financial cost, the broader repercussions on family function, and lastly, the possibility of employer responsibility are just a few of the consequences [23, 28, 31]. For instance, studies show a strong correlation between a high level of stress and poor patient safety in the hospital environment [30, 32]. Another study found that hospital staff members' cybersecurity activities were riskier the more stressed they were[10]. These investigations support a previous study that found stressed-out people learn things slowly and may make less lucrative judgments[33]. Therefore, stress management is important.

The first step of stress management is to monitor an individual's stress level. Knowing their personal level of stress may help people stay alert, feel more in control of how they react to circumstances, and know when to unwind or take action to handle it effectively [19]. Additionally, this monitoring can aid in the early detection of mental diseases and illnesses. The measurement of physiological responses to stress using various sensors has become one of the most prominent methods for assessing stress levels in recent years. One of the best gadgets for stress monitoring, particularly in the workplace, is smartwatches. Apart from its high level of social acceptance, this type of gadget has several integrated sensors, such as a skin temperature sensor (ST), an accelerometer (ACC), an electrodermal activity sensor (EDA), and a blood volume pulse sensor (BVP), that may be used to detect stress using multiple sensor fusion [18, 27].

Multiple sensor fusion can provide richer information by synthesizing data from many perspectives on the environment in order to improve the stress detection system [5]. This is accomplished by integrating redundant and complementary measurements of the environment by using several sensors. There are many strategies in terms of multiple sensor fusions. The most popular strategy is to combine them at the feature level (e.g., [11, 14, 24, 25]). In this chapter, we propose some multiple sensor fusion strategies for stress detection. The sensors employed include skin temperature, accelerometer, electrodermal activity, and blood volume pulse sensors. The classifier used is Logistic Regression because of its ability to provide good performance for stress classification tasks [4, 9, 17]. The contribution of this study is to provide a comparative analysis between single and multiple sensors. In addition, we also propose several fusion methods and analyze the comparison between them.

2 Materials and Methods

2.1 Dataset

This study made use of a publicly accessible dataset called WESAD (Wearable Stress and Affect Detection) [24]. Data were collected from 15 individuals; 12 of whom were male and 3 of whom were female. The demographic data of the dataset's participants are shown in Table 1.

During the data collection for the WESAD dataset, three different states of mind are recorded: neutral, stress, and amusement. The participants were tasked to read given magazines while seated or standing during the neutral situation. Meanwhile, in the stressful condition, the participants underwent a Trier Social Tension Test (TSST) [16], which was designed to elicit stress. In the amusing scenario, the participants viewed a series of humorous video clips. For this study, we merged the neutral and amusement sessions into one class called non-stress. Therefore, the problem in this study was binary classification (stress and non-stress).

The data were actually collected simultaneously utilizing an Empatica E4 wristwatch and RespiBAN chest-worn sensors. This study only uses Empatica E4 data because smartwatch sensors are the primary subject of this investigation. Chest-worn sensors are not convenient for working environments like a hospital setting. The smartwatch comes equipped with an accelerometer (ACC), an electrodermal activity (EDA) sensor, a skin temperature (ST) sensor, and a blood volume pulse (BVP) sensor.

2.2 Features

In order to extract features from the raw data, the signal data from each sensor were segmented using a sliding window with a sliding step of 0.25 seconds and a window length of 60 seconds. In addition to the 3-dimensional accelerometer signal data (x-, y-, and z-axis), we also extracted their magnitude using Eq. 1. In the next step, we created five additional signals for every original sensor's data. The first two signals are its first and second derivatives. The other three signals are transformed form of the signal using a Discrete Wavelet Transform (DWT) with the Haar wavelet at three different frequencies (1 Hz, 2 Hz, and 4 Hz). In total, 42 different signals were constructed including 6 ST signals, 24 ACC signals, 6 EDA signals, and 6 BVP signals. In the final stage, 10 statistical features were extracted

Table 1 Demographic information of the WESAD dataset participants ($N = 15$)

