EAI/Springer Innovations in Communication and Computing

Surbhi Bhatia · Ashutosh Kumar Dubey
Rita Chhikara · Poonam Chaudhary
Abhishek Kumar *Editors*

Intelligent Healthcare

Applications of AI in eHealth

EAI
RESEARCH MEETS INNOVATION

Springer

EAI/Springer Innovations in Communication and Computing

Series Editor
Imrich Chlamtac, European Alliance for Innovation, Ghent, Belgium

Editor's Note

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.

More information about this series at http://www.springer.com/series/15427

Surbhi Bhatia • Ashutosh Kumar Dubey
Rita Chhikara • Poonam Chaudhary
Abhishek Kumar
Editors

Intelligent Healthcare

Applications of AI in eHealth

Springer

EAI
RESEARCH MEETS INNOVATION

Editors
Surbhi Bhatia
Computer Sciences and Information Technology
King Faisal University
Alhasa, Saudi Arabia

Ashutosh Kumar Dubey
Computer Science and Engineering
Chitkara University
Rajpura, Punjab, India

Rita Chhikara
Computer Science and Engineering
The North Cap University
Gurugram, Haryana, India

Poonam Chaudhary
Computer Science and Engineering
The North Cap University
Gurugram, Haryana, India

Abhishek Kumar
Computer Science and Engineering Department
Chitkara University
Rajpura, Punjab, India

ISSN 2522-8595 ISSN 2522-8609 (electronic)
EAI/Springer Innovations in Communication and Computing
ISBN 978-3-030-67050-4 ISBN 978-3-030-67051-1 (eBook)
https://doi.org/10.1007/978-3-030-67051-1

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

This book is dedicated to all the editors and their family members. Without their support and cooperation, this book would not have become a reality.

Preface

This book is a small yet significant step towards exploring recent advances, disseminating state-of-the-art techniques, and deploying novel technologies in intelligent healthcare services and applications. It is ideally designed for medical professionals and researchers. The book is divided into three parts, namely Medical Expert Systems, Machine Learning in Healthcare, and Case Studies on Recent Pandemic: COVID 19.

The proliferation of huge data sources (e.g., genomes, electronic health records (EHRs), mobile diagnostics, and wearable devices) and breakthroughs in artificial intelligence applications have unlocked the doors for diagnosing and treating multitudes of rare diseases. There is tremendous scope in the field of intelligent healthcare, and this is evident from the fact that the WHO, in partnership with the International Telecommunication Union (ITU), has launched a Focus Group on Artificial Intelligence in Health (FG-AI4H). The group's objective is to establish a standardized assessment framework for the evaluation of AI-based methods for health, diagnosis, triage, or treatment decisions. Economists have found that medical research can have an enormous impact on human health and longevity, and research to identify and cure disease at early stages should be encouraged.

In view of there being tremendous scope in future of intelligent systems for healthcare, this book fosters a scientific debate for sophisticated approaches and cognitive technologies (such as deep learning, machine learning, and advanced analytics) for enhanced healthcare services. The widespread adoption of intelligent health-based systems could help overcome challenges such as shortages of staff and supplies, accessibility barriers, lack of awareness on certain health issues, identification of patient needs, and early detection and diagnosis of illnesses, creating the potential to accelerate new drug discovery, precision health, and operational savings.

Although artificial intelligence revolutionized the healthcare management, it is extremely difficult to handle and deliver the big health data anytime and anywhere. With the advent of big data, handling of humongous data became a possibility and mobile devices have merged with great potential to renovate the healthcare industry and provide cost-effective, high quality, and more accessible healthcare

Health research has high value to society, and different approaches to research provide complementary insights. Health research has grown exponentially over the years and has led to significant discoveries, the development of new therapies, and a remarkable improvement in healthcare and public health. However, with new diseases like COVID-19 coming up now and then, the scope of research in healthcare continues to be the requirement of the hour. Therefore, the healthcare systems have to be upgraded with new capabilities in emerging areas such as machine learning, big data, and mobile computing for providing mankind with more intelligent and professional healthcare services.

In this book, various researchers have emphasized on the latest and potential health-related topics medically and technologically also. It provides an in-depth knowledge of the interdisciplinary area of healthcare research with more focus on the recent pandemic of COVID-19 which has disrupted our lives and continues to be a threat.

Alhasa, Saudi Arabia — Surbhi Bhatia
Rajpura, Punjab, India — Ashutosh Kumar Dubey
Gurugram, Haryana, India — Rita Chhikara
Gurugram, Haryana, India — Poonam Chaudhary
Rajpura, Punjab, India — Abhishek Kumar

Acknowledgment

The editors would like to acknowledge all the authors for their contributions. We present our sincere gratitude to those who contributed their time and expertise to this book. We also want to thank the reviewers who took part in the review process and gave their valuable comments to the contributors to refine their work and adhere to the quality, coherence, and content of the chapters. Without their support, this book would not have become a reality.

Contents

A Healthcare-Based Intelligent Monitoring Paradigm in Quantum Dot Cellular Automata (QCA) to Protect Against Novel Corona Outbreak

Suparba Tapna

1 Introduction

At present, it tends to be said without a doubt that the risk of novel coronavirus is the most serious threat to humans. In this way, to battle against the spread of novel corona virus (COVID-19) in the network, we have some settled models with the goal that we can vanquish this insidiousness and can limit insignificant harm to humanity. India is presently in Stage 2 of the novel corona virus spread. We have to attempt sincerely to crush this danger with the goal that we do not enter into community transmission stage. We were at incomplete first phase of lockdown stage throughout India until 14 April 2020. After the excessive spread of novel corona virus, the government decided to have the second phase of lockdown till 3 May 2020. It is high time to confine the infection with the goal that we do not enter into community transmission stage. We were confronting issues like individuals not under not control and gathering in large numbers in public places. Considering the above situation, we have made a little exertion to come out with a calculation so that we can battle against this malevolence. This calculation is not new – it is an existing paradigm, yet we have attempted to execute it to overcome the spread of this novel corona virus in India.

S. Tapna (✉)
Department of Electronics and Communication Engineering, DIATM, Durgapur, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_1

2 Background

In this segment, a conceptual outline about fundamental highlights of QCA and associative memory, which is the proposed healthcare-based monitoring paradigm to protect from SARS-COV-2, is introduced.

QCA Concept

QCA is one of the latest and exciting device designs that have a nanometre game plan [2]. This depends on a cell methodology that provides a new system for computing and modifying information [1, 5]. As a consequence of the conditions under which each electron in the cell is located [1], QCA store base is not in any event explicitly used in CMOS. There are four spots at the end of a square cell in a quantum cell. The count of the extra electrons in the quantum dabs is known by the Columbic correspondence. Each dot is a quantity square with a nanometre at any cell side. Quantum will be in a position to tunnel between points through two additional electrons usable in each molecule. The excess electrons require turnstiles due to electron repulsion from one corner to the other. Equal cell information is encoded through action courses of these two electrons (also called cellular polarizations P). They make two extraordinary polarization; $P = -1$ along with $P = +1$ identifying with two-fold states, 0 and 1, independently. Figure 1 shows the two sorts of QCA cells (45-degree, as well as 90-degree) with their two-fold lead.

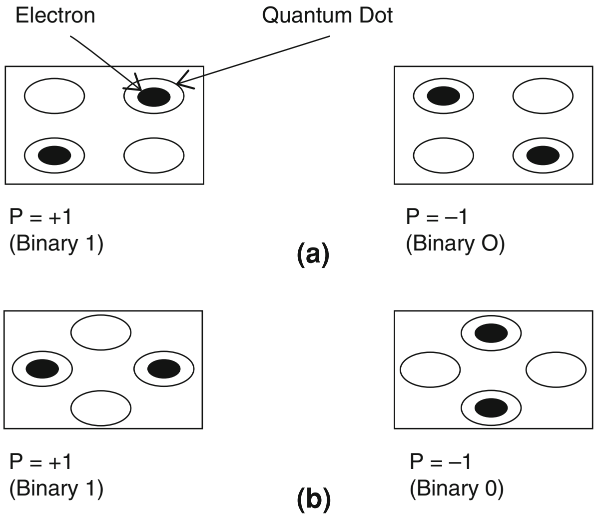


Fig. 1 Polarization conditions of QCA cells: (**a**) 90-degree cells and (**b**) 45-degree cells

The information needed to coordinate data stream bearing (as a figure) needs to be synchronized in order to build progressively complex QCA structures [6]. The QCA clock is used to execute this part and will ensure proper QCA circuit movement. The clock should be used for cell social affairs (date areas). In every region, a bunch of QCA cells under their circumstances shift, and their yield is utilized as an undertaking for the next clock zone.

Associative Memory Computation

Associative memory gets to data utilizing the information substance itself. This is besides recognized as familiar memory, agreeable stockpiling, or partnered bunch. Diverged from finding a thing by its area, the necessary time to find a thing can be diminished widely on the off chance that we recognize the sub-position of that particular item [3, 23]. Due to its agreeable memory affiliation, the memory is phenomenally proper for equivalent endeavours by data affiliation. In cooperative memory, every region as shown in Fig. 2a along with 2b has a limit reasoning circuit's method for coordinating its substance with an external conflict. The all out development of a connected memory to take a solicitation term as the key and in this way returns the memory field [7].

The QCA timing schedule for adiabatic [8, 14] is illustrated in Fig. 3, and expert support was strongly recognized. The clock signal in this structure has four basic stages: Turn, Keep, Release and Relax, each of 90 degrees. The phone is satisfied with the Columbic partnership in the Transfer stage with adjacent cells. The phone stores a mutt lease condition during the hold process by taking a specific polarization. The polarization of cells is decreased and destroyed in the release and relax stages over the long period [14, 15].

CAM Approach Towards Confinement of Associative Memory

CAM gets to data utilizing the data substance itself. This is additionally called associative memory, familiar limit or partnered cluster. Stood out from finding a thing by its area, the essential chance to find a thing can be reduced broadly if we know the item substance [8, 13]. Due to its familiar memory association, the memory is especially suitable for equivalent interests by data alliance. In CAM, every region as outlined in Fig. 2a, b has a limit reason circuits to coordinate substance with an outside contention. The general action of the CAM is to use a term of interest as a key to correctly return the organizational memory region [16].

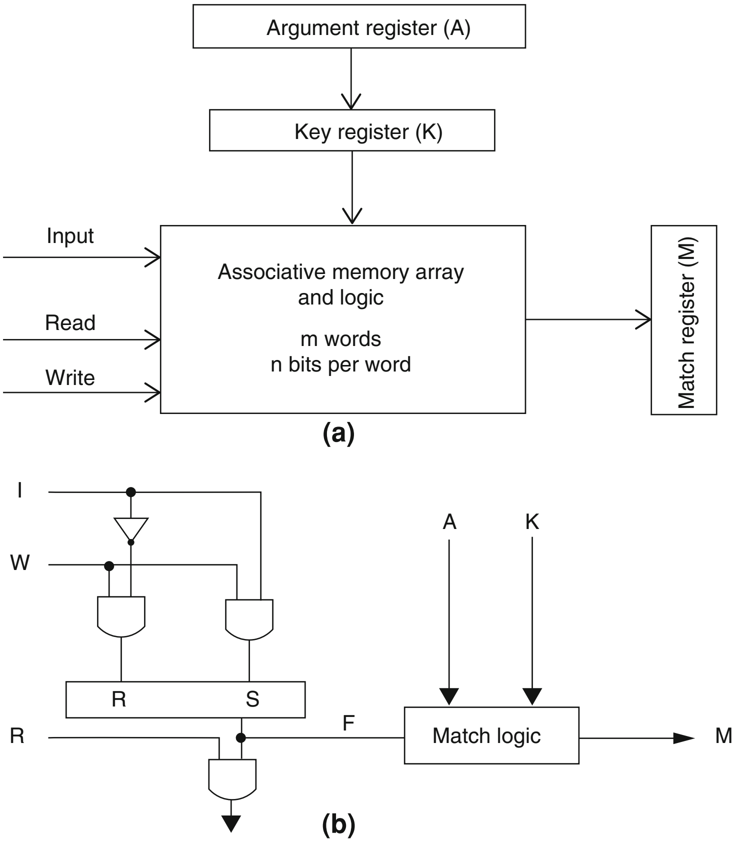


Fig. 2 (**a**) block diagram for associative memory computation. (**b**) match logic for Content Addressable memory

3 QCA Circuits

Here are illustrated the basic sections of the circuits of QCA. The simple QCA methods are then presented. Similarly, QCA evaluates different kinds of memory cell designs.

Basic Elements

The wire, inverter and three-input majority gate are the crucial pieces of QCA circuits. As in Fig. 4a, the 90-degree QCA wire is created by falling QCA cells to cause a multiple respect starting with one side then onto the next de-swinging on Coulomb associations. Figure 4b speaks to 45-degree QCA wire where the polarization of every cell will be reverse of its neighbour cell [17]. Co-planar wire crossing is cultivated using these two sorts of wires in a symmetrical structure as

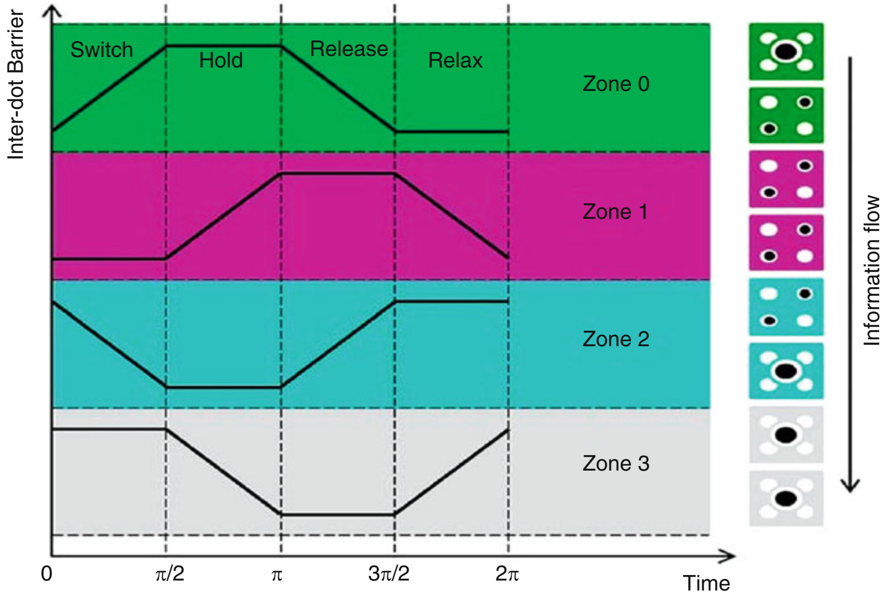


Fig. 3 QCA clocking scheme and its effect on a QCA wire

shown in Fig. 4c [18]. In this procedure, equivalent qualities are prompted in two wires separation.

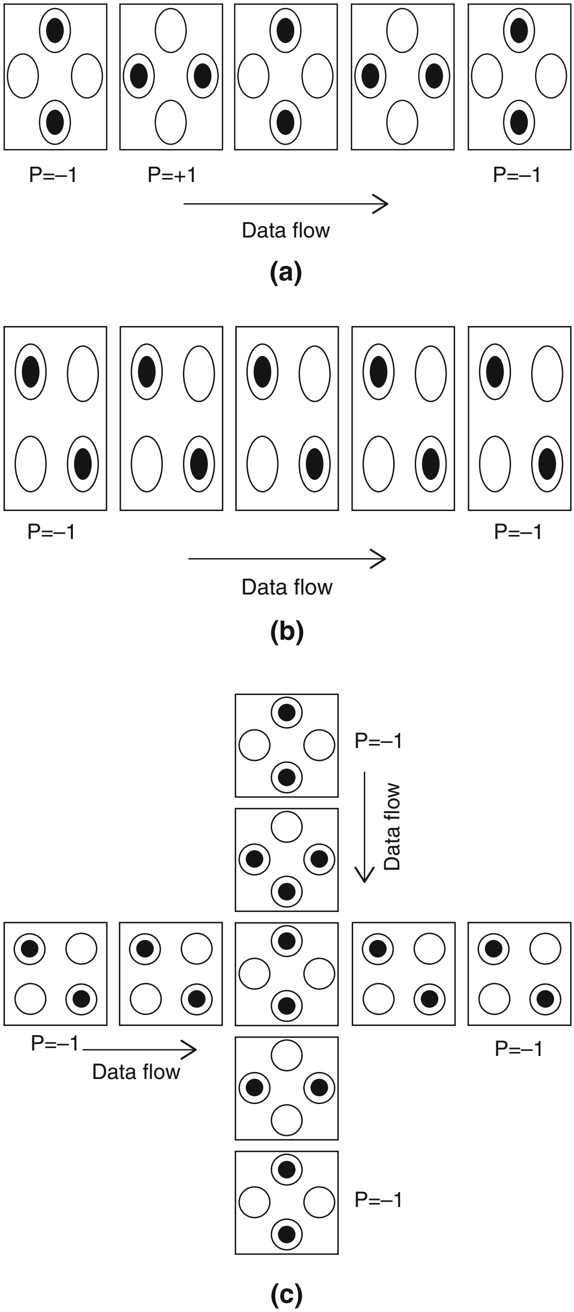
Fundamental Gates

Two key QCA passages, inverter and larger part entryways, are shown in Fig. 5, whereas Fig. 5a represents the inverter's QCA execution. The data signal is isolated into two QCA wires and in this manner, its enhancement appears at the combining point [23]. As in Fig. 5b, QCA three-input larger part gateway finishes the majority gate offer constraint of its three wellsprings of information A, B and C as Maj(A, B, C) = AB + AC + BC. By setting the polarization of one information cell to predictable assessment of −1 or +1, prevailing part passage value acts like a two-input AND OR entryway, independently.