Characteristic	Value, mean (SD)
Age (years)	27.5 (2.4)
Height (cm)	177.6 (6.7)
Weight (kg)	73.1 (10.3)

Table 2 Extracted statistical features from each signal

No.	Statistical features
1	Mean value of the signal
2	Minimum value of the signal
4	Maximum value of the signal
4	Median value of the signal
5	Maximum signal amplitude
6	Signal variance
7	Standard signal deviation
8	Absolute signal deviation
9	Signal kurtosis
10	Signal skewness

from each signal using the Python BioSPPy and Numpy packages [13], as shown in Table 2, to produce 420 features to be used for the stress detection task. In addition, we conducted a feature normalization using a Min–Max normalization technique due to its ability to improve classification performance.

$$ACC_{norm} = \sqrt{ACC_x^2 + ACC_y^2 + ACC_z^2} \quad (1)$$

2.3 Classification Strategies

In terms of the number of sensors employed, generally, we can divide classification strategies into two types: individual sensor and multiple sensor fusion. The individual sensor strategy only utilizes data from one single sensor to feed the machine learning (ML) algorithm. In contrast, the multiple sensor fusion strategy combines several sensor data in order to capture richer and more diverse information to feed the ML method. This combination can be conducted at the feature level, decision level, or score level. The ML algorithm used in this work is Logistic Regression and is implemented using the Scikit-learn library [22].

2.4 Individual Sensor Strategy

Generally, classification tasks can be divided into two main stages: training and testing. Using the individual sensor strategy, as pictured in Fig. 1, in the training stage, the raw data from the sensor are transformed into features. These features are then fed into the ML algorithm to train the model. The final result is an ML model. Since we have four sensors, we produce four different models using the individual

Fig. 1 Training process using individual sensor

sensor strategy. Furthermore, this model will be used to detect stress levels from testing data as depicted in Fig. 2. The decision from each model can be different.

2.5 Feature-Level Multiple Sensor Fusion Strategy

This method combines all of the features from each sensor into a single vector once they have been extracted from each sensor. In the training process, these combined features are then used to train the ML algorithm to generate an ML model as shown in Fig. 3. In the testing phase, this model is used to detect stress levels as displayed in Fig. 4. The features used in the testing stage are also the combination of all the extracted features from each sensor.

2.6 Decision-Level Multiple Sensor Fusion Strategy

Decision-level multiple sensor fusion strategy combines the decisions from each sensor's ML model. This method's primary objective is to use a group of separate classifiers' redundancy to combine their outputs in order to increase robustness.

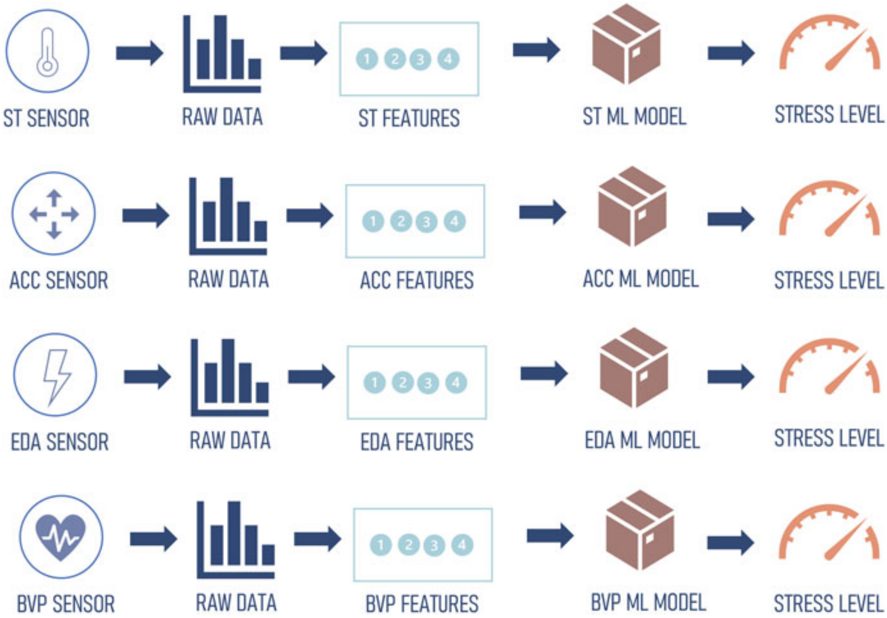


Fig. 2 Individual sensor ML model testing

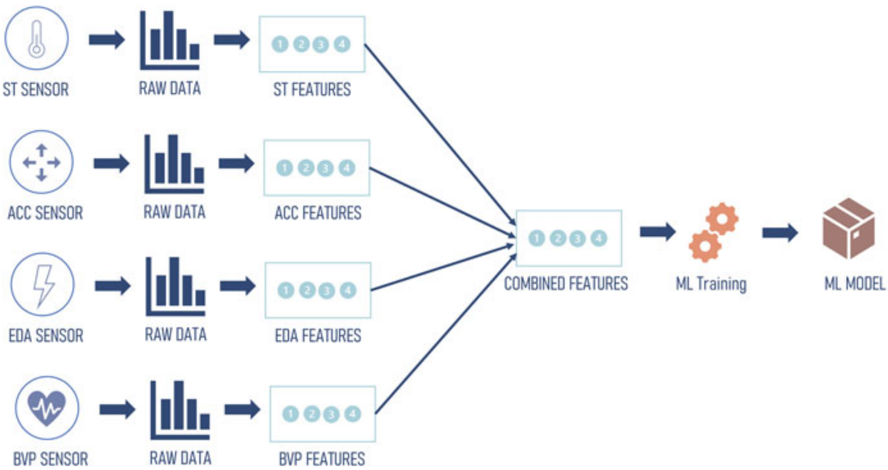


Fig. 3 Training process using feature-level multiple sensor fusion

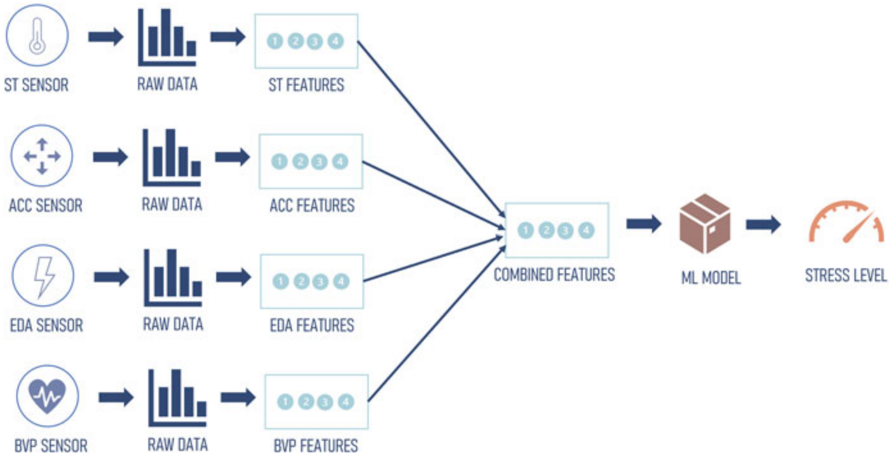


Fig. 4 Testing process of feature-level multiple sensor fusion ML model

In this study, we use hard voting to determine the final decision. Therefore, the number of models used must be odd. Since the number of sensors in this study is four, only three of them will be used for this strategy. The training process for this strategy is almost similar to the training process using the individual sensor strategy as displayed in Fig. 1. The difference is that we only use three sensors instead of four and finally produce three individual ML models. These three models are then used in the testing phase to detect stress levels as depicted in Fig. 5. The features from each sensor are extracted before being fed into each sensor's individual model. Then, each model will decide whether the testing data point belongs to the stress or non-stress class. The final decision will be determined using hard voting. In other words, each sensor model has one vote, and the class of the data is decided by the majority vote [8].