Three sorts of single-stage five-input greater part entryway were introduced in [19, 21], which are shown in Fig. 6. The basis limit of five info larger part door is

$$\begin{aligned}\text{Maj}\,(\text{A, B, C, D, E}) = {} & \text{ABC} + \text{ABD} + \text{ABE} + \text{ACD} + \text{ACE} + \text{ADE} + \text{BCD} \\ & + \text{BCE} + \text{BDE} + \text{CDE}\end{aligned} \tag{1}$$

Fig. 4 Data propagation with (**a**) 90-degree QCA wire, (**b**) 45-degree QCA wire and (**c**) coplanar wire crossing



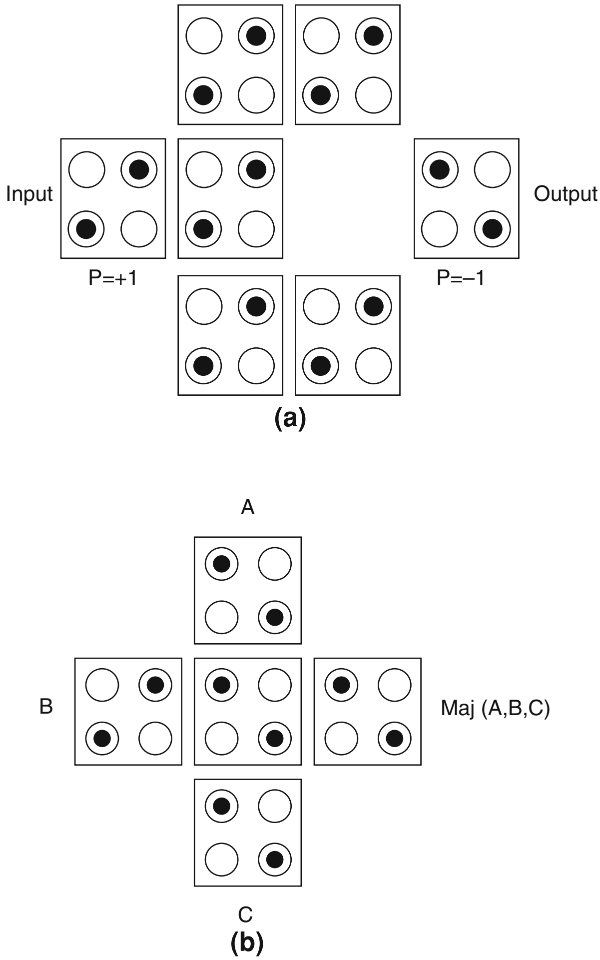


Fig. 5 Two fundamental QCA gates: (**a**) inverter gate and (**b**) three-input majority gate

This is seen in the primary structure in [19], and in Fig. 6a, the larger component five-input entrance is executed using only 10 cells. The give of this game plan can be hence acquired in another layer of the data cells. The next one is seen in [20], and in the Fig. 6b, neighbouring cells are input cells. The improvement plan is presented in [21]. In Fig. 6c, larger part entryway takes every one of the five data flags a solitary route at the principal timing zone.

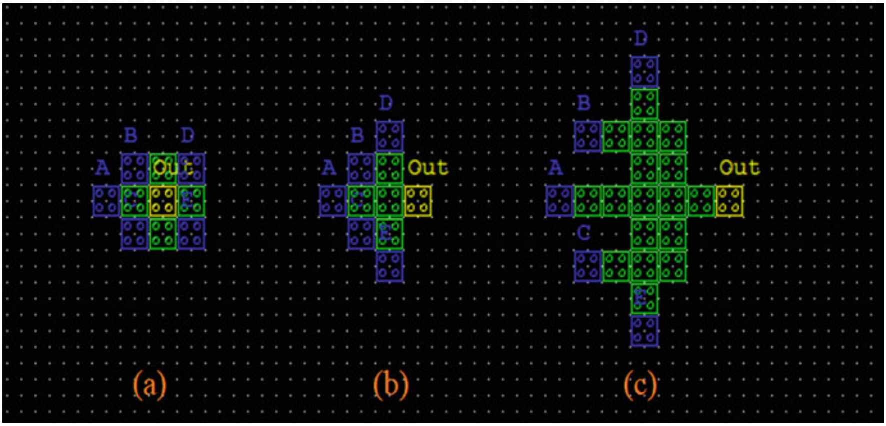


Fig. 6 QCA five-input majority gate: (**a**) the presented structure in [20], (**b**) the presented structure in [21] and (**c**) the presented structure in [22]

Memory Cell Structures

As previously mentioned, QCA is an enticing invention for high thickness but low-power memories as a consequence of the key features of nanotechnology. Along with the circle memory cell [9], the QCA operating directly is called line-based memory cell [10]; in general, there are two types of memory cell designs in QCA. The row-based memory cell stores data bits scattered back and forth on a QCA line [10, 11], as seen in Fig. 7a. Line-based memory cells need additional zones that hinder their implementation. Regardless, the memory cycle-based cell stores information bits broadcast to a QCA cell subtleties anyway that is found in Fig. 7b. Becoming more encouraged when running a circle-based memory cell, they do not require any additional clock areas.

In [12], improved line-based storage cell was suggested. Just two new time zones are essential in this new configuration of the memory cell. Besides, reading throughput for each clock cycle is increased to one step.

Similarly, an addressable S-RAM cell has been implemented in the SQUARES formalism as a memory structure subject to notice multi-content addressable transfer registers. It is not advised to use SQUARES on large circuits because there is an added expense for both time and resource abundance.

A typical unpredictable access (RAM) structure configuration using QCA was seen in [9]. The layout described relies on D-lock, and a circle is used to control the memory information. The data sign to the yield is also extended to give double-crossing intervals.

QCA has been seen as a supporting volatile access memory without a reset and fixed limit [22]. This framework also has a circular-based instrument, which is SR-lock-based. This arrangement has been based on a coplanar wire-crossing technique.

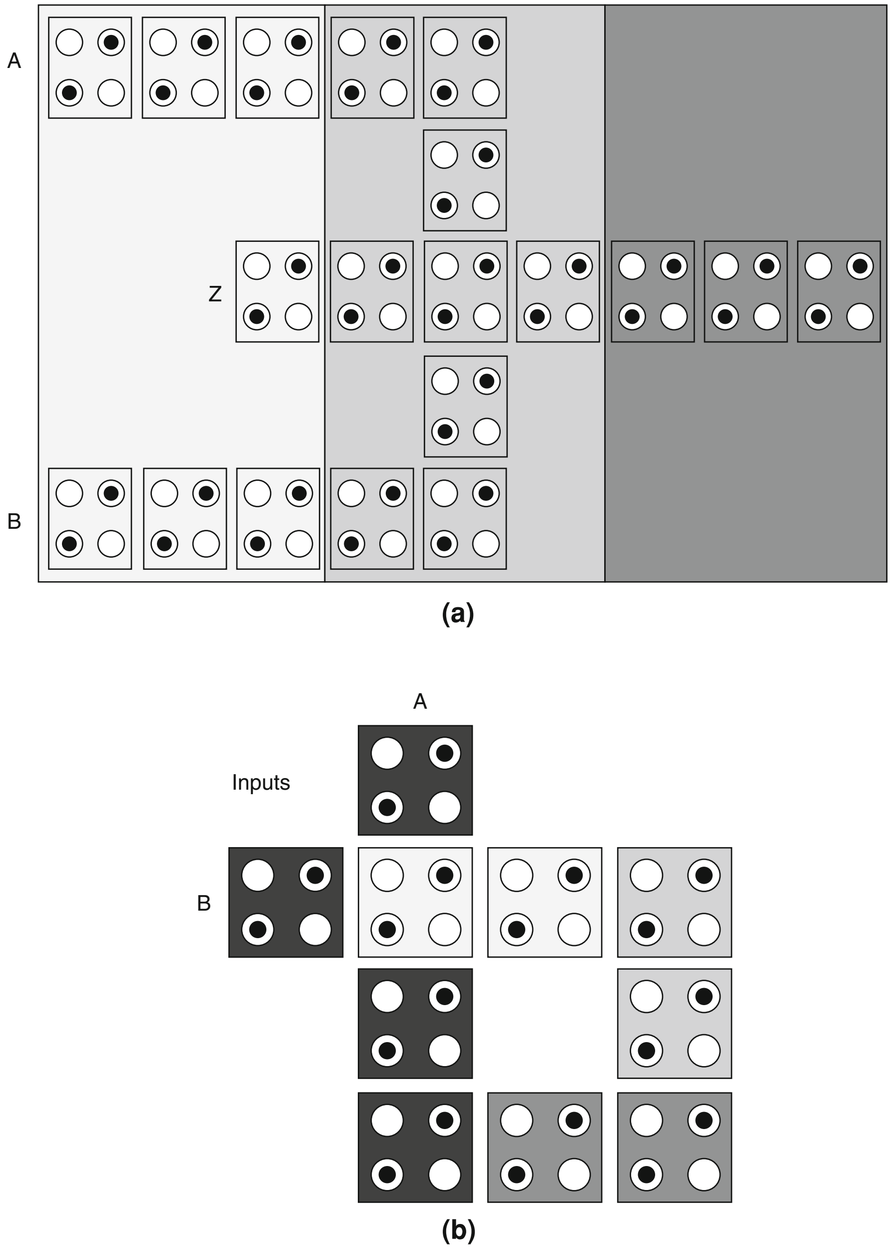


Fig. 7 Two common types of memory cell architectures in QCA: (**a**) line-based memory cell and (**b**) loop-based memory cell inside the cell. The burrowing boundaries' statures are constrained by QCA clock

A comparative behaviour is seen in [22] in another system for RAM cells. This structure was designed on a D-layer with a three-input, unbalanced structure of the doorway.

The set and reset limit structure of the Ram cell was introduced in [23]. Two 2:1 multiplexers make up this framework. A modern cording method has been implemented for the new architectures of Flip Flops and RAM for a reduced number of cells [24]. An energetic five-input standard component entrance is implemented in a revolutionary work [13], which is planned to use simple and efficient QCA circuits alone.

A new RAM cell structure with fixed and reset cap was suggested by the use of this structure. Our work depicts a CAM that does not necessarily compare to RAM, since any memory zone in the CAM is tuned to the substance as opposed to its addresses, not necessarily like the works depicted above that use the RAM model, where the memory region is in a specific location.

Five-Input Minority Gate

In QCA proposals, essential structures used to render the various essential QCA entries are historically referred to as the dominant component gateway. A five-input minority entrance is consequently presented around there. The five-input minority entrance is focused on

$$\begin{aligned} \text{Min}\,(\text{A, B, C, D, E}) = {} & (\text{ABC} + \text{ABD} + \text{ABE} + \text{ACD} + \text{ADE} + \text{BCD} \\ & + \text{BCE} + \text{BDE} + \text{CDE})' \end{aligned} \tag{2}$$

A three-input NAND passage, as well as three-input NOR entry, can be modelled separately by setting two of the five data cells' polarizations to -1 or $+1$. Table 1 provides a table of realities of a five-input passage of the minority subject to all data sources.

Five-input minority passage may be made after changing the five-input lion's share gateway. In the chief arrangement of the five-input minority entryway that is shown in Fig. 6a [19], the data cells thus prevent admittance to it in one layer surround cell. The next one is seen in [20], and in Fig. 6b, input cells border each other while inducing unintended results (e.g. inputs B and D or sources C and E). This is seen in the action plan in [21] as seen in Fig. 6c, and each of the five data signals takes one direction at the chief planning area, the entrance to Lion's bid. The findings are not generous enough that this arrangement is extended to create a minority entrance so that a minority gateway is not feasible.

Figure 8 demonstrates our suggested configurations for the five-input prevailing entry and five-input minority doors. In the planned five-input greater entry, there are 20 and 22 cells, dividing five-input minority entry. Input cells consist of five

Table 1 Truth table for proposed five-input minority gate dependent on the whole of the information esteems

Sum(A,B,C,D and E)	Min(A,B,C,D and E)
0	1
1	1
2	1
3	0
4	0
5	0

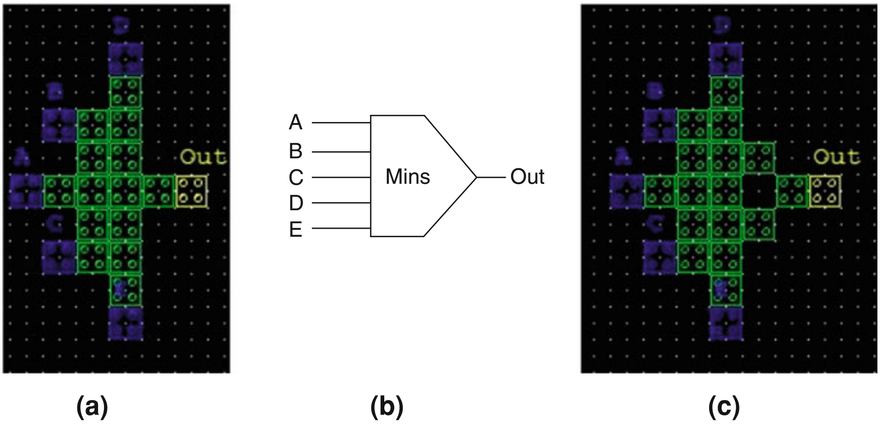


Fig. 8 Advanced designs: (**a**) QCA layout of five-input majority gate, (**b**) schematic of five-input minority gate and (**c**) QCA layout of five-input minority gate

Table 2 Data for healthcare monitoring in test feedback and outcomes [23]

Types of operation	R/W	I	Previous F	F	O
Write	0	1	x	1	0
Write	0	0	x	0	0
Read	1	x	1	1	1
Read	1	x	0	0	0

cells denoted as A, B, C, D and E, one of which is the input cell and the other cells are the gadget cells in the two systems. In addition, information cell polarization is constant, and gadget cells and the yielding cell can be modified (Table 1).

Table 3 Data for healthcare monitoring in SARS-COV-2 detection

K	A	F	M
0	x	x	1
1	0	0	1
1	0	1	0
1	1	0	0
1	*1*	*1*	*1*

Five-Input Minority Gate-Based Multilayer CAM Cell

The expected and QCA execution of the CAM cell is suggested in Fig. 8a, b. The device consists of six key doors with three inputs and one minority portal with five inputs, which is seen in Fig. 8c. There are two portions in this circuit, one being a memory device and the other being an organization. The memory unit gets the read/make signal (as R/W and I). The simple unit receives memory cell information (as F) and the conflict and main signals. The memory unit gets a read/make signal (as F). The circuit offers the content of the cell as an output (divided into O) when a data was reviewed, and the expansion partner decides the corresponding signal (divided into M) when the information was detected. Table 3 displays the brain behaviour reality, and Table 4 displays the coordination function reality.

The input data are commuted to the F yield when the R/W signal is set to ‘0’ in the proposed plan and activities are thus produced. The R/W symbol was set to ‘1’. The perused motion was performed on both sides. Estimation of past F is submitted in reading motion to yield F and yield O. Only as the K signal is set to ‘0’, the equivalent symbol M is set to ‘1’ and will send little brain to the evaluation of An and F. If K is set to ‘1’, the match signal M will be set to ‘1’ on the off chance that the assessments of A and F are equivalent, and where the assessments of A and F are not proportionate, it will be set to ‘0’. Otherwise it is set to ‘1’.

4 Scope

This section describes the entire methodology, which is based on the healthcare-based monitoring to protect from novel corona virus (COVID-19). The possibilities of nanotechnology-based worldview is enhanced about the surroundings associated with memory computation does rely on a controlling mechanism into a ventilator device and functioning in an intelligent healthcare system (Fig. 9).

Table 4 Simulation parameters

Parameters	Value
Temperature	1.000000K
Relaxation time	1.000000e-015s
Time step	1.000000e-016s
Total simulation time	7.000000e-011s
Clock high	9.800000e-022J
Clock low	3.800000e-023J
Clock shift	0.000000e+000
Clock amplitude factor	2.000000
Radius of effect	80.000000nm
Relative permittivity	12.900000
Layer separation	*11.500000nm*

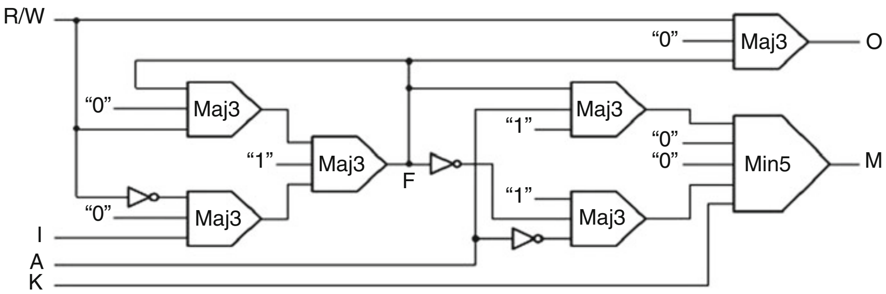


Fig. 9 Proposed paradigm for ideal healthcare monitoring

5 Proposed Paradigm of Healthcare Monitoring

To carry out the proposed worldview for intelligent healthcare monitoring of associative memory-based computation into ventilator controlling device the parameters to such following the below phenomenon.

R/W = fetch the test data from samples/study thoroughly the samples data and compare to CORONA symptoms.

I = controlling the operation.

A = for conveying the message for searching the corona symptoms from a patient's body.
K = specifies which portion is majorly affected in patient's body.
F = feedback for test.
Maj3 = sample from patient's body.
0 = report negative
1 = case positive
Min5 = sample taken finally from the patient's body for rapid test.
O = possible outcome.
M = match the data either '1' = > case positive, when '0' = > report negative.

6 Entire Operation for Process Flow of Implemented Design in Quantum Dot Cellular Automata (QCA) Nanotechnology

The entire process flow of implementing design describes the phenomenon related to ventilator-based controlling in intelligent healthcare memory monitoring system involved with the associative memory implication towards high-performance computing for protecting the spread of novel corona in the entire world. The implementation in QCA-based nanotechnology does rely on at given input, which is to be considered as to collect the sample from a human body. Later on it is to get at read operation to fetch the test data from the patient's body and study, the tested sample for SARS-COV- 2 thoroughly. At the end, the possible outcome to this methodology is to realize that if say '1', the patient is corona affected and if '0', the patient is out of danger from this pandemic disease.

To synchronize the design, there are different clock regions in QCA clocking. Also, different colour zones have been associated with the segment of the area to protect against COVID-19 pandemic. The first colour zone 0 'green' is free of novel corona disease; second, zone 1 'pink' is recognized as the hotspot area and complete lockdown during the vital stage. The next zone 2 'blue' is considered as it might get affected, but growth is possibly less. The last considerable zone 3 'grey' relies on the rate of growth in the entire 24 h of operation (Fig. 10).

7 Results

Recreation is checked in this region using the QCA Designer type 2.0.3 [4] according to the impact of the proposed structures. The structures were rebuilt using the Coherence vector reproduction motor along with Bistable Approximation re-enactment motor with default parameters, and comparable findings from the two re-enactment motors have been obtained. In our proposed QCA model, the standard parameters were considered (Fig. 11).

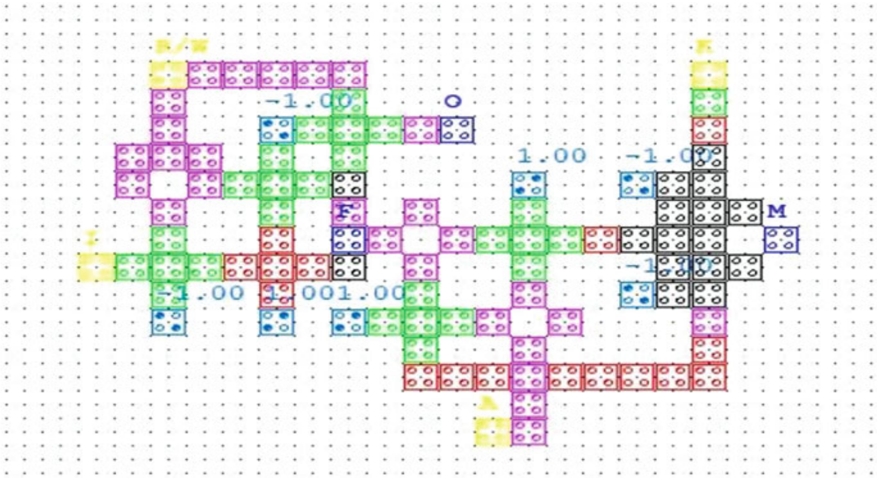


Fig. 10 Proposed paradigm for ideal healthcare monitoring in QCA

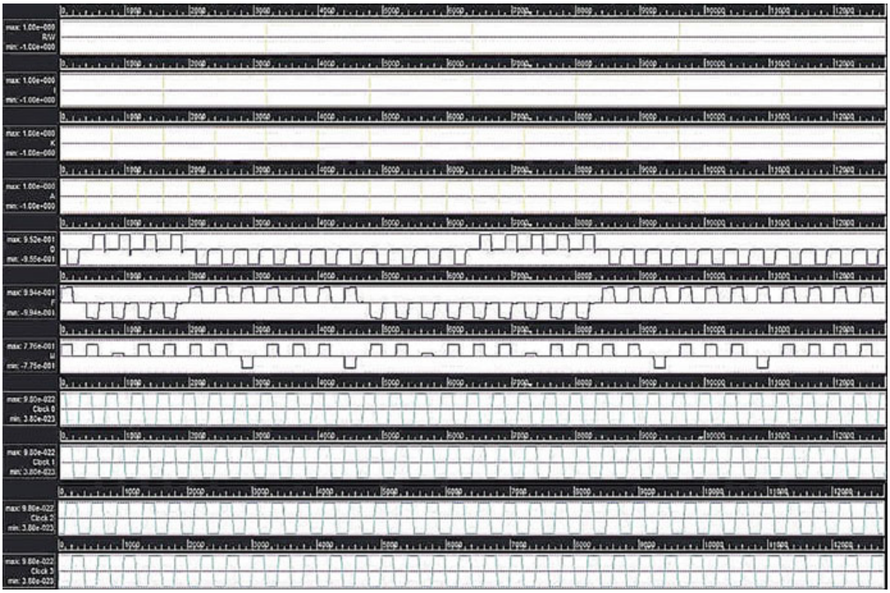


Fig. 11 Simulation of the proposed model for ideal healthcare monitoring in QCA

The above suggestion towards the technique adopted in QCA for monitoring intelligent healthcare system, we have to show two case strategies in Fig. 12, that is, for the feedback of the test report and probable outcome. In another depiction, Fig. 13, case relies on for the identification of SARS-COV-2-affected patient.

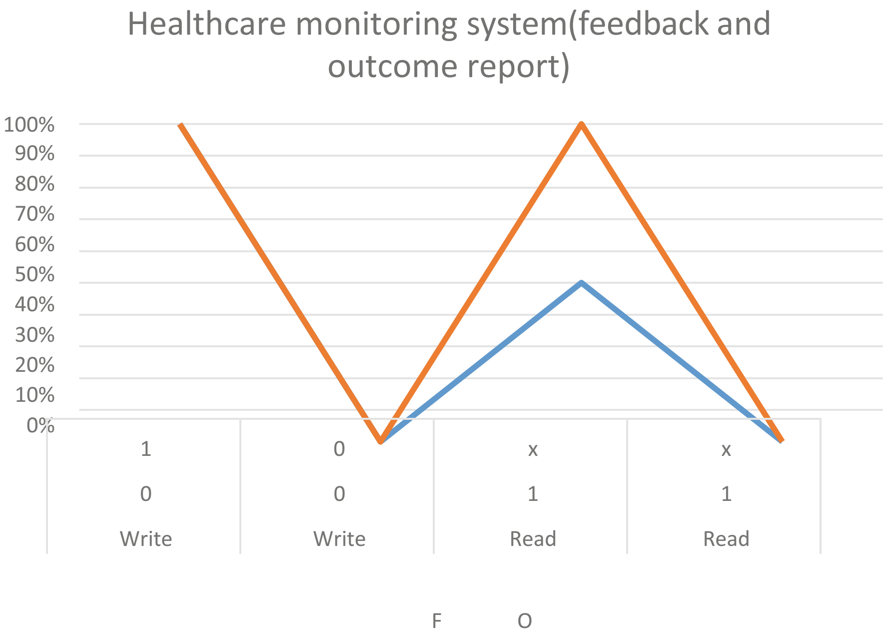


Fig. 12 Healthcare monitoring in feedback and probable outcome

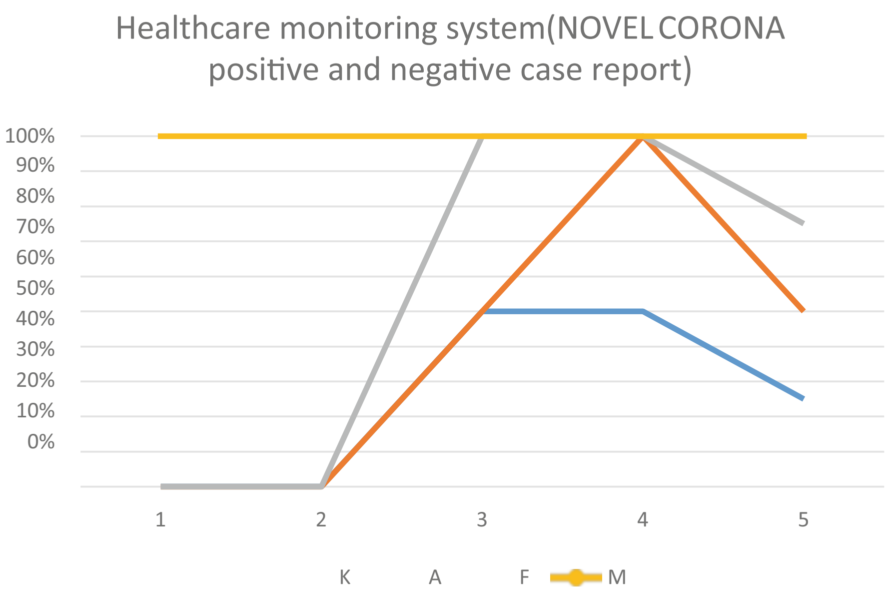


Fig. 13 Healthcare monitoring for detection of novel corona

Risks

(a) Government workers are suppressing realities or some way or other green-coded representative gets CORONA contaminated.

- Mobile or ISP ought to guarantee consistently web network in the portability of worker.
- His/her portable area chronicles to discover where he had visited the past 7 days.
- All the spots ought to be cleaned where he/she had visited, and all the people who came in contact to be set in isolation.

(b) Proper treatment of the tainted representative.

- Any resident is suppressing realities or some way or other green-coded resident gets corona contaminated.
- Mobile or ISP ought to guarantee consistently web availability in the portability of green-coded resident.
- His/her portable area chronicles to discover where he had visited the past 7 days.
- All the spots ought to be disinfected where he/she had visited, and all the people who came in contact to be set in isolation.
- Proper treatment of the contaminated resident.

(c) Legitimate well-being checking measures to yellow-coded residents or representatives.
(d) If any yellow-coded resident gets contaminated, all the relatives need to be isolated.
(e) Neighbourhood organization will guarantee that yellow-coded individuals ought not to go out from home. In case ignores do accordingly, by then indistinct steps to be taken from green-coded individuals if any yellow-coded individuals got contaminated.

8 Conclusion

Let us trust in the best to abstain from spreading the crown infection into community transmission stage in our nation. Legitimate cooperation and appropriate solidarity among us are extremely fundamental for the accomplishment of this model. All hands should be brought together firmly to battle against this insidiousness. Any commitment opening with respect to anybody will have a genuine impact and will provoke frustration of this model in this crisis circumstance.

References

1. P. Tougaw, C. Lent, Logical devices implemented using quantum cellular automata. J. Appl. Phys. **75**(3), 18181825 (1994)

2. C. Lent, P. Tougaw, A device architecture for computing with quantum dots. Proc. IEEE **85**(4), 541557 (1997)
3. L. Chisvin, R.J. Duckworth, Content-addressable and associative memory: Alternatives to the ubiquitous RAM. IEEE Comput. **22**, 5164 (1989)
4. K. Walus, T.J. Dysart, G.A. Jullien, R.A. Budiman, QCA Designer: A rapid design and simulation tool for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **3**(1), 2631 (2004)
5. A.O. Orlov, I. Amlani, G.H. Bernstein, C.S. Lent, G.L. Snider, Realization of a functional cell for quantum-dot cellular automata. Science **277**, 928930 (1997)
6. C.S. Lent, P.D. Tougaw, W. Porod, Bistable saturation in coupled quantum dots for quantum cellular automata. Appl. Phys. Lett. **62**(7), 714716 (1993)
7. K. Pagiamtzis, A. Sheikholeslami, Content-addressable memory (CAM) circuits and architectures: A tutorial and survey. IEEE J. Solid State Circuits **41**(3), 712–727 (2006)
8. R. Compano, L. Molenkamp, D.J. Paul, Technology roadmap for nanoelectronics, in *European Commission IST Programme, Future and Emerging Technologies*, (2000)
9. C.S. Lent, P.D. Tougaw, W. Porod, G.H. Bernestein, Quantum cellular automata. Nanotechnology **4**(n1), 49–57 (1993)
10. M.T. Niemer, P.M. Kogge, Problems in designing with QCAs: Layout = timing. Int. J. Circ. Theory Appl. **29**, 49–62 (2001)
11. G.H. Bernstein, A. Imre, V. Metlushko, A. Orlov, L. Zhou, L. Ji, G. Csaba, W. Porod, Magnetic QCA systems. Microelectron. J. **36**, 619624 (2005)
12. J. Huang, F. Lombardi, *Design and Test of Digital Circuits by Quantum-Dot Cellular Automata* (Artech House Inc, Norwood, 2007)
13. W. Liu, L. Lu, M.O. Neill, E.E. Swartzlander Jr., Design rules for quantum-dot cellular automata, in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, (Rio De Janeiro, Brazil, May 2011), pp. 2361–2364
14. K. Kim, K. Wu, R. Karri, Towards designing robust QCA architectures in the presence of sneak noise paths, in *Proceedings of the IEEE Design, Automation and Test in Europe Conference and Exhibition (DATE)*, (Munich, Germany, March 2005), pp. 1214–1219
15. D. Berzon, T.J. Fountain, A memory design in QCAs using the squares formalism, in *Proceedings of the Great Lakes Symposium on VLSI*, (March 1999), p. 166169
16. V. Vankamamidi, M. Ottavi, F. Lombardi, A line-based parallel memory for QCA implementation. IEEE Trans. Nanotechnol. **4**(6), 690698 (November 2005)
17. B. Taskin, B. Hong, Improving line-based QCA memory cell design through dual phase clocking. IEEE Trans. VLSI **16**(12), 16481656 (2008)
18. K. Kim, K. Wu, R. Karri, Quantum-dot cellular automata design guideline. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E89**(A(6)), 16071614 (2006)
19. V. Vankamamidi, M. Ottavi, F. Lombardi, Two-dimensional schemes for clocking/timing of QCA circuits. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **27**, 3444 (2008)
20. I. Amlani, A.O. Orlov, G. Toth, G.H. Bernstein, C.S. Lent, G.L. Snider, Digital logic gate using quantum-dot cellular automata. Science **284**, 289291 (1999)
21. K. Kim, K. Wu, R. Karri, The robust QCA adder designs using composable QCA building blocks. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **26**, 176183 (2007)
22. M.A. Dehkordi, A. Shahini Shamsabadi, B. Shahgholi Ghahfarokhi, A. Vafaei, RAM ovel cell designs based on inherent capabilities of quantum-dot cellular automata. Microelectron. J. **42**, 701708 (2011) (Elsevier)
23. S. Hashemi, K. Navi, New robust QCA D ip op and memory structures. Microelectron. J. **43**, 929940 (2012) (Elsevier)
24. S. Angizi, K. Navi, S. Sayedsalehi, A.H. Navin, Efficient quantum dot cellular automata memory architectures based on the new wiring approach. J. Comput. Theor. Nanosci. **11**(11), 2318–2328 (2014)

Intelligent Healthcare

Yogesh S. Kale, Shubhangi V. Rathkanthiwar, and Prachi D. Gawande

1 Introduction

Remote Intelligent Healthcare, also known as telemedicine, requires diligent care and effort and is the largest growing area of study and expertise with growing applications in Biomedical Engineering and Medical Sciences using the latest technology – the Internet of Things (IoT). The basic definition of telemedicine is an exchange of clinical/medical-related information from one destination to another destination using information and communication technology (ICT). Population analysis indicates the percentage of people over the age of 65 will increase by approximately 38% in Japan, almost 29% in Western Europe and 22% in the U.S. by 2050. There will be an increase in health care needs related to the aging population in society. This chapter predicts the development of intelligent health care will increase in proportion with the increase in the aging population. Some important areas where telemedicine can be proficiently used are:

- Tele Support
- Tele Monitoring
- Tele Diagnostics
- Tele Treatment
- Tele Consultation
- Tele Education
- Tele Training

Y. S. Kale (✉) · P. D. Gawande
Department of Electronics & Telecommunication Engineering, Yeshwantrao Chavhan College of Engineering, Nagpur, India

S. V. Rathkanthiwar
Department of Electronics Engineering, Yeshwantrao Chavhan College of Engineering, Nagpur, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_2

Fig. 1a A powerful & friendly Bluetooth blood pressure monitor module

Tele monitoring enables medical experts to remotely examine the patients and their needs within the comfort zone of the patient as well as medical expert. With the help of tele monitoring, transportation charges, transportation time and waiting time will be reduced to zero; in most cases, the transportation charges are more than the doctor fees. This new method of investigation has exploded over the past two decades because of the increasing doctor to patient ration, aging populations and, in countries such as India, because of the inaccessibility of hospital centres. Most of the research indicated that tele monitoring, telemedicine, tele diagnostics, tele support, tele consultation, tele education and tele training are more cost effective than hospitalization. Tele monitoring is based on the broadcast and reading of medical indicators. The broadcast reading analysis may lead to the judgment that the patient should be hospitalized or that only advice is required.