Assume that M_t is used to represent a sensor model in the ensemble E such that $t = 1, 2, ..., n$ and ensemble $E = M_1, M_2, ..., M_n$. The decision of the t_{th} model (M_t) on a testing data x is represented by $d_t(c, x) \in 0, 1$, where $c = 1, 2, ..., k$ and k is the number of classes. The result will be $d_t(c, x) = 1$, if t_{th} model determines that x belong to the class c, and $d_t(c, x) = 0$ otherwise. The final class of data x in majority voting is determined using Eq. 2 [6].

$$class(x) = \underset{c \in \mathcal{C}}{\text{argmax}} \sum_{t=1}^{n} d_t(c, x) \tag{2}$$

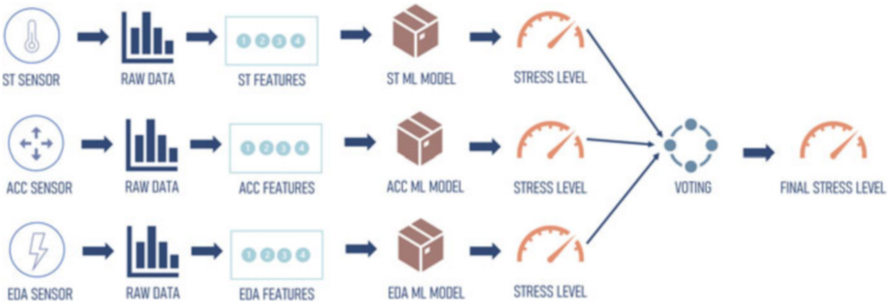


Fig. 5 Testing process of decision-level multiple sensor fusion ML model

2.7 Score-Level Multiple Sensor Fusion Strategy

Score-level multiple sensor fusion strategy computes the probability of each class using the ML method of each sensor and then combines them. This method provides a softer combination than decision-level by utilizing probability information. Using this strategy, the number of models used does not have to be odd. Therefore, all of the models from all sensors are used for this strategy in this study. The training process for this strategy is completely similar to the training process using the individual sensor strategy as displayed in Fig. 1. The end product of the training phase is four individual ML models. These models are then used in the testing phase to detect stress levels as depicted in Fig. 6. The features from each sensor are extracted before being fed into each sensor's individual model. Then, each model will compute the probability of each stress and non-stress class. These probability values of each class from the four models are then averaged. Finally, the final class of the testing data is the one with the highest average probability value.

Assume that M_t is used to represent a sensor model in the ensemble E such that $t = 1, 2, ..., n$ and ensemble $E = M_1, M_2, ..., M_n$. The probability of testing data belonging to the class c computed using the t_{th} model (M_t) is denoted by $p_t(c, x) \in [0, 1]$, where $c = 1, 2, ..., k$ and k is the number of classes. The final class of data x in soft voting is determined using Eq. 3 [15].

$$class(x) = \underset{c \in \mathcal{C}}{\text{argmax}} \sum_{t=1}^{n} \frac{p_t(c, x)}{n} \tag{3}$$

2.8 Evaluation

The leave-one-subject-out (LOSO) cross-validation (CV) method was used to evaluate all of the strategies. This evaluation method uses each individual participant as a "test" set. It is a particular form of k-fold cross-validation where the number of k