Tele monitoring supports continuous monitoring for urgent situations and serious care patients, and it includes tele care for elderly people, continuous monitoring, examination of chronic diseases, such as for people with cardiac diabetes or dysfunction, while the continuous medical care of pregnant women sometimes requires their hospitalization, or they need to visit hospital every day in order to use cardiotocography (CTG) and pulse oximeters. The tele monitoring system uses physiological sensors to measure blood pressure, weight, temperature, blood oxygen saturation, etc. Some of the I Health Care System Modules are shown in Figs. 1a and 1b [1].

In the 1990s, basic tele monitoring used standard telephone lines, and it relied on a very limited set of vital signs. At the beginning of this century, it moved to the internet and now is incorporated into various services, especially for chronic patient

Fig. 1b Gluco-monitoring system with Android application

follow-up. In the era where everything is available with a single click, this system is available on wireless and mobile technologies and is also called the portable and home-based system. The next update in this system is the use of wearable sensors in which sensors can provide the reading in an easily readable form. Presently, the overall research focus is on emergent implantable and wearable sensors, which will play a crucial role in achieving continuous measurement of vital sign parameters throughout the entire day without limiting the mobility of patients.

Tele monitoring will help to address the increasing patient to doctor ratio that is expected to reach 12.9 million in the next 10 years, as suggested by the World Health Organization (WHO) in 2013 [1]. The above condition is owing to the rapid population increase, and when the total population increases, the aging and disabled population also increases. In addition, if a country has a good healthcare system, then it is possible to reduce the number of hospital visits and hospitalizations with the help of Information and Communication Technology (ICT). The UK has had excellent results after implementing intelligent healthcare according to the UK National Health Service (NHS), and by using intelligent healthcare, they have saved seven billion pounds per year [2].

In a similar way, to better analyse how patients could avoid needless admissions to the Intensive Care Unit (ICU) because of chronic conditions, Veterans Health Administration (VHA) conducted a Care Coordination/Home treatment (CCHT) survey from 2003 to 2007. Further, they analysed medical-related data regarding the stage of schedule and pre-eminence from a cohort of 17,025 CCHT patients who participated in the survey and found there was a 25% decrease in the number of ICU bed days and a 19% decrease in hospital admissions. Veterans Health

Administration (VHA) received a very good response to achieve the technological change in the medical field, and the survey concluded that home tele-health accomplishes providing an extremely caring and reasonable approach for assisting chronic care patients (CCP) in rural and urban areas [3].

Now, the upcoming technology is wireless, and the new tele monitoring research is based on a wireless patient monitoring system (WPMS). In the medical field, contactless treatment should always be a secure treatment and this system plays a vital role in early diagnosis of diseases and physical conditions by relating information to the healthcare experts [4]. The main function of WPMS is to competently: (1) gather medical-related clinical data using predefined medical sensor devices; (2) recognize the medical-related clinical data in consideration mode; (3) obtain the early warning score (EWS) from medical professionals who have access to the data; (4) carry out appropriate action through medical experts or professionals [5, 6]. This is the way to shift from the traditional method and adopt the smart method, and it will include calculating EWS, which helps to predict early detection of chronic diseases [7]. It has been found that ambulatory blood pressure monitoring (ABPM) offers nonstop measurement of blood pressure at a predefined time gap for 24-h periods or longer. Mainly, this system helps to identify white-coat hypertension, which shows unusual systolic and diastolic averages in those who are not taking antihypertensive medications. This system gives measurement readings in any window size frame, which will be very helpful for the medical professionals; it will also help clinical supervisors to categorize and identify chronic diseases such as hypertension and boost the correctness of diagnosis [8]. The proposed intelligent healthcare is a very cost effective IoT-based system. The main function of this system is to easily connect with digital medical sensor devices and send the measurements to the cloud. Most of the time patient essential parameter readings such as body temperature, pulse rate, diastolic and systolic pressure will be sent to medical professionals through a global cloud server.

Figure 1c shows the workflow of an intelligent healthcare system for remote patient monitoring; the smart home consists of IoT devices and medical sensors. Digital medical sensors sense the patient readings and, for the purpose of analysis, send it to the cloud server through the intelligent healthcare system, and this clinical data is received at, for example, a hospital where medical authorized professionals analyse and compare the patient parameter readings with the data that are already available in the hospital data base and convey to the patient to visit hospital or not. If the patient readings are normal or nearest to normal, then there is no need to visit hospital; for the condition of nearest to normal, medical professionals could remotely prescribed medicine to the patient.

2 Research Method

Many researchers have published papers on intelligent healthcare and many more are working on it; this segment presents some of the significant works on intelligent healthcare. Digital medical devices use Bluetooth (BT) and USB from the intelligent

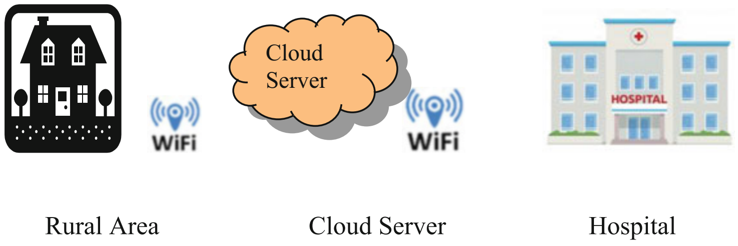


Fig. 1c Workflow of an intelligent healthcare system for remote patient monitoring

healthcare domain, and the limitations include the number of devices connected, range and power [9, 10]. Today's researchers are focused on the need for wireless technology. For example, the Zig-Bee device has a good transmission and reception range, allows connectivity for a good number of devices, has low power consumption and supports intelligent healthcare domains on a solo wireless technology [11]. Alesanco [12] worked on electrocardiogram (ECG) signals and using 3G cell phone network simulations and analyses. The author developed a transmission procedure and transmission algorithm to fulfil the medical trial condition regarding ECG signals, and as per the medical trial's criterion, 4 s is the maximum common delay in ECG signal. Trigo et al. [13] obtained a trial object implementation of concurrent ECG transmits through the IEEE 11073 standard family, where synchronization on behalf of the ECG device branch of learning continues. The same practice formation was not clinically evaluated, but put into practice in standard feasibility [14, 15].

3 Labelling or Classification of Vital-Sign Data

Data classification is a very important aspect of vital-sign detection, and accurate labelling of normal and abnormal data is required for the data used to calculate the early warning score (EWS). There are two types of data classification: one is one-class classification, where only the abnormal label is used for demonstration. When we consider one-class classification, the total calculation depends upon abnormal data only, but for the analysis of patient data, normal data must also be in the data set. On the other hand, in two-class classification (TCC), if we have an ample sample of abnormal data, joint normal and abnormal sample data are used, and the abnormal bunch is evidently modelled. To accurately calculate or predict the EWS, data labelling is a very important aspect when we compare the patient data sets with the available clinical data sets. A huge quantity of data is collected from uninterrupted patient monitoring, and thus realistically managing clinical data sets for every patient in the hospital is monotonous work for any medical professional. Therefore, to conquer this task, it is better for medical clinicians to mark only those patients for whom known abnormalities occurred [16], occasionally only those patients having simple univariate alerting criteria, for example, heart rate (HR) $\geq$ 120 beats per minute, were renowned or accepted for the period of retroactive trying [17, 18]. Many authors want to predict the early warning score (EWS) from the vital signs; for the detection of chronic diseases, there is always a possibility of over-scoring as well as under-scoring. Here, under-scoring creates delay in the detection of worsening patient health and over-scoring leads to gratuitous calling of medical personnel [19]. It is not totally dependent on the exercise of stiff patient results; conceivably, on discharge, collection labels may be offered because many times it happens that patients have negative results such as unanticipated cardiac arrest or urgent situation admission to ICU; there can be periods of normal physical processes, believably prior to their deterioration in the patient health, and in a similar way with the periods of abnormal physiology prior to complete renewal.

Many review papers discuss the 4-D dataset investigation involving breathing rate (BR), oxygen saturation (SpO2) and heart rate (HR); the systolic–diastolic average (SDA) is the arithmetic mean of the two blood pressures. Three successive run-throughs conducted between November 2006 and August 2007 at the Presbyterian Hospital, University of Pittsburgh Medical Centre (UPMC), collected a dataset involving over 18,000 h of 4-D vital-sign data samples from 332 patients that were admitted into the ICU because of various diagnoses [20]. The main aim of this study was to obtain all basic vital-sign measurement parameters, including oxygen saturation (SpO2), heart rate (HR), diastolic and systolic blood pressure (BP) that were generated by their proposed remote system with a sampling rate of 20 s. Most of the time blood pressure is also measured noninvasively with a hot-air balloon pump attached to the bedside screen, and diastolic and systolic blood pressures are simultaneously measured by it. The conclusions of this survey were determined after consulting with medical experts or professionals and rejection criteria included rejecting sample data outside the following ranges: SDA 20–180 mmHg, HR 30–300 bpm, SpO2 60% and higher was acceptable [20].

Existing One-Class Classification (OCC) Methods

Existing one-class classification (OCC) is used to outline a model of normal data, and afterwards detect deviations from that model and document them as being abnormal according to previous works. A regular approach is emerging which involves using vital-sign sample data [21–23].

Existing Two-Class Classification (TCC) Methods

The conclusions regarding TCC approaches of labelled patient's data are insufficient because there are very few examples of TCC approaches to the problem of patient vital-sign monitoring (PVSM). Parati G et al. concluded that a huge set of data is needed to allow the abnormal class to be correctly modelled. It has been noted that at the boundary level, TCC is achieved in which all types of abnormalities are explicitly modelled. To create such modelled requisite, adequate sample data that would also not group up to data sets of higher dimensionality requires connotation, summing together a unique vital sign would start on further modes of abnormality due to further vital signs; hopefully, covariance can be decided by modes of abnormality starting from other vital signs [24].

4 Results

The major aim of this chapter is to comprehend telemedicine programs in the framework of chronic diseases that are accurately linked with the existing medical or clinical settings. This chapter discusses the implementation of a basic model, at a very low cost, for telemedicine by using the latest technology – the Internet of Things (IoT) and ThingSpeak. ThingSpeak is a widely accepted and open source IoT platform where medical professionals can remotely obtain basic parameter readings to observe and analyse patient vital signs. This platform also allows users to collectively analyse and visualize live data streams in the cloud. This IoT module helps in continuous measurement of basic parameters involved with chronic diseases and the basic parameters such as blood pressure (BP), which includes diastolic and systolic, pulse rate and body temperature throughout the entire day without limiting the patient mobility or removing them from their comfort zone. Figure 2 shows the basic module for an intelligent healthcare system. This cost effective system consists of ESP32. Figure 3 shows the detailed pin diagram of ESP32, which is a very popular controller because of its small size, low cost and having microcontroller as well as Wi-Fi module capability in the same package. ESP32 was invented by Espressif Systems, a Shanghai-based Chinese company, and manufactured using TSM Cusing, which is their 40 nm process [25, 26].

A demonstration of a low-cost IoT module for a smart healthcare system with blood pressure sensor module is shown in Fig. 4.

Graphical Analysis of Measurement of Systolic, Diastolic and Pulse Rate Parameters

Blood Pressure (BP) is the force of blood against the walls of arteries. Blood is pumped in through a patient's arteries every time their heart beats. There are two types of BP: systolic and diastolic. The systolic pressure is a measure of when a patient's heart pumps blood, and a patient's BP is highest; the diastolic pressure is a measure of when the patient's heart relaxes between beats, and the patient's BP

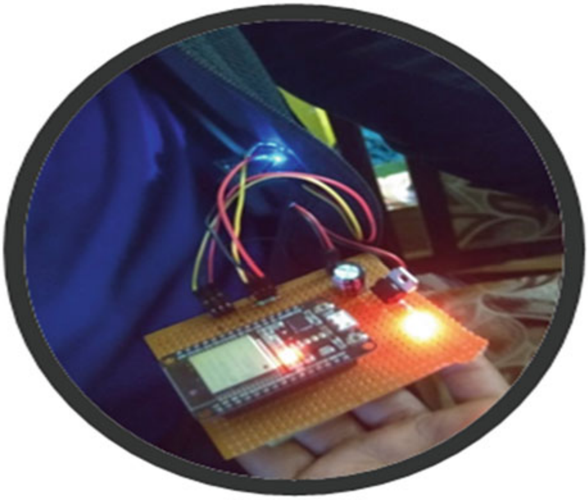


Fig. 2 Assembly of low-cost IoT module for a smart healthcare system

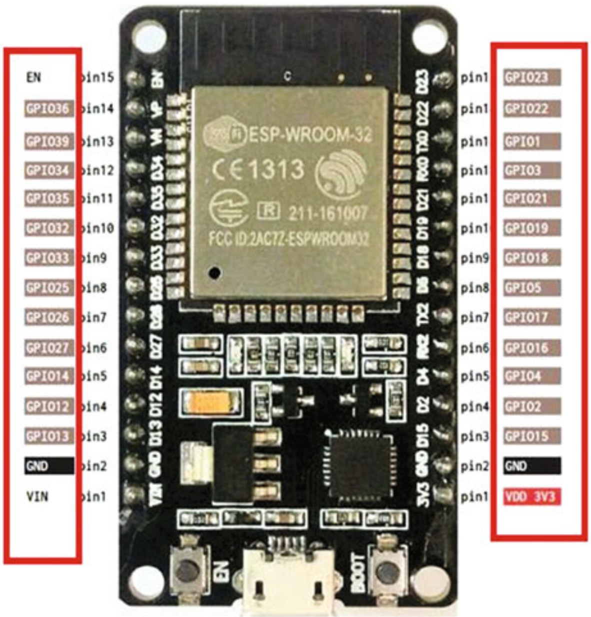


Fig. 3 Standard pin diagram of ESP32

Fig. 4 Demonstration of a smart healthcare system with low-cost IoT module

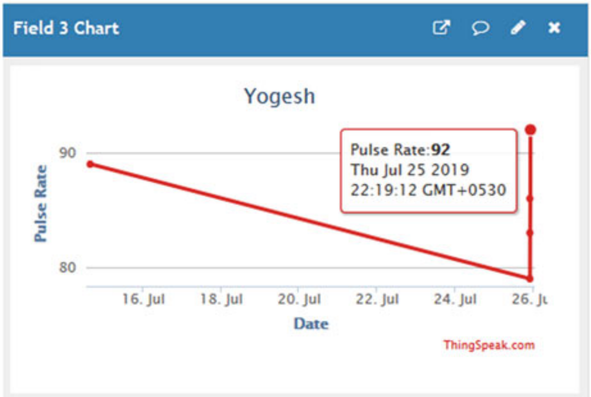


Fig. 5a Graph of pulse rate parameter from ThingSpeak platform

falls [27]. Figures 5a, 5b, and 5c show graphs of pulse rate, systolic and diastolic parameters from ThingSpeak platform.

Tables 1a, 1b, and 1c show the excel sheet of systolic, diastolic and pulse rate readings with date and time, directly from ThingSpeak, and it is very easy to evaluate the clinical data.

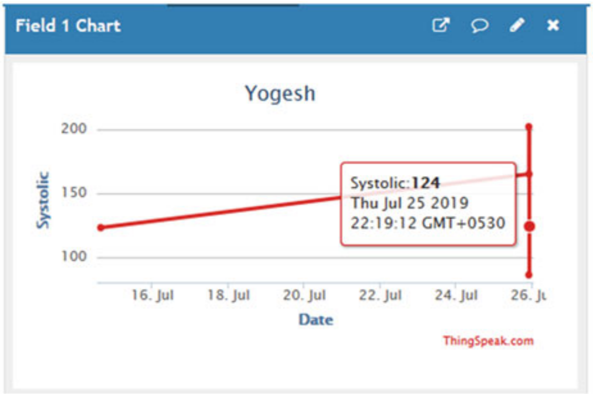


Fig. 5b Graph of systolic parameter from ThingSpeak platform

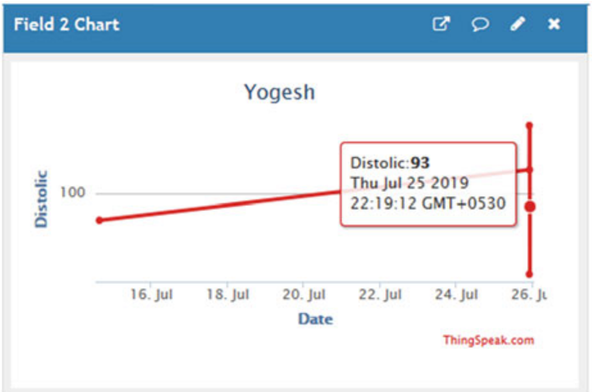


Fig. 5c Graph of diastolic parameter from ThingSpeak platform

Table 1a Excel sheet of systolic parameter from ThingSpeak platform

Sheet for systolic reading		
Date and time	Yogesh	Field1 (Systolic)
2019-07-25 22:19:12 GMT + 0530	1	124

5 Discussion

Often times the transportation charges related to a hospital visit are more than the doctor’s consultation fees, and it is time consuming to personally visit the doctor at hospital. Using Information and Communication Technology (ICT) and the IoT, this proposed system will help to overcome this cost and the geographical distance between rural people and the doctor. Typically, the system easily connects with

Table 1b Excel sheet of diastolic parameter from ThingSpeak platform

Sheet for diastolic reading		
Date and time	Yogesh	Field2 (Diastolic)
2019-07-25 22:19:12 GMT + 0530	1	93

Table 1c Excel sheet of pulse rate parameter from ThingSpeak platform

Sheet for pulse rate reading		
Date and time	Yogesh	Field3 (Pulse Rate)
2019-07-25 22:19:12 GMT + 0530	1	92

digital medical sensors, and the measured parameter readings are made available to medical professionals globally by sending the readings over the cloud server.