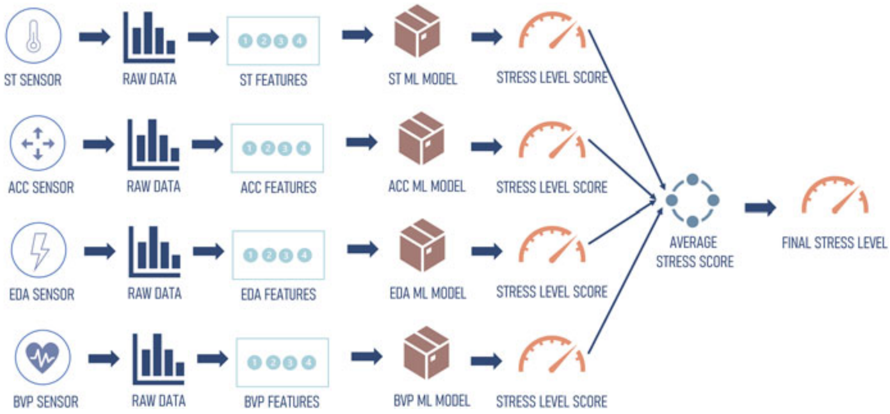


Fig. 6 Testing process of score-level multiple sensor fusion ML model

Fig. 7 Confusion matrix

		Predicted	
		Stress	Non-stress
Actual	Stress	TP	FN
	Non-Stress	FP	TN

is the same as the number of participants in the dataset. In this study using WESAD data, the value of k is 15. Therefore, we have 15 iterations to evaluate the machine learning model. On each iteration, the model is trained on 14 participants and tested on the “left out” participant. Finally, the evaluation metric for each iteration will be averaged. Several evaluation metrics are computed including Accuracy (Acc), Precision (P), Recall (R), and F_1-measure (F_1) on the basis of the confusion matrix shown in Fig. 7. The following equations show the formula for each measurements:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{4}$$

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$F_1 = 2\frac{P \cdot R}{P + R} \tag{7}$$

3 Result

The stress classification results using individual sensor strategy are displayed in Table 3, while the results using multiple sensor fusion are depicted in Table 4. In terms of accuracy, all of the individual sensor models work quite well to conduct a stress detection task with the ACC sensor model achieving the best accuracy with 0.866, while the ST sensor model obtained the lowest value with 0.823. However, since this task is conducted on an imbalanced dataset, we have to focus more on precision, recall, and F_1-measure. In terms of F_1-measure, the ACC sensor model also had the best performance with 0.758. The BVP sensor model sat in the second place with 0.726 accuracy, while the ST sensor model came behind it with a value of 0.684. In the last place, the EDA sensor model obtained only 0.611 accuracy. Furthermore, in terms of accuracy and precision, the ACC sensor model also got the best result compared to the other sensor models.

Table 4 shows that all models of the multiple sensor fusion methods achieved better performance compared to all of the individual sensor models. By combining some sensor data, richer information can be used to decide the stress levels. Generally, decision-level fusion strategy got the worst performance compared to other fusion strategies. This result is reasonable because the decision-level fusion strategy only uses three sensors compared to others that use four of them. The best combination for decision-level fusion strategy is ACC+ST+BVP with 0.802 of F_1-measure. It is reasonable because this combination did not employ the EDA sensor model that had the worst performance according to Table 3. Since the ACC sensor model is the best among other individual sensor models, the combination that did not include it got the worst result with 0.762 of F_1-measure. The best result

Table 3 Stress detection result using individual sensors

Sensor	*Acc*	*P*	*R*	F_1
EDA	0.828	0.556	**0.813**	0.611
ACC	**0.866**	**0.753**	**0.813**	**0.758**
ST	0.823	0.661	0.743	0.684
BVP	0.853	0.702	0.792	0.726

Bold means the highest score compare to the other method/strategy

Table 4 Stress detection result using multiple sensor fusion

Multiple sensor fusion strategy	*Acc*	*P*	*R*	F_1
Feature level	0.891	**0.814**	0.855	0.812
Decision level EDA+ACC+ST	0.897	0.731	0.920	0.793
Decision level EDA+ACC+BVP	0.901	0.721	0.937	0.794
Decision level EDA+ST+BVP	0.880	0.698	0.885	0.762
Decision level ACC+ST+BVP	0.893	0.762	0.874	0.802
Score level	**0.921**	0.781	**0.951**	**0.844**

Bold means the highest score compare to the other method/strategy

Table 5 Comparison with other studies

Method	Accuracy
Schmidt et al. [24]	0.883
Siirtola [25]	0.874
Alshamrani [2]	0.850
Fauzi and Yang [9]	0.871
Zhu et al. [34]	0.864
Our best method	**0.921**

Bold means the highest score compare to the other method/strategy

was obtained by score-level fusion strategy with 0.844 of F_1-measure and 0.921 of accuracy. This fusion strategy also got the best recall score with 0.951. Meanwhile, feature-level fusion strategy came quite close behind score-level fusion with 0.812 of F_1-measure. This strategy even obtained the best precision value of 0.814.

The comparison between our research and earlier studies that employed smartwatch data from the WESAD dataset for stress detection tasks with two classes (stress and non-stress) is shown in Table 5. Since almost all the previous works only use accuracy as the evaluation metric, we only compare the accuracy. Table 5 shows that our best method got the best accuracy compared to the other methods from prior work. However, this comparison may not really be fair because the features and machine learning methods used by each study are different.

4 Conclusion

Many hospital employees, according to several studies, experience acute stress. As a result, stress detection is crucial for stress management. Knowing their personal stress level may help people stay alert, feel more in control of how they react to circumstances, and know when to unwind or take action to handle it appropriately. Smartwatch is one of the devices that is suitable for use in the hospital setting because of its usability for the workplace. Smartwatch also provides several built-in sensors that can be utilized to create a robust stress detection system.

In this chapter, we propose some multiple sensor fusion strategies to classify stress employing Logistic Regression and four sensors including skin temperature, accelerometer, electrodermal activity, and blood volume pulse sensors. In addition, we also analyze the performance of each individual sensor model as well as each multiple sensor fusion strategy. The experiment results show that accelerometer sensor models had the best performance compared to other sensor models with 0.758 of F_1-measure and 0.866 of accuracy. The results also report that all of the models of the multiple sensor fusion methods achieved better performance compared to all of the individual sensor models. The best result was achieved by score-level fusion with 0.844 of F_1-measure and 0.921 of accuracy.

WESAD dataset only provides two classes: stress and non-stress. In the future work, this fusion method can be tested on a stress detection dataset that provides more stress levels (e.g., low stress, moderate stress, and high stress). In addition, it will also be interesting to propose different ways to do multiple sensor fusion (e.g., a weighted combination method, etc.).

References

1. Acosta-Ramos, S., Ramirez-Martinez, F.R., Manriquez, I.J.R., Galindo-Odilon, M., Estrada-Esparza, S.Y., Trejo-Franco, J., Flores-Padilla, L.: Burnout syndrome and association with work stress in nursing staff in public hospital of the northern border of Mexico. Archiv. Psychiatric Nurs. **35**(6), 571–576 (2021)
2. Alshamrani, M.: An advanced stress detection approach based on processing data from wearable wrist devices. Int. J. Adv. Comput. Sci. Appl **12**, 399–405 (2021)
3. Amanullah, S., Ramesh Shankar, R.: The impact of COVID-19 on physician burnout globally: a review. In: Healthcare, vol. 8, p. 421. MDPI, Basel (2020)
4. Can, Y.S., Chalabianloo, N., Ekiz, D., Ersoy, C.: Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. Sensors **19**(8), 1849 (2019)
5. Chavez-Garcia, R.O., Aycard, O.: Multiple sensor fusion and classification for moving object detection and tracking. IEEE Trans. Intell. Transport. Syst. **17**(2), 525–534 (2015)
6. Dogan, A., Birant, D.: A weighted majority voting ensemble approach for classification. In: 2019 4th International Conference on Computer Science and Engineering (UBMK), pp. 1–6. IEEE, Piscataway (2019)
7. Ebrahimi, H., Jafarjalal, E., Lotfolahzadeh, A., Kharghani Moghadam, S.M.: The effect of workload on nurses' quality of life with moderating perceived social support during the COVID-19 pandemic. Work **70**, 1–9 (2021)
8. Fauzi, M.A., Bours, P.: Ensemble method for sexual predators identification in online chats. In: 2020 8th International Workshop on Biometrics and Forensics (IWBF). pp. 1–6. IEEE, Piscataway (2020)
9. Fauzi, M.A., Yang, B.: Continuous stress detection of hospital staff using smartwatch sensors and classifier ensemble. In: pHealth 2021, pp. 245–250. IOS Press, Amsterdam (2021)
10. Fauzi, M.A., Yeng, P., Yang, B., Rachmayani, D.: Examining the link between stress level and cybersecurity practices of hospital staff in Indonesia. In: The 16th International Conference on Availability, Reliability and Security, pp. 1–8 (2021)
11. Garg, P., Santhosh, J., Dengel, A., Ishimaru, S.: Stress detection by machine learning and wearable sensors. In: 26th International Conference on Intelligent User Interfaces, pp. 43–45 (2021)
12. Gerada, C.: Doctors, suicide and mental illness. BJPsych Bull. **42**(4), 165–168 (2018)
13. Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al.: Array programming with numpy. Nature **585**(7825), 357–362 (2020)
14. Indikawati, F.I., Winiarti, S.: Stress detection from multimodal wearable sensor data. In: IOP Conference Series: Materials Science and Engineering, vol. 771, p. 012028. IOP Publishing, Bristol (2020)
15. Karlos, S., Kostopoulos, G., Kotsiantis, S.: A soft-voting ensemble based co-training scheme using static selection for binary classification problems. Algorithms **13**(1), 26 (2020)
16. Kirschbaum, C., Pirke, K.M., Hellhammer, D.H.: The 'trier social stress test'-a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology **28**(1–2), 76–81 (1993)