6 Conclusion

This chapter presented telehealth monitoring systems with different wireless technologies and related their physiognomies. This chapter also indicated that the two-class method is always better than the one-class method. We designed a low-cost IoT module which mends the classification performance, compared it with the existing Ambulatory Blood Pressure Monitoring system (ABPM) and demonstrated that our proposed system is more useful than the existing one.

The main advantages of this system are it is very cost effective, easily interfaced with BP sensors and body sensors, the measured reading can be easily accessed globally and it has user-friendly operation. The only limitation is the availability of the internet. This system can be easily used for better coordination between doctors and patients not only in urban areas but also in rural areas. It can help to overcome the geographical barriers between rural patients and medical specialist by enabling remote patient monitoring that is totally contactless. The results proposed in this chapter require external validation and clinical trials.

References

1. B.G. Celler, I.E.E.E. Fellow, R.S.S. Mashohor, Important paper home telemonitoring of vital signs – Technical challenges and future direction. IEEE J. Biomed. Health Inform. **19**(1), 82–91 (January 2015)
2. Global Health Workforce Shortage to Reach 12.9 Million in Coming Decades. Accessed 27 March 2017
3. M.M. Alam, E.B. Hamida, Surveying wearable human assistive technology for life and safety critical applications: Standards, challenges and opportunities. Sensors **14**(5), 9153–9209 (March 2014)

4. A. Darkins et al., Care coordination/home telehealth: The systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed. E-Health **14**, 1118–1126 (December 2008)
5. M.M. Baig, H. Gholamhosseini, M.J. Connolly, mHealth monitoring systems for hospitalised older adults – Current issues and challenges, in *(Accepted) Mobile Health (mHealth): The Technology Road Map*, ed. by S. Adibi, (Springer, New York, 2014)
6. F. Lau, C. Kuziemsky, M. Price, J. Gardner, A review on systematic reviews of health information system studies. JAMIA **17**(6), 637–645 (2010)
7. M. Chan, D. Estève, C. Escriba, E. Campo, A review of smart homes – Present state and future challenges. Comput. Methods Prog. Biomed. **91**(1), 55–81 (2008)
8. V.L. Yu, L.C. Madoff, ProMED: An early warning system for emerging diseases. Clin. Infect. Dis. **39**(2), 227–232 (2004)
9. J.R. Turner, A.J. Viera, D. Shimbo, Ambulatory blood pressure monitoring in clinical practice: A review. Am. J. Med. **128**(1), 14–20 (January 2015)
10. Health Informatics – Personal Health Device Communication – Application Profile – Optimized Exchange Protocol, ISO/IEEE 11073–20601 (2014)
11. Continue health alliance design guidelines. http://www.continuaalliance.org/products/design-guidelines
12. M. Clarke, J. de Folter, V. Verma, H. Gokalp, Interoperable end-to-end remote patient monitoring platform based on IEEE 11073 PHD and ZigBee health care profile. IEEE Trans. Biomed. Eng. **65**(5), 1014–1025 (May 2018)
13. A. Alesanco, J. Garcia, Clinical assessment of wireless ECG transmission in real-time cardiac telemonitoring. IEEE Trans. Inf. Technol. Biomed. **14**(5), 1144–1152 (September 2010)
14. J.D. Trigo, F. Chiarugi, A. Alesanco, M. Martınez-Espronceda, C.E. Chronaki, J. Escayola, I. Martınez, J. Garcıa, Standard compliant real-time transmission of ECGs: Harmonization of ISO/IEEE 11073-PHD and SCP-ECG, in *Proceedings of the Annual International Conference of IEEE on Engineering in Medicine and Biology Society (EMBC)*, (2009), pp. 4635–4638
15. Health informatics – Personal health device communication – Part 10406, Device specialization – Basic electrocardiograph (ECG) (1- to 3-lead ECG), in *IEEE Standard*, (2012), pp. 11073–10406
16. S. Thelen, M. Czaplik, P. Meisen, D. Schilberg, S. Jeschke, Using off-the-shelf medical devices for biomedical signal monitoring in a telemedicine system for emergency medical services. IEEE J. Biomed. Health Inform. **19**(1), 117–123 (January 2015)
17. S. Khalid, D.A. Clifton, L. Clifton, L. Tarassenko, Two-class approach to the detection of physiological deterioration in patient vital signs, with clinical label refinement. IEEE Trans. Inf. Technol. Biomed. **16**(6), 1231–1238 (November 2012)
18. J. Quinn, C.K. Williams, N. McIntosh, Factorial switched linear dynamical systems applied to physiological condition monitoring. IEEE Trans. Pattern Anal. Mach. Intell. **31**(9), 1537–1551 (September 2009)
19. D.A. Clifton, S. Hugueny, L. Tarassenko, Novelty detection with multivariate extreme value statistics. J. Signal Process. Syst. **65**, 371–389 (2010)
20. S. Hugueny, D.A. Clifton, L. Tarassenko, Probabilistic patient monitoring with multivariate, multimodal extreme value theory. Commun. Comput. Inf. Sci. **127**, 199–211 (2011)
21. L. Tarassenko, A. Hann, A. Patterson, E. Braithwaite, K. Davidson, V. Barber, D. Young, Multiparameter monitoring for early warning of patient deterioration, in *Proceedings of the 3rd IEE International Seminar on Medical Applications of Signal Processing*, (London, 2005), pp. 71–76
22. L. Tarassenko, D.A. Clifton, M.R. Pinsky, M.T. Hravnak, J.R. Woods, P. Watkinson, Centile-based early warning scores derived from statistical distributions of vital signs. Resuscitation **82**(8), 1013–1018 (2011)
23. C.P. Subbe, M. Kruger, P. Rutherford, L. Gemmel, Validation of a modified early warning score in medical admissions. Q. J. Med. **94**, 521–526 (2001)
24. G. Parati, G. Bilo, G. Mancia, Blood pressure measurement in research and in clinical practice: Recent evidence. Curr. Opin. Nephrol. Hypertens. **13**(3), 343–357 (May 2004). https://doi.org/10.1097/00041552-200405000-00013

25. C.M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006)
26. L. Mhatre, N. Rai, Integration between wireless sensor and cloud, in *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, (Palladam, 2017, 2017), pp. 779–782. https://doi.org/10.1109/I-SMAC.2017.8058285
27. A. Maier, A. Sharp, Y. Vagapov, Comparative analysis and practical implementation of the ESP32 microcontroller module for the internet of things, in *2017 Internet Technologies and Applications (ITA)*, (09 November 2017). https://doi.org/10.1109/ITECHA.2017.8101926

Yogesh S. Kale is currently Assistant Professor in the Electronics and Telecommunication Department. He received his Bachelor of Engineering in Electronics from YCCE, RTMNU, Nagpur and M.Tech. in Communication Engineering from YCCE Autonomous, RTMNU, Nagpur. Presently he is working towards his PhD from RTM Nagpur University in Telemedicine. His area of Interest is Communication and Embedded Systems. He has published more than 4 papers in both international conferences and journals and filed 4 patent applications, 2 of which have been published. He has approximately 11 years of experience in academics.

Dr. Shubhangi V. Rathkanthiwar is currently Professor in the Electronics Engineering Department. She is the Recipient of Best Scientist Award 2017 (RTMNU), Best Teacher Award 2010 (Maharashtra Teacher's council), Best Research paper presentation Award 2011 (IET, UK), Shikshak Ratna Award (2011) * Editorial board member of 7 international journals, Reviewer of 6 IEEE Transactions and one Science Direct journal, TPC Member of many IEEE conferences *and has 5 Books, 3 Book chapters, 92 Research papers, 3 patent applications, 2 copyrights, and 300+ articles to credit.

Big Data

D. Karthika and K. Kalaiselvi

1 Introduction

Anything on the internet nowadays is open. You will check for feedback or opinions on how you buy an issue. Consumers may also be misled by the better dependence on certain studies in their ads. Consequently, a consulting program offers people with the potential to discuss curiosity and acceptability [1]. The attributes of an item, the wishes of customers, and market knowledge are used to create this context. Across these development approaches that capture these data and deliver insightful intelligence, the users' questions, insights, decisions, and behaviors continually generate a massive amount of knowledge. Big data and analysis are not an unusual topic. And then the characteristics remain accurate. Many techniques have been established to productively gather vast quantities of data because of the data processing and implementation of many unstructured and unprocessed data.

The security industry is a prime example of how big data will be included in increasing factors. A treatment programmed, depending on a user's choice, can determine if a person is buying a topic or not. Dependent on the user profile or style, this method is really to incorporate. Its chapter discusses the process for collective public participation screening, offering useful detail on the consumer product profile. Other platforms and networking sites are now accessible on the Internet where users can share opinions, suggestions, reviews, and product prices. For users who do not give ratings to receive user reviews for each object, the recommendation system opts for users. Many e-commerce platforms utilize a recommendation framework to improve sustainable sales. Millions of clients are buying their products from e-commerce web portals [2].

D. Karthika (✉) · K. Kalaiselvi
Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (Formerly Vels University), Chennai, Tamil Nadu, India
e-mail: kalairaghu.scs@velsuniv.ac.in

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_3

They offer their thoughts or review on this commodity in the respective web forum after they have purchased goods. Therefore, both businessmen are predominantly neutral in increasing wages. With this suggestion system process, our sales productivity in this sector can be improved. While consumer preferences can be low-risk, decision-making in some industries may have a larger effect on consumers. Life-impact choices in the health industry should be rendered when they lead to health and safety services. It will also track and dispense patients, manage vital signs, and operate together on a central network in real-time. The Advice Mechanism would also only enhance decision management to prevent danger and pause. The adequacy of clinical counseling services is enhanced by this.

2 Recommendation System and Its Basic Concepts

Under the recommendation framework, two major actors, that is, goods and end consumers, perform an energetic role. Consumers may assess certain preferences about such items and use the data obtained. The data obtained were known as a utility matrix describing the value of the level of priorities of each pair of customer goods. The two major forms of advice are user-based and item-based. Customers have their preferences and scores for items in the user-based recommendation system. You will propose this item in terms of device resemblances to the consumer who has not been defined by this device using a user-centered recommender. The article developed a recommending organization with correlations between things (not within users) to generate user predictions. The first phase of the prediction is for the recommendations method to gather data [3].

Phases of Recommendation System

Information Collection Phase

This method collects key user details and generates user profiles that rely on client role, background, or service. Without a well-defined user profile, the suggestion engine cannot work properly [4]. A suggestion framework is based on the answers that are received from different interactions including explicit feedback, tacit feedback, and data. Users give experienced input combined with their company preferences, while tacit input enhances the curiosity of the customer by the study of consumer behavior. 1. Hybrid feedback is considered a mixture of overt and implicit feedback (Fig. 1).

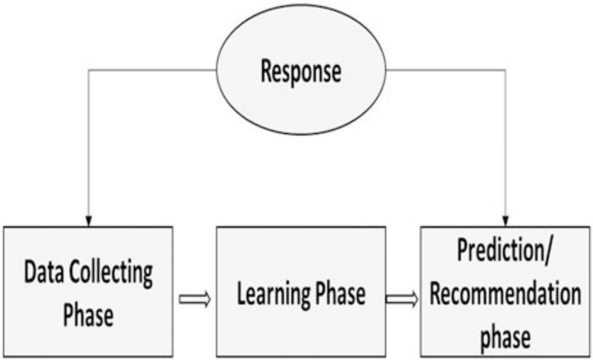


Fig. 1 Recommendation system phases

Step in Learning

This approach uses the input obtained in the previous phase to process and the characteristics of the consumer utilizing a testing algorithm.

Step of Prediction/Recommendation

Preferred items are recommended for users in this phase. The framework forecasts model, memory, or observed consumer behaviors by analyzing the feedback gathered during the knowledge collection process.

Methodism

Techniques for Filtering

End-users must be offered helpful guidance and an appropriate process of clearance. This unit describes threesome procedures primarily used to render product suggestions to customers. Figure 2 demonstrates the hierarchy of recommendation mechanisms focused on multiple filtering technologies [5].

A. Content-Based Filtering Technique

The approach to content filtering relies on a review of the functionality even in making forecasts. Typically, content-based filtering is included in the recommendation. For the filtering method, the consumer information sets out instructions. The contents of the consumer discuss the numerous features and prior sales background of the object. The good, bad, or neutral result offers customers their choice. The framework proposes good quality goods for customers in this methodology.

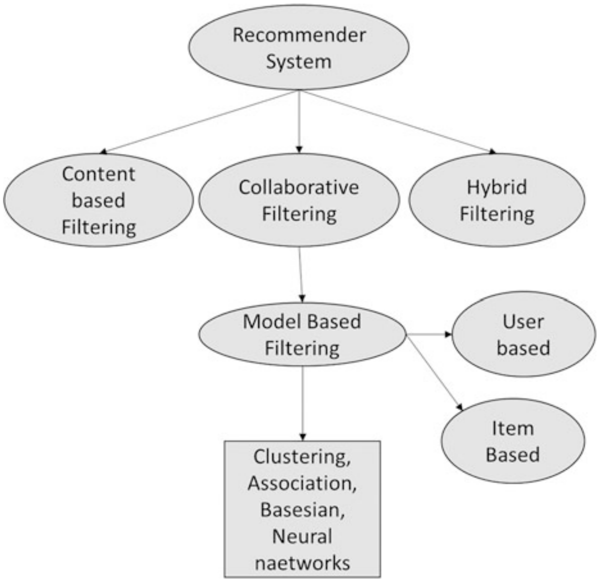


Fig. 2 Filtering hierarchy of a recommendation system

B. Collaborative-Based Filtering Technique

This method utilizes user-based reviews to define correlations between goods instead of explaining the similitudes of product features and attributes. If all customer feedback has been issued, the framework uses a competition index to connect certain outcomes to other customers and provides the highest-ranking of solutions. Different remote markers, including the Jaccard interval, Cosine gap, and Peterson coefficient, are used to define the consumer similarity. The system of screening is generally used to suggest goods based on e-market customer reviews [6].

C. Hybrid-Filtering Technique

This methodology incorporates these two approaches to enhance the recommendation system's efficiency and efficiency. The following techniques should be considered as an alternative to dual filters: to build a recommendation scheme that incorporates double methods: to implement a content-based approach to collective filtering; and to use a content-based approach to filtering. This technique utilizes multiple hybrid systems, including cascades, measuring, combining, and hybrid stages [6–8].

3 Health Recommendation System

Data mining and analytics are quickly growing and the usage is growing in numerous areas. This health sector is one of the exciting fields for the focus and appreciation of Big Data Analytics and its application. These three primary features are the volume, the speed at which data are produced, collected in such a way that it is transmitted, and the richness (the availability of data from several different sources), which characterize large data in health data. Recommendation mechanisms are common for the study of huge databases and capabilities for vast volumes of untreated data and knowledge fatigue.

The health network must be strengthened with additional medical support services (MSS) programmed to cope with knowledge from many people with various issues concurrently. The advice method focuses on predicting and generating customer goods. The predictive analysis may be used for purposes through its framework. A crucial element to be used in the suggestion framework is community review. The clinical recommendations framework is a decision-making tool that provides health professionals and patients as clients with accurate clinical knowledge. The scheme helps individuals to manage illnesses correctly and to eliminate health complications and offers useful knowledge on medication recommendations and good quality procedures of patients of healthcare practitioners. These hours can be reliable and effective to support end consumers with the software in Fig. 3 [9].

The HRS is made up of different steps to say an object. This involves planning, user accounts, nostalgic study, preservation of personal privacy, and guidance. These are the phases of the user profile. First, we need health details to broaden the classification methods compilation. Include profile health records (PHRs) and a consumer data network much of this time. PHR is of considerable significance as an interface for the decision engine in the prediction and evaluation of patients

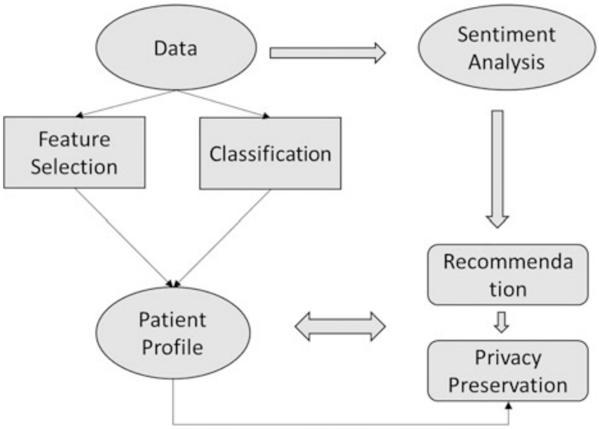


Fig. 3 Health recommendation system (HRS)

on clinical care. The software design associated with PHR incorporates valuable information when choosing services. The classification system is used to analyze and position the information repository. The consolidation mechanism, the learning method, and the appraisal are the three sub-stages of the recommendation process. Following the usage of the consumer database in the advice process, patients and physicians can consider those professional recommendations and increase the quality of health care. The process of HRS empathy unites individuals with a psychological degree of the correct decision. It helps users to learn about their opinions on a topic. The Hour's security mechanism should not change important details. Hours are quite helpful for medical applications in measuring the service provided to society [7, 8].

Designing the Health Recommendation System

Many various methodologies have been built in this increasingly emerging field. Here we have placed in place a fair and logical process. A problem statement is created in the first phase by the product development team to define project priorities. A summary of the meaning of the project accompanies the sentence. A feasibility analysis will be undertaken by the design team, involving technological assessments, cost estimates, and risk forecasts. The team will shift to the next step of the project development process after the definition of the problem has been understood. The team reports on the aspects of the project. The elevated project costs, when contrasted with normal ventures, suggest that the team would execute a realistic economic study to support its cost-effectiveness [10]. The project team should also have background knowledge on the problem and recent studies and interventions in this area. Then in stage 3, the design process continues and begins. The debate regarding the dilemma is several stages. Self-governance and variables or calculation criteria are specified in the same area. Data sources are often identified; data are processed, evaluated, and converted for data processing. Adequate machine instruments like Hadoop and Cloudera need to be obtained at this period. Step 4 includes testing and checking the prototypes and their output. Using automatic feedback loops is a step-by-step method that reduces the chance of deceit.

Framework for HRS

This calls for a framework in which patients, doctors, nurses, and medical workers will be incorporated into a shared consulting system. The software architecture contains three elements: expertise selection, information production, recovery, and show. The next time is to determine the data. Collected details: diagnosis records, data sampling and arrangement of medical instruments, data on illnesses, monitoring and research, hospitals, clinical data, drugs, prescription, CT, and X-rays; (ii) specialist equipment semi-standard data. The figures are straightforward programmers, clinics, labs, students, government departments, and so on. Public reports, governmental papers, hospitals, personal statistics, vital signs, CT scans, fingerprint biometrics, rays and fingerprints, diagnostic codes, etc. [9, 10].

An analysis is given. The results are valid at present. Healthcare automation systems are an integrated information field that uses computational tools and analysis to provide specialists with guidance. Firstly, it defines forms of guidance, as with most other sectors of instruction. Dietary data: Dietary advice production is categorized into numerous categories. The doctor should change food habits such that nurses get adequate medication to prevent their disease. Suggestions can involve healthy eating, new food, aggressive food, or natural additives. Daily practice: people who need help are expected to obey directions and routine to be easier to heal and to reflect on users' wishes. The needs of the consumer might be a place, disease, weather, etc. Diagnosis: medical strategy focused on the symptoms of individual situations. Diagnosis: medication/therapy: advice about specific aspects of infection or recovery. The medication analysis method is the second part of the framework.

During the testing method, health guidelines could be made. This refers to consumers who utilize the area first. The main users of the programming are wellness experts, clinicians, and patients. To help those such as pharmacists, doctors, and physicists, Hours will also benefit these end-users. Those advisory systems may require the cost of medical care. Calls for analytical techniques for MapReduce is used and implemented in many medical monitoring systems. It raises the detection rate and establishes the appropriate standards for doctors to assess the patient's illness type and to evaluate the patient's condition. Visualization is the third aspect of the system. The description of the goods presented in this section is given with elements. Simulation methods and information representation are used to transfer mining expertise to end-users. The healthiest option is preferred, but the evaluation of the treatment depends on thematic criteria. Information-driven techniques are being used to extract information through evidence mining and critical learning from heterogeneous data. Hospital treatment is focused on scientific data, experience, and patterns [11].

Visual knowledge extraction and deep learning may be used for medical database exploration of measurements:

(i) Focus on teaching
(ii) Generation of patient profile
(iii) Analysis of emotion
(iv) Suggestion
(v) Retention of privacy

(i) Training Phase

Physicians conduct clinical experiments on individuals to identify common diseases such as TB, cholera, measles, and so on. Therefore, doctors require the skills to research, assess, and cure numerous diseases utilizing criteria and variables.

However, there is a huge increase in the amount of information produced in healthcare. The compilation and review of data are parts of each process. Nevertheless, the absence of sufficient data collection and computing software will hamper the whole operation. All of this involves the compilation of numerous medical details and statistics, community details, diagnostics, studies, clinical experiments, public records on patients, real-time clinics as well as information on health care services, such that the protocols for collecting real-time data are more effective.

(ii) Patient Profile Generation

A user profile with different details is generated for each at this stage. The claimant may have a medical background report of the patient. Knowledge is included from different sources: clinicians, nurses, medical studies, CT scans, and X-ray imaging. The phase begins at the very beginning when new patients are admitted when data are collected and new health records are generated. The device changes the data to suit the patient's specific requirements.

(iii) Sentiment Analysis

Patients must be truthful in the initiative to preserve the health to the protection of medical records to support patients with healthcare care advice. There is reliable and no abuse of data gathered by patients with or without sufficient medical knowledge.

(iv) Recommendation

Guidelines can be created with user contexts and rules removed. Personalized advice is accessible to patients. These recommendations can include prevention and correction steps, causes, or potential disease detection.

(v) Privacy Preservation

Hours must provide multiple specialist expertise to boost the superiority of guidance in healthcare. It is essential in clinical science to preserve the privacy of the individual's records. The solution suggested would maintain the dignity of this information while protecting personal identities effectively.

Methods to Design HRS

The system contains several strategies that serve a given software parameter's domain interests. The two options are primarily based on customer needs and hence must meet market requirements. The use of participatory architecture is the first approach to the method. The owners' core assets are in threat. Customers engage strongly in the implementation of the technique since consumer input assesses the conceptual system to remove any issues in the new model. Patient evaluations are quite relevant regularly [10–12]. Doctors include success criteria ahead of the advantages and incentives of hiring firms while creating a referral scheme, which may even act as informal guidance to resolve serious health risks. The most overwhelming challenges are that the infrastructure today is structured to help the broad presence of healthcare professionals and practitioners. Such instruments can lead to the diagnosis without active intervention by physicians. The second main advantage of holidays is the usage of privacy.

The integrity of the devices is guaranteed by contractual secrecy. This is used to expose the medical background of the patient which may not disclose the identity of the client. The consumer is also trustworthy. Sometimes, doctors cannot take long-term safety into account. Access to an integrated healthcare system is vital to patients' safety. The degree of understanding of privacy issues is so broad on the internet that technology is often connected to different viewpoints on vulnerabilities and consumer experience. Therefore, consumers are less willing to expose private evidence. A third method is of combining them, both pleasant and productive interactions. Bidirectional dialog (consumer relation and method of recommendation). It makes confidence between doctors to allow consumers to know the symptoms of patients and to alert patients about their diseases. The collecting of data serves the function of the operating system and promotes knowledge of patients, doctors, and their needs [13].

A sound process for evaluating alternatives would allow users to determine and explain comprehensive recommendations in the Recommendation System. In this area, therefore, more study is important as discrimination may play a crucial role in the health sector. Any of the common big data resources in healthcare are accessible from software cleaners, apache Hadoop, and Cassandra repositories. It is a leading instrument of enormous potential in the field of broad numerical results. Hive is one of the eco-components of Hadoop, which enables programmers to build comprehensive Hadoop datasets. Broad collections of data may be updated and reviewed. Data Cleaner is an effective data management device data disclosure programmed platform [14]. This process is usually used to wash, organize, and combine the results. Cassandra is often used for the safe processing of vast volumes of data. These methods are used to interact with the Big Data Analytics recommendation engines.

Evaluation of HRS

To ensure the performance of the recommendation method, the criteria to evaluate the framework should be chosen. Due to the historic assessment of the requirements resulting from information processing, advisory mechanisms were applied. In the estimation, common parameters used are as follows:

- Precision: an approximation of each case received.
- Remember the proportion of products that are not included in the approved list.
- F-Measurement is a precision metric of the evaluation that varies from the weighted vocal mean for precision and the research retrial.

- Receiver operating characteristic (ROC) curve: It indicates that the false-positive rate is truly hopeful. It utilizes the relationship between sensitivity and character.
- Root mean square error (RSME): The standard deviation from residual error determines this equation which is the difference between the limited which expected expectations.

To decide the standard of time according to consumer acceptance and satisfaction, the appraisal criteria for a recommendation framework are required. The Framework can work to avoid problems that lead to better health research because of customer preferences by tailoring software for individual users. Interface complexity analysis includes an HRS uniformity index (UI) control as well as a consumer-defined performance. In the current recommendation method, the pretension of objective metrics and the absence of factors like serendipity and diffusion pose significant challenges. There are unique health issues, but data collection and associated incidents need to be enhanced. Therefore, the two linked shows must be held for hours. Another big concern is trust. The doctor will customize your prescription to restore trust as things get worse. To schedule the HRS and devise a policy according to the specifications, the individual involved is pragmatic [15].

The study underpins the efficiency of the HRS. Tracking improvements in health services, for example, and giving follow-up recommendations regarding recovery patterns. The feasibility of the machine is not measured, since certain patients will feed without checking the system, even though patients are subject to feed restrictions. The health guidance will always promote actions and should always be taken up. The machine will continue to track the patient and see if the drug is working after starting the medicine. To promote rapid rehabilitation, the mechanism must also take steps. It is necessary to accept advice that has no harmful consequences, since failure to take account of one parameter of health may contribute to another condition, which may contribute to bodyweight loss due to diet habits (superficial health parameter). Customer-friendly and efficient applications must be produced before practicing the process. We are trying to guarantee that the device results in real time.

4 Proposed Intelligent-Based HRS

The framework consists of four primary modules: the first component involves the processing and storing of data from diverse sources such as clinics, community facilities, etc. The data databases for all patients contain population, diagnosis, personal history, test, clinical research, and so on. Records are exchanged by easily centralized databases and accessible to staff at the health facility. Each record consists of a single file that can be modified to allow doctors to easily make improvements and increase access. Dual replication of data is not required due to a shortage of redundant archives [14–16].

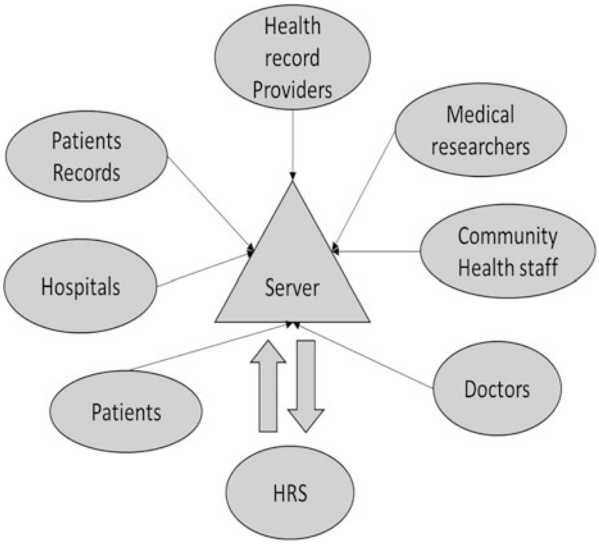


Fig. 4 Health recommendation system (HRS) architecture

The second section of the system involves data preprocessing. The vast volume of data obtained is processed and analyzed in this portion. A range of approaches, including methods for the collection of attributes and methods for transforming data, are used to classify and clean up discarded data. Detailed and redundant data attributes are skipped that do not offer predictive model accuracy. This method is also regarded as data purification. To avoid the creation of unknown or problem models and to enhance the efficiency of the learning model, data cleaning is necessary. The data are often translated into a kind of classification. Therefore, data purification is an effective step to ready the raw data for the next phase (Fig. 4).

The framework consists of four primary modules: the first component involves the processing and storing of data from diverse sources such as clinics, community facilities, etc. The databases for all patients contain population, diagnosis, personal history, test, clinical research, and so on. Records are exchanged by easily centralized databases and accessible to staff at the health facility. Each record consists of a single file that can be modified to allow doctors to easily make improvements and increase access. Dual replication of data is not required due to a shortage of redundant archives [17].

The second section of the system involves data preprocessing. The vast volume of data obtained is processed and analyzed in this portion. A range of approaches, including methods for the collection of attributes and methods for transforming data, are used to classify and clean up discarded data. Detailed and redundant data attributes are skipped that do not offer predictive model accuracy. This method is also regarded as data purification. To avoid the creation of unknown or problem models and to enhance the efficiency of the learning model, data cleaning is

necessary. Data are often translated into a kind of classification. Therefore, data purification is an effective step to ready the raw data for the next phase.

Big Data Analytics and Intelligence Healthcare Perspectives

In comparison, there have been a variety of implications on healthcare: person assessment based on therapeutic imagery; objective physical fitness, to list only a handful. However, for many purposes addressed in the study, healthcare absorbs the benefit of all electronic data repositories worldwide by postponing the implementation of broad-based approaches to knowledge. Secret data will alter the patient's existence or, in significant amounts, change the environment itself. The simplest, most safe, and most useful way to learn about human science is to uninstall this knowledge [18].

In general, large data systems utilize programming algorithms and deep learning to evaluate huge data. This large multi-goal monitoring machine is an emerging big data problem. In healthcare, breadth and multidimensional awareness are rising. To prevent additional costs for medical and safety errors, heterogeneous protection care archives, such as words, images, videos, etc., have request process, viewed, and analyzed. Increasing data growth enables sophisticated computing structures to be rapidly built. "Big data" has several meanings in scientific science. The realistic description of the usage of human health data includes: "massive health knowledge," comprising immense amounts, a range of biological, psychological, environmental, and lifestyle particulars, from persons to vast populations.

A wider concept of Wide Data involves vast datasets that cannot obtain, archive, process, and validate software for standard data processing. "Most patient data ought to be evaluated in a medical sense, to contribute to the advancement of information and disease predictions". This chapter addresses the utilization of high volumes of patient records, the implementation of multimodal information from multiple locations, and rising labor and healthcare costs. A wide-ranging data collection utilizing a particular output of facilities is a significant part of the health sector decision-making process.

Architectural Outline Designed for Big Data Analytics in Health Care

An architectural context is an essential prerequisite for any method of information exploration. The proper use of big data in health care data needs to be analyzed and created [19]. A computer model proposes data acquisition at a high rate of data processing during the research process. Few strong ideas for analysis of large-scale healthcare data analyses. This is a nod to the above. A clear and easy frame is

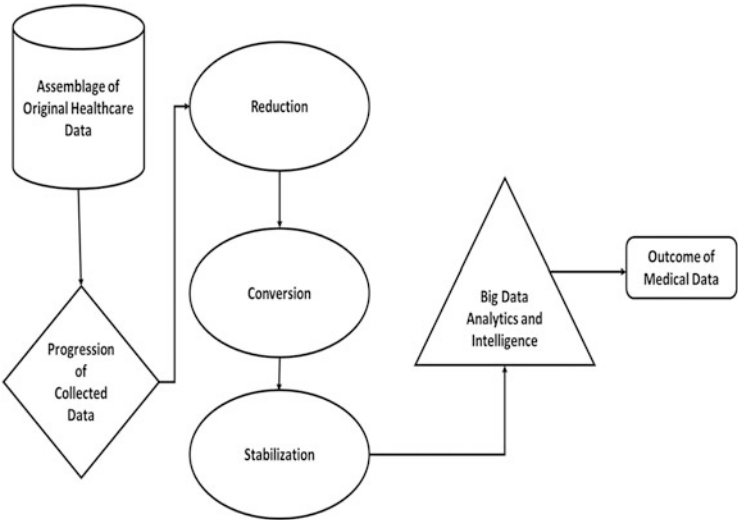


Fig. 5 Big data analytics in healthcare outline

important for efficient study. Second, an architecture system for large-scale data processing of healthcare data is provided at level 0 (Fig. 5). Secondly, in the diagram, an architectural stage 1. 6.

Figure 6 is the highest. International data streams include social networks media networks, machine data such as analysis from multiple sensors, biometric data, and personal details like medical paper reports, medical history, or papers. Those sources of data may be secondary sources of data. These networks create massive amounts of facts every day, resulting in big data growth. Data sources can also be located at various geolocations in various sizes such as ASCII/email, flat files, csv, similar tables, etc. Big medical data are used in various contexts than conventional big data analyses. The peculiar essence of raw medical evidence and the heterogeneous experience of growth is the main explanation for these data. For data cleaning, data standardization, and data transformation another preprocessing phase is needed. To enhance and ultimately enhance medical diagnosis, different technologies for creating large-scale data in the medical field should be applied.