17. Kurniawan, H., Maslov, A.V., Pechenizkiy, M.: Stress detection from speech and galvanic skin response signals. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. pp. 209–214. IEEE, Picataway (2013)
18. Lazaro, M.J.S., Lim, J., Kim, S.H., Yun, M.H.: Wearable technologies: Acceptance model for smartwatch adoption among older adults. In: International Conference on Human-Computer Interaction, pp. 303–315. Springer, Berlin (2020)
19. Liao, W., Zhang, W., Zhu, Z., Ji, Q.: A real-time human stress monitoring system using dynamic Bayesian network. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-workshops, pp. 70–70. IEEE, Picataway (2005)
20. Luan, X., Wang, P., Hou, W., Chen, L., Lou, F.: Job stress and burnout: a comparative study of senior and head nurses in China. Nursing Health Sci, **19**(2), 163–169 (2017)
21. Lucchini, A., Giani, M., Elli, S., Villa, S., Rona, R., Foti, G.: Nursing activities score is increased in COVID-19 patients. Intensive Crit. Care Nurs. **59**, 102876 (2020)
22. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
23. Pickering, T.G.: Mental stress as a causal factor in the development of hypertension and cardiovascular disease. Curr. Hypertens. Rep. **3**(3), 249–254 (2001)
24. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., Van Laerhoven, K.: Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction, pp. 400–408 (2018)
25. Siirtola, P.: Continuous stress detection using the sensors of commercial smartwatch. In: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, pp. 1198–1201 (2019)
26. Sorokin, M.Y., Kasyanov, E.D., Rukavishnikov, G.V., Makarevich, O.V., Neznanov, N.G., Morozov, P.V., Lutova, N.B., Mazo, G.E.: Stress and stigmatization in health-care workers during the COVID-19 pandemic. Indian J. Psychiatry **62**(Suppl 3), S445 (2020)
27. Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., Gamberini, L.: Measuring user acceptance of wearable symbiotic devices: Validation study across application scenarios. In: International Workshop on Symbiotic Interaction, pp. 87–98. Springer, Berlin (2015)
28. Tennant, C.: Work-related stress and depressive disorders. J. Psychosom. Res. **51**(5), 697–704 (2001)
29. Tsai, Y.C., Liu, C.H.: Factors and symptoms associated with work stress and health-promoting lifestyles among hospital staff: a pilot study in Taiwan. BMC Health Serv. Res. **12**(1), 1–8 (2012)
30. Tsiga, E., Panagopoulou, E., Montgomery, A.: Examining the link between burnout and medical error: a checklist approach. Burnout Res. **6**, 1–8 (2017)
31. Wang, Y., Chen, R., Zhang, L.: Reliability and validity of generalized anxiety scale-7 in inpatients in Chinese general hospital. J. Clin. Psychiatr. **28**, 168–71 (2018)
32. Welp, A., Meier, L.L., Manser, T.: Emotional exhaustion and workload predict clinician-rated and objective patient safety. Front. Psychol. **5**, 1573 (2015)
33. Wemm, S.E., Wulfert, E.: Effects of acute stress on decision making. Appl. Psychophysiol. Biofeedback **42**(1), 1–12 (2017)
34. Zhu, L., Spachos, P., Gregori, S.: Multimodal physiological signals and machine learning for stress detection by wearable devices. In: 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–6. IEEE, Picataway (2022)