5 Intelligence-Based Health Approval Classification via Big Data Analytics

Big data analytics, like health breakthroughs, would enable up doors. This type of big data analytics will overcome major capacity challenges and greatly boost healthcare system efficiency and sustainability, and take measures toward that. Every imaginable jargon is used in the understanding of big data, technical

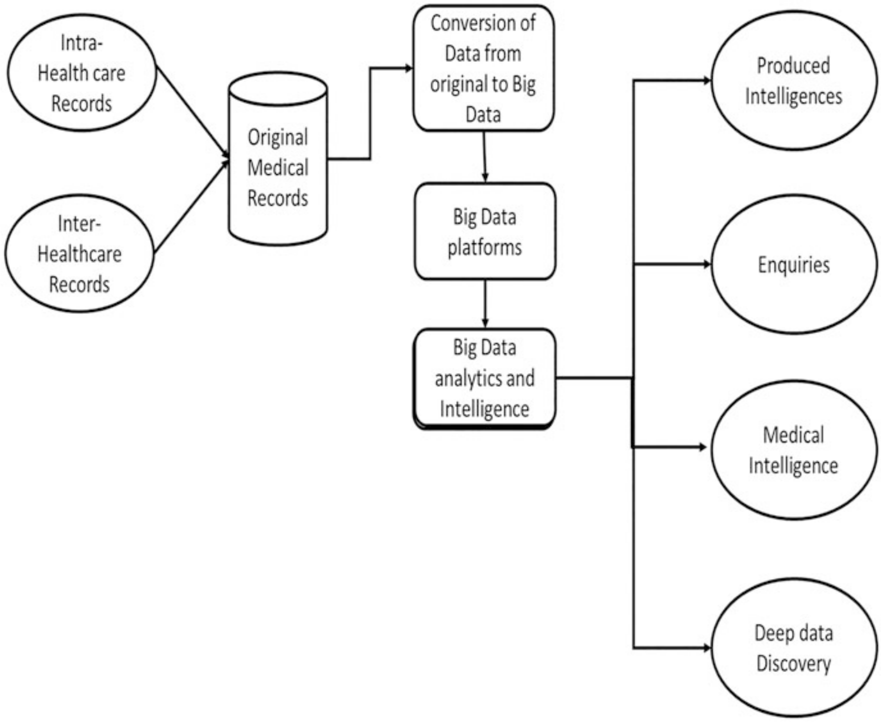


Fig. 6 Big data intelligence in healthcare

information, and the computational system for massive data processing and multiple networks. Health treatment is one of the main fields throughout today's global era. To obtain knowledge and comprehension of the condition, certain health specifics of the care setting can be investigated. The food, fitness, and social behaviors of the patient should be measured, and caring for the client should be carefully estimated. The Protection Recommendation Framework (Model) is a central forum for health care. The decision to render instruments in the healthcare sector has therefore become critical for intelligent health systems [20].

6 Enhancing Workflows in Healthcare

Events are also predicted to arise in the industrial industry. The circumstances in the city, however, are quite diverse and sometimes intertwined, involving workers, offices, workers, and facilities. The dynamic condition renders quality progress in any patient field very complicated for medical professionals and managers without a good explanation of the hospital service. A clinical service provider must therefore be provided with multi-data streams to evaluate the ongoing functionality of the clinic, such as real-time monitoring devices, medical electronic records, personnel, patient monitoring, laboratory data, and machine logs.

7 Healthcare Knowledge Bases

A directive that recognizes the effect of rising demand on healthcare services is of vital significance, along with a dynamic study and a multidisciplinary approach to learning. Problems in health detection and resolution involve the use of vast quantities of data, non-conformities, heterogeneous outlets (heterogeneous dissemination and formats), a reliance on nuanced, multidisciplinary data processing, and rich memantine data trends. Ontology-driven initiatives have introduced health policymakers effectively. Memantine awareness creation has very strong potential and practical health implications. The details can be incorporated from different heterogeneous sources, filtering systems can be built and knowledge is explored. The Limit of Detection (LOD) project has been applied thoroughly over the current centuries and is a traditional method in which structured Web68,69 data are shared and transmitted. LOD has the option to use data for forecast, ranking, diagram, and publishing in a few ways. The realization of this awareness will establish interrelationships and correlations and new hypotheses [21].

Making Better Doctors

The artificial intelligence (AI) dictates what a doctor wants. AI is better and saves patients' life by integrating into a pharmacy. To promote effectiveness or to help its actions, AI supports physicians and health care practitioners in making educated choices and seeking facts, from time to priority.

Optimization

Inadequate systems lose capital and resources. Education, infrastructure, and clinical resources are as good as practicable for healthcare. The IA plan to organize and streamline patient communication should reduce the pressure on doctors to spend more time on matters of urgency. If the patient is facing a medical condition, for example, the AI will examine the patient and plan appointments.

Diagnosing Disease

"Training radiologists cannot be achieved," said Professor Geoffrey Hinton, a professional in the Network, since the Visual Identification Method is complicated and maybe more sophisticated than a person, "health monitoring, identifying type 2 signs of diabetes, the identification of representation environment for translated network analyses (RETNA) and the incidence of coronary diseases, and spot indicators of breast cancer." However, the detection of a condition needs not just clinical evidence. But this does not indicate that all AI features are used. IA devices are also rendered without peer evaluation and methodological rigor. Significant information includes explanations, which involve the framework architecture, the planning and testing of datasets, related data and output measurement, and expectations of neural networks. The randomized control trials (RCT) analysis of medicines identified as AI diagnostic instruments is likely to be performed by AI. Breathing, money, and time by earlier diagnosis and response may be removed from AI detection [22].

Drug Discovery

The production of pharmaceutical medication is an expensive enterprise and one in three experimental products will be put on the market. The perception that medical trials have failed may be harmful: lowering product values, shrinking workplaces, and those workers. Therefore, AI is progressively used by the pharmaceutical and life sciences sectors to accelerate product research and growth. Pharmaceuticals should not threaten to destroy humans. You can see why the failure rate of innovative AI is so huge. AI technologies that can detect novel medicines that are able to increase production, enhance drug growth, and speed up the path to modern pharmaceutical marketing are being developed. A designed solution will save many lives in pandemic situations, including AIDS or Swine Flu.

3D Printing

3D printing technology expects a major change in the healthcare industry, where 3D printing is becoming increasingly inexpensive and available and is questioning conventional paradigms. As its name implies, 3D printing involves the development of three-dimensional structures from a computer model that constructs a container by additive processes. 3D printing, sometimes called conservatory engineering, uses the creation of iterative layers on top of each other. The exact simulation is feasible and mistakes may be minimized. While 3D printing techniques are still in infancy, the potential to be implemented is thrilling. Some AI specialists suggest the potential usage of robotics and bioprinter materials for humans. People will eventually print nearly anything from prosthetics and smartphones to bioengineered body sections and replacements for organs. 3D processing advances are changing patients' lives.

Bioprinting and Tissue Engineering

Liver and cardiac transplants should have long been part of that. 3D bioprinting is the advent of an age where donor lists belong to the past. Instead of utilizing stem cells as a printing media, 3D printed organs may be processed in the same manner as 3D imprinting techniques. In addition to printing cells, bioprinters typically create a gel to cover cells. If published, these organoids may be produced inside people's bodies. A 3D bionic ore printing technology has been developed at Princeton University to track frequencies in other individuals. The ability for repulsion decreases printable, customizable devices from human cells. Another bioprinting process is 3D printing on human tissue. For example, the option of their disfigured bodies is poor for patients with burns and acids [23].

Drug and Facilities

3D printing challenges existing approaches to pharmacology and product design. In the beginning, their efficacy may be tested for drug and other therapies in human replicated cell tissue. 3D printing builds patient-specific drugs. Printed three-dimensional drugs will increase medical effectiveness, mitigate harmful impacts, and encourage adherence to decrease the clinical variability. The same goes for branded products. For instance, 3D home-use strips can be printed by diabetic patients with blood glucose checks. There's a low chance.

Gene Care

Genetic diseases, like places for organ transplants, could have become a thing of the past. The use of nanotechnology between monitoring and genetic modification makes it possible to regulate gene expression at the cell level. For example, gene editing of clinical immune cells is commonly used for the prevention of HIV infection. Gene engineering has the potential to please or remove the disorder of one of 25 children born with a genetic defect. The faulty embryo genes may be changed and the children's desperately required care can be tailored. Patient cell DNA modulation can manage genetic conditions like cystic fibrosis, anemia of the sickle cells, and muscular dystrophy [20–23]. There is no shortage of ethical issues surrounding gene editing: with the analyses and modification of genomes, numerous people are debating the ethical and moral of human embryos. Nanotechnology and genome engineering are now active, far from being solely accessible to wealthier individuals who would exasperate the health and care divide. Nanotechnology can greatly improve the degree of medication and regeneration with the maximum potential for avoiding clinical illness.

The Virtual and Growing Reality

The CT scans would easily enable surgeons to look at patients' bodies and their more real-life conditions; technological students would digitally feel their hearts through virtual reality. Diving in a truly immersive world has been used mostly as a virtual experience in competition. Innovative emerging and hybrid technology guarantees smooth interaction between simulated and physical worlds for both habitats and their control. There are more and more immersive, scalable, and streamlined environments for a variety of medical applications. The way health treatment is given will now be changed because medical care is broader. That is now obvious.

Health Treatment and Delivery

Augmented reality (AR) is included in regular everyday preparation. Improving human skills may be utilized to increase productivity in other forms of technology. At the Alder Leys Kids' Hospital in England, for example, a 360-degree virtual reality (VR) headset is used for teaching children in challenging circumstances. Important judgment recommendations from colleagues and workers may be updated and evaluated. More generally, more practical approaches, such as anatomy and rehabilitation, are used. Real-time enhanced tracking and reviews may be carried out. The first AR procedure was performed at the Royal London Hospital in the world in 2016. You will take part in 3D 360 tourists. The learning environment, particularly in low-income countries, is unique and can interfere with medical education. Doctors utilize interactive and hybrid systems to increase the wellbeing of patients. Doctors perceive vital information in real time through blurred images and lenses. A Tufts Medical Center in Boston uses VR to bring potential nervous patients to therapy from an intervening department of cardiology.

Internet and Classroom Meetings

The costs of growing and mixed reality systems are increasingly making immersive meetings more cost-effective. Better sessions improve the entry burden, save resources and environmental costs, and place emphasis on health staff, when possible. Online meetings are often common with webinars that permit stakeholders to engage without travel. In comparison to one-to-one meetings, the interactive and blended worlds provide an immersive learning atmosphere.

Logging In

A big move forward in-patient treatment is the change from conventional medical reports to EMR. The typical dangers of centralized data collection are reduced by digitalization. Each model also incorporates the medical history of the maker. Blockchain technology is likely to transfigure data access and governing accountability and is common through Bitcoin cryptocurrencies [24]. Blockchain is yet to be found in the area of conventional health care. Blockchain, basically a database set, is a list with many preexisting structures with the signature characteristics with blockchain: an unchanging popular repository that can be exchanged by anybody for legitimacy. As the main financial transaction's authorities, for example, PayPal, Visa, and Mastercard operate as intermediary providers for security. The land register is seen as a confident data repository for property possession. Blockchain technology seeks to address data access, data safety and security challenges, interoperability, and flow of data between physicians, healthcare centers, and insurance providers through the usage of a shared cryptographic ledger. Patients need better access to health records, like IOS 11.3 EHR apps.

Supply Chain Verification

The first step is the supply chain control. Based on the auditable and structured ways it is developed, blockchain can be used to track goods for each component in the stock-bond. For example, a blockchain registry can trace service transactions or processes to identify fraudulent activity, malfunctions, and chain disruptions on data entry or IoT systems. The vendor may be tested, for example, that the cold supply chain insulin in a pharmacy is not chilled properly. In developed nations, it is commonly used to tackle spurious pharmaceutical items. The technology is also being built to enhance the safety of genomic data to overcome the security of privacy issues of massive genomic data [25].

Entry to Medical Record

Patients ought to remain linked to their protection documents. The concern is when the medical proof is distributed to unknown third parties. Third parties are often concerned with safeguarding data security while protecting the dignity of the patient. In the case of a correctly official preserved medical report, cryptographic documentation of the data content will be generated without the intervention of individuals. Transactions can be carried out by providers or customers who have safety data. In the implementation of a private access key, users obtain a signature and a timing. Both databases can be categorized and used via digital signatures to establish a detailed health record for patients. The usage of digital signatures and cryptography technology guarantees secure data transactions and is only accessible to those with access keys. The addition of Blockchain technologies can be tracked, the transaction background can be untouched and audited, and the current iteration of the ledges is stored. Patients should assess any attempts at collecting or processing data [26]. Blockchain's decentralized architecture offers all registered users with a secure connection to the network.

Robot Movement

The use of robots increases wellbeing, but time in the development of developments is quite cost-effective. Machines also completely replace people in medical environments. Technological risks cannot be covered in hospitals and healthcare facilities today; electronic workers cannot delete connections because several would say they can't.

Surgery with Robotic Aid

Robotic surgery enables improved vision, precision, and power. The reach of the robot-assisted operation is now restricted, as access obstacles such as technology and planning costs are increased. Professional medical professionals might need to know how to use the program in the future. The robotic service blurs the fault lines that would hinder adoption.

Drones

Drones can change pharmaceutical flow. Drones can be used for the delivery of drugs, vaccinations, and therapy in countries, war zones, and remote populations. Drones may be used for this purpose. Drones are used in prescription medication quest. Time-sensitive items like blood, body fluids, and organs can travel at less time in or off campuses. Drones for positioning and identifying can also be used, particularly in remote areas. In emergencies, the drone state may be used as a toolbox. For medical therapies, it is essential to establish cure and mortality within minutes after stroke, accident, and heart failure. A heart defibrillator, medications, and two-way radio were used in the Tu Delft ambulance drone unit, which was then delivered to the hospital to boost the response before the first interrogate. Present Limitations are not in use for rising ability, escape flight restrictions, legislation, and technical challenges such as long life and time-consuming hundreds of drones [27].

Intelligent Locations

Smart houses, clinics, places, and stuff are changing our lives. Unbiased networking and an expandable sensor base offer a range of supports for people, patients, and clinicians. Centennials have risen over the past 30 years as the landscape of the industry in the area of health care has improved. The number has increased by 65% in the United Kingdom alone. Smart workplaces and intelligent materials create versatile opportunities for better wellness and care at the same time enhancing medical satisfaction and rising costs per user. Related sites utilize digital sensors that do not need the consumer to constantly communicate and capture large volumes of real-time data instead of linking and managing apps, computers, or machines. Facial recognition, speech recognition, fast alerts, and data recommendations are important [27]. In a constantly evolving world, workflow and management systems streamline and simplify decision-making. These devices offer real-time guidance to patients and healthcare providers: forget your kid.

Hospitals Intelligent

Like a smart home, there is a smart hospital. The smart hospital strives to achieve therapeutic success, a well-structured control of the source, and an impressive, technologically appropriate human viewpoint. Continuous learning programmed from documentation and records to new technology, digital technology, 3D printing, unstructured content, and rigorous study may be utilized by knowledgeable hospitals. Social networking and IA would of course allow patients to become more personally engaged in their medical choices. Internet patient visits with a preferential AI atmosphere that helps the right practitioner to work with the client with appropriate qualifications and experience. Physicians shall track and sustain their involvement and accountability for off-line, physical, and interactive therapy. Growing analytics and large-scale automation can continue to personalize patient interactivity [28].

Automated and streamlined enrollment decreases patients' hospital adherence demands. Patients are classified to be able to track main outcomes during their stay by way of a standardized professional ranking. These operations are sent wirelessly to the device of the medical staff. It is a necessity to recognize certain events or problems. The data-partnering facilities, company generation, issue reduction, research, and decision management progress are translated from facilities and systems into data repositories. As a benefit, patients who encourage cryptocurrencies to sustain a safe and diligent lifestyle are supported with health treatment. Every relevant patient's encounter is anonymized and made accessible to Digital Health Services and internal agencies. The knowledge is stored electronically and is usable [29].

8 Advantages and Disadvantages of the Proposed Health Recommendation System Using Big Data Analytics

The new hours have certain disadvantages. In the event of pulsation, disorientation, etc. irregularities, the in-built sensors enable real-time remote monitoring of critical signals. Big data processing enables clinicians to view patient records easily and cuts the costs of diagnostic research by 50%. Sophisticated cloud-based evaluations of patient knowledge and information systems can offer a highly effective model of operation for all healthcare bases. Physicians are now quicker to enhance findings and illness control. It also changes systems that are stronger and more mature. The proposed health advisory framework still has some limitations. When there are a growing number of users and applications, the collaborative filtering (CF) algorithm would have grave scalability issues. Another big concern is that cold-start complications occur because healthcare resources (HRs) may not have adequate experts and medical knowledge to make the best choices. A synonym is encountered because certain objects have names or entrances that are close or somewhat similar [30].

9 Conclusion and Future Research

Big data analytics is becoming extremely relevant to the healthcare sector. Medical visualization may often depend on medical data either in part or directly to be dubbed manual-diagnostic enhancement so a correct evaluation of a significant condition will involve the ongoing review of most diagnostic data obtained from the numerous clinics of geolocation of patients with identical indications. It is therefore important to select the best apps and resources. During the review process, some other topics must be discussed. While a broad-based health data review is expected to work, challenges must be faced and fixed. These concerns would be discussed in future studies and steps will be taken in a modern system for successful medical big data processing. First research would be performed on a large-scale, raw cleaning and normalization of science data structures. Many medical datasets contribute to the development of complex, fuzzy analysis – an essential part of future research-based techniques to improve this kind of medical image database. Following the preprocessing testing point, a wider dataset, including conventional Big Data Review, would rely on preprocessing measures for more study.