Index

A
Agent-based simulation (ABS), 39–45, 52, 109–163
Anylogistix, 41, 44, 49, 53, 54, 59

B
Blockchain, 182, 184, 186, 188, 189, 192, 195–197, 218

C
Catastrophic health spending, 247–255
Corporate culture, 239
COVID-19, v, 52, 103, 247–255, 274
CVRP, 3–13

D
Data modelling, 80, 85
Doxorubicin (Dox), 258–261, 266–269

E
E-commerce, 101–106
Electrical design, 79–86
Emotion detection, 169–179
Entrepreneurship, 102, 106, 186

F
Fusion, 172, 273–284

G
Geometric design, 63–74, 259, 268
Green field analysis, 39–45

H
Health 4.0, 217–240
Hospital, 221, 224, 247, 254, 273–284

I
Information systems, 8, 13, 64, 81–83, 85, 105
Intelligent technology, 63–74
Inventory policies, 47–59

L
Layout conditions, 90
Literature review, 10, 15–33, 49–52, 101–106, 170–173, 181–198, 217–240
Logistic regression, 235, 248, 274, 276, 283

M
Machine learning (ML), 30, 74, 169–179, 218, 248–250, 252, 255, 276–281, 283
Mathematical models, 10–11, 25, 41–42, 65, 66, 69, 74, 114, 115, 117, 162, 264, 268, 269
México, 4, 45, 110–112, 126, 203, 248
mHealth, 218, 223, 225–230, 232–240

J. A. Marmolejo-Saucedo et al. (eds.), *Computer Science and Engineering in Health Services*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-031-34750-4

N
Nanocomplex, 257–269
National Institute of Geography and Statistics (INEGI), 110, 162
Noncommunicable diseases (NCDs), 217–223, 230, 239, 258
Non-fungible tokens, 181–198

O
Operations research (OR), 39, 259
Optimization, v, 6, 12, 17, 26, 27, 29, 49, 54, 64, 65, 67, 69–74, 89, 96, 117, 259, 261, 264, 266, 267
Optimization placement problem, 64, 67
Optimized packing, 72, 74, 89–96, 268

P
Packing, 16, 64, 65, 70, 72–74, 89, 90, 92–96, 259, 261, 264, 266–269
Parkinson's disease (PD), 203–215, 227
Pervasive health, 217
Phi-function technique, 66–68, 72, 259, 262–264
Physical activity (PA), 218
Physical therapy, 203–215
Power systems, 80
PRISMA protocols, 102, 103, 185, 187
Private insurance, 248, 249

R
Remote sensor, 203–215
Resilience in supply chains, 47–59
Reverse logistics, 3–13
Routing, 6, 7, 9, 13, 16, 25, 29, 30, 32, 33, 80

S
Secretary of Environment and Natural Resources (SEMARNAT), 162
Sensor, 203–215, 222, 223, 226–229, 231, 235–237, 273–284
Smartwatch, 227, 274, 275, 283
Soft polygons, 89–96
Solution methods, 7–8, 10–12, 19, 20, 32, 70
Startup, 101–106
Storage location assignment problem (SLAP), 15–33
Stress detection, 273–283
Supply chain, 4, 15, 16, 39–45, 47–59, 106, 112, 115
Systematic review, 16, 24, 101–106, 185, 222, 233

T
Text analysis, 173, 179

W
Warehouse management, 15–33
Wearables, 218, 221–223, 225–240, 275
Word2Vec model, 170, 174, 175, 177–179
Workplace, 232–234, 239, 274, 283