References

1. A.K. Sahoo, S. Mallik, C. Pradhan, B.S.P. Mishra, R.K. Barik, H. Das, Intelligence-based health recommendation system using big data analytics, in *Big Data Analytics for Intelligent Healthcare Management*, (Academic, London, 2019), pp. 227–246
2. A.S.G. Gupta, G.S. Prasad, S.R. Nayak, A new and secure intrusion detecting system for detection of anomalies within the big data, in *Cloud Computing for Geospatial Big Data Analytics*, (Springer, Cham, 2019), pp. 177–190
3. M. Al Ahmad, S.S. Patra, R.K. Barik, Energy-efficient resource scheduling in fog computing using SDN framework, in *Progress in Computing, Analytics and Networking*, (Springer, Singapore, 2020), pp. 567–578
4. A.K. Sahoo, C. Pradhan, H. Das, Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making, in *Nature Inspired Computing for Data Science*, (Springer, Cham, 2020), pp. 201–212
5. H. Das, B. Naik, H.S. Behera, Disease classification using linguistic neuro-fuzzy model, in *Progress in Computing, Analytics and Networking*, (Springer, Singapore, 2020), pp. 45–53
6. C. Rout, A. Panigrahi, J.C. Badjena, C. Pradhan, H. Das, An approximation solution to the NP-complete joint problem in multi-radio WMN, in *Smart Intelligent Computing and Applications*, (Springer, Singapore, 2020), pp. 385–396
7. A.S.G. Gupta, G.S. Prasad, S.R. Nayak, A new and secure intrusion detecting system for detection of anomalies within the big data, in *Cloud Computing for Geospatial Big Data Analytics*, (Springer, Cham, 2019), pp. 177–190
8. A.K. Sahoo, C. Pradhan, H. Das, Performance evaluation of different machine learning methods and deep-learning based convolutional. Nat. Inspire. Comput. Data Sci. **871**, 201 (2019)
9. A. Asokan, J. Anitha, Satellite image enhancement using hybrid denoising method for fusion application, in *Progress in Computing, Analytics and Networking*, (Springer, Singapore, 2020), pp. 115–123
10. K.P. Panigrahi, A.K. Sahoo, H. Das, A CNN approach for corn leaves disease detection to support digital agricultural system, in *2020, 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184)*, (IEEE, Piscataway, 2020, June), pp. 678–683
11. D. Das, A.K. Rath, D.K. Bera, B. Bisoyi, Design thinking on geospatial climate for thermal conditioning: Application of big data through intelligent technology, in *Cloud Computing for Geospatial Big Data Analytics*, (Springer, Cham, 2019), pp. 251–263
12. M.N.Q. Bhuiyan, M. Shamsujjoha, S.H. Ripon, F.H. Proma, F. Khan, Transfer learning and supervised classifier-based prediction model for breast cancer, in *Big Data Analytics for Intelligent Healthcare Management*, (Academic, London, 2019), pp. 59–86
13. A.K. Sahoo, C. Pradhan, H. Das, Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making, in *Nature Inspired Computing for Data Science*, (Springer, Cham, 2020), pp. 201–212
14. N. Dey, A. S. Ashour, H. Kalia, R. T. Goswami, H. Das (eds.), *Histopathological Image Analysis in Medical Decision Making* (IGI Global, Hershey, 2018)
15. M. Rout, A.K. Jena, J.K. Rout, H. Das, Teaching–learning optimization based cascaded low-complexity neural network model for exchange rates forecasting, in *Smart Intelligent Computing and Applications*, (Springer, Singapore, 2020), pp. 635–645
16. H. Das, B. Naik, H.S. Behera, Medical disease analysis using neuro-fuzzy with feature extraction model for classification. Inf. Med. Unlocked **18**, 100288 (2020)
17. S.K. Meher, Efficient pattern classification model with neuro-fuzzy networks. Soft. Comput. **21**(12), 3317–3334 (2017)
18. N. Dey, H. Das, B. Naik, H. S. Behera (eds.), *Big Data Analytics for Intelligent Healthcare Management* (Academic, London, 2019)

19. S.K. Meher, S.K. Behera, E.R. Rene, H.S. Park, Comparative analysis on the application of neuro-fuzzy models for complex engineered systems: Case study from a landfill and a boiler. Expert. Syst. **34**(6), e12215 (2017)
20. J. Tanha, H. Salarabadi, M. Aznab, A. Farahi, M. Zoberi, Relationship among prognostic indices of breast cancer using classification techniques. Inf. Med. Unlocked **18**, 100265 (2020)
21. A.K. Verma, S. Pal, S. Kumar, Comparison of skin disease prediction by feature selection using ensemble data mining techniques. Inf. Med. Unlocked **16**, 100202 (2019)
22. J. Nayak, B. Naik, A.K. Jena, R.K. Barik, H. Das, Nature inspired optimizations in cloud computing: Applications and challenges, in *Cloud Computing for Optimization: Foundations, Applications, and Challenges*, (Springer, Cham, 2018), pp. 1–26
23. L. Khairunnahar, M.A. Hasib, R.H.B. Rezanur, M.R. Islam, M.K. Hosain, Classification of malignant and benign tissue with logistic regression. Inf. Med. Unlocked **16**, 100189 (2019)
24. H. Das, A.K. Jena, J. Nayak, B. Naik, H.S. Behera, A novel PSO based back propagation learning-MLP (PSO-BP-MLP) for classification, in *Computational Intelligence in Data Mining-Volume 2*, (Springer, New Delhi, 2015), pp. 461–471
25. M. Nilashi, H. Ahmadi, L. Shahmoradi, O. Ibrahim, E. Akbari, A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique. J. Infect. Public Health **12**(1), 13–20 (2019)
26. A.K. Sahoo, C. Pradhan, H. Das, Performance evaluation of different machine learning methods and deep-learning based convolutional. Nat. Inspire. Comput. Data Sci. **871**, 201 (2019)
27. H. Das, B. Naik, H.S. Behera, Disease classification using linguistic neuro-fuzzy model, in *Progress in Computing, Analytics and Networking*, (Springer, Singapore, 2020), pp. 45–53
28. S. Selvam, B. Selvam, J. Naveen, Root-cause analysis using ensemble model for intelligent decision-making, in *Machine Learning for Intelligent Decision Science*, (Springer, Singapore, 2020), pp. 93–114
29. H. Das, B. Naik, H.S. Behera, S. Jaiswal, P. Mahato, M. Rout, Biomedical data analysis using neuro-fuzzy model with post-feature reduction. J. King Saud Univ. Comput. Inf. Sci. (2020). https://doi.org/10.1016/j.jksuci.2020.01.007
30. P. Roy, R. Chakrabortty, I. Chowdhuri, S. Malik, B. Das, S.C. Pal, Development of different machine learning ensemble classifier for gully erosion susceptibility in Gandheswari Watershed of West Bengal, India, in *Machine Learning for Intelligent Decision Science*, (Springer, Singapore, 2020), pp. 1–26

Implication of Statistical Methods on Patient Data: An Approach for Cancer Survivability Prediction

Sudip Mondal, Namrata Bhattacharya, Troyee Das, Zhumur Ghosh, and Sunirmal Khatua

1 Introduction

In recent years, cancer has become the second leading cause of death across the globe. Cancer researchers worldwide have made major advances to investigate in depth regarding the mode of diagnosis, treatment, and patient survival rate. Cancer is the unrestrained proliferation of abnormal human cells that develops when the normal cellular control mechanism behaves differently. The primary cause of cancer is DNA damage within the cells mainly attributed to environmental risk factors such as smoking, radiation, viruses, toxic chemicals (carcinogens), etc. A multitude of cancer type exists ranging from common and frequently occurring breast, lung, liver, colon, prostate, and skin cancer, malignant brain tumor, etc. to lesser common and rare ones like leukemia, lymphoma, osteosarcoma, and thyroid and testicular cancer to name a few. The proportion of male-to-female cancer ratio varies widely based on geographical and socioeconomic differences in developing countries like India and other developed countries like the USA. Statistical studies reported by the National Cancer Registry Program in India point out that the probability of developing cancer between the age group of 35 and 64 and life risk in males is 4.67% and 9.05% respectively, whereas in females, it is 6.55% and 10.2%, respectively

S. Mondal · S. Khatua (✉)
Department of Computer Science and Engineering, University of Calcutta, Kolkata, India
e-mail: smcse_rs@caluniv.ac.in; skhatuacomp@caluniv.ac.in

N. Bhattacharya
Department of Computer Science and Engineering, Indraprastha Institute of Information Technology, Delhi, India
e-mail: namratab@iiitd.ac.in

T. Das · Z. Ghosh
Division of Bioinformatics, Bose Institute, Kolkata, India
e-mail: troyee@jcbose.ac.in; zhumur@jcbose.ac.in

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_4

[1, 2]. In the USA, approximately 1,806,590 new cases and 606,520 deaths related to cancers, based on sex and cancer type, have been reported by the American Cancer Society in the year 2020. Estimated numbers of new cancer cases and death for males are 49.46% and 52.95%, respectively, whereas in females, they are 50.53% and 47.04%, respectively. Cancers are more prevalent in females due to the high risk of development of breast cancer and gynecological cancers [3]. In many cases, the initial stages of disease progression are often asymptomatic, and diagnosis at a later stage often limits the scope to restrain its uncontrolled growth, hence reducing chances of patient survival. However, with the advancement in patient screening and cancer treatment strategies, survival rates are improving for various types of cancer. The overall survival rate for all cancers diagnosed in the USA during 2009–2015 was 67%. Survival percentage is highest for prostate cancer, skin cancer, and breast cancer which are approximately 98%, 92%, and 90%, respectively, and lowest for esophagus cancer, lung cancer, liver cancer, and pancreatic cancer which are approximately 20%, 19%, 18%, and 9%, respectively [3].

In the year 2016, the Public Health Foundation of India (PHFI) and Indian Council of Medical Research (ICMR) reported approximately 59 to 60 persons per lakh died due to cancer in all the Indian states. The number of cancer cases has increased double in India between 1990 and 2016. According to the report from regional hospitals, almost 75–80% of patients have different types of advanced cancers among the South Indians at the time of diagnosis [4]. Cancer has greater prevalence due to its late diagnosis at an advanced stage (stage III or stage IV)[5]. The most prevalent cancers are stomach cancer, breast cancer, lung cancer, lip and oral cavity cancer, nasopharynx cancer, colon cancer and rectal cancer, blood cancer, cervical cancer, esophageal cancer, brain cancer, and nervous system cancers. Among various types of cancers, breast cancer is the major cause of morbidity and mortality among women worldwide with rare occurrence in male also [6]. Age groups $\geq$ 60 years are at higher risk of developing cancer in contrast to age groups 30–39 years. Other than genetic inheritance, overweight, low physical activity, and sedentary lifestyles have given rise to several breast cancer cases [4]. However, in India, tobacco usage is the major cause of cancer deaths followed by breast cancer. According to a report from the Indian Council of Medical Research (ICMR) in 2018, cancer of the lip and oral cavity is the leading cancer type followed by breast cancer [7]. The risk of gastric cancer increased with the number and duration of tobacco usage among male cohorts in South India.

According to the World Cancer Report released by WHO, India had an estimated 1.16 million new cancer cases in 2018 [8]. One out of ten Indians will develop cancer during their lifetime, and among them, 1 out of 15 will die. Estimated 7 lakh new cancer cases are added every year with 3 lakh cancer death cases. According to WHO, by 2026, cancer burden in India is estimated to be about 14 lakhs given the expected increase in life expectancy [9]. Data regarding comprehensive childhood cancer remains limited compared to global standards due to significant barriers in recognition, diagnosis, and cure. According to population-based cancer registry programs, there is lower prevalence of childhood cancer especially in leukemia and central nervous system (CNS) tumors due to skewness in data [10].

In the current era, genomic data plays an important role in the treatment of cancer patients. Physicians are provided with vast amount of genomic data generated by next-generation sequencing platforms [11]. The genomic and biomedical researchers play an important role in analyzing such patient data to apply statistical and machine learning methods. Among them, survival analysis is a widely used statistical method for data analysis, now being heavily used in the field of healthcare to estimate the risk of disease progression in patients [12]. It calculates the survival rate as well as the treatment failure rates of cancer patients and predicts the impact of medications and treatment algorithms and the timing of medications and procedures. Therefore, this technique is utilized to develop a prediction model to study cancer progression which aids in cancer treatment [13]. There are numerous types of models for survival analysis. Survival function, life tables, Kaplan-Meier curves, and hazard function models illustrate the survival time of members within a group. To compare the survival times of two or more groups, a model such as log-rank test is used, which is a special case of Cox proportional hazard analysis. Cox proportional hazard regression, parametric survival models, survival trees, and survival random forest models are used to describe the effect of categorical or quantitative variables on survival. For categorical predictor variables, Kaplan-Meier curves and log-rank tests are fruitful, and for quantitative predictors, Cox proportional hazard regression analysis methods can be used [14, 15].

Currently, a fundamental challenge is the accurate prediction of clinical outcomes from these high-throughput profiles. The Cancer Genome Atlas (TCGA) database is used to build deep survival models for analysis of high-dimensional genomic, clinical, and molecular data. Clinical statistical data usually reveal patient age, cancer stage, and different time points like start of treatment, tumor recurrence, patient demise, or leaving the survey [16]. From genomic and molecular data, people have utilized gene expression pattern, DNA methylation, mutation, copy number variation, and small as well as long non-coding RNA (lncRNA) expression data for predicting survival probability of cancer individuals [17]. Determining the key features from these huge genomic data obtained from clinical follow-up by employing deep learning methods can be advantageous for survivability prediction. Existing tools such as SurvivalNet explore these deep learning methods for survival modelling [18]. Due to extensive heterogeneity and complexity of cancer, it is still a challenge to predict survivability of cancer patients more precisely.

In this chapter, we shall discuss the basic concepts of survival data analysis mining human genomic data. Particularly, we have considered three genes, namely, breast cancer 1 (BRCA1), breast cancer 2 (BRCA2), and ataxia telangiectasia mutated (ATM), which are known for playing vital roles in both breast and ovarian cancers [19]. In such context, we will describe the popular statistical methods, such as the genetic algorithm and Cox proportional hazard regression model, and their strengths and limitations toward analyzing such data and put forward its future perspectives. We will also emphasize on some popular survival analysis models such as Kaplan-Meier plots, log-rank tests, and Cox regression. Finally, we shall discuss the usage of these methods toward analyzing cancer datasets which can be accessed from online repositories including TCGA.

2 Cancer Statistics

Here, we will discuss the statistics and recovery rate of two most popular female cancers, breast and ovarian cancer.

Breast Cancer

Breast cancer has the highest frequency of occurrence among all cancers and is the leading cause of cancer-related death worldwide among women. Breast cancer incidences have increased in all urban registries in the world over the last 20 years between the age group 35 and 65 [20]. In the year 2020, approximately 42,690 breast cancer-related deaths are estimated [21]. The American Cancer Society formulated a 5-year survival rate to estimate the proportion of people that will survive at least 5 years after detection of the cancer [22]. According to the report, the average 5-year survival rate for women with invasive breast cancer is 91%. If the cancer is located only in the breast, then the 5-year survival rate of women is 99%. 62% of women are diagnosed at the early stage prior to metastasis. Metastatic breast cancer is the major cause of mortality in women. The 5-year survival rate for metastatic breast cancer is below 25% [23–25]. This is a major challenge for breast cancer management. The same study shows that the average 10-year survival rate is 84% for women with invasive breast cancer. Early diagnosis of breast cancer in adolescent and young adult females between the age group of 15 and 39 is only 47% compared to women who are older than 65 which is only 68%, due to delays in screening in younger women. Survival rates are approximately 9% higher in white women compared to black women.

Prognosis of breast cancer can be performed by doctors using computer programs. Patient-specific information about breast cancer and data from large research studies can be integrated by programs to estimate prognosis. Survival benefit from treatments such as chemotherapy can also be estimated. These programs aid the treatment of breast cancer. Genomic assays (also called gene expression profiling or gene assays) can be used to analyze groups of genes expressed in breast cancer [26]. It can equip us about the severity of the diseased state owing to the aberrant expression patterns of certain prognostic genes [27].

Ovarian Cancer

Ovarian cancer is the sixth most common gynecologic malignancy in women worldwide and third leading cancer in India [28] with a highly aggressive natural history and causing over 140,000 deaths every year [29]. It is caused due to rapid advancement of cell division within the reproductive glands in which the ova and the female sex hormones are made. Ovarian cancer is common in female between the age group of 40 and 50 years old; however, malignancy is more common in postmenopausal women between the age group of 51 and 60 years [30]. Worldwide an estimated 18.1 million new ovarian cancer cases and 9.6 million ovarian cancer deaths were reported in the year 2018 [22]. India itself has recorded around 5.8 lakh new cases of ovarian cancer in 2018. While the survival rates of women with early-stage ovarian cancer are high, most cases are diagnosed late, when the likelihood of successful therapy is low [31, 32]. It is estimated that the incidence of ovarian cancers will increase to 55% per year by 2035 and mortality will increase by 67% [33]. The 5-year survival rate is around 25.4% for ovarian cancer [34]. Ovarian cancer requires improved methods for its early detection since it has a poor survival rate due to late diagnosis.

3 Basics of Survival Analysis

Survival data require a different analytical approach for predicting the survivability of patients. In this analysis, comparison of a time-specific event, e.g., death, disease, or recurrence, with respect to two or more groups of patients is of interest. In this section, we will introduce a few terminologies which are frequently used in this type of analysis.

Censoring

Censoring refers to incomplete data about the "time to event" which is a missing data problem [35]. The time to event is not observed for reasons such as the individual doesn't experience the event when the study is completed, the individual doesn't follow up during the study period, or the withdrawal of the individual from the study. This is common in survival analysis [36]. There are three types of censoring, namely, right censoring, left censoring, and interval censoring. Right censoring occurs when the information about the time to event is incomplete because the subject did not have an event during the time when the subject was studied and the actual failure time exceeds the observation time [37]. For example, the date of birth of the subject is known, and the subject is still alive when the study ends or withdraws from the study. The patient is censored since the exact survival time of the subject is not known. Hence, the observed survival time is always less than or equal to true survival time. Left censoring occurs if the event has already happened before the subject is included in the study; however, it is not known when the event occurred. For example, during estimating the emergence distribution of tooth in subjects, the tooth has already emerged in the subject prior to the start of a dental study. In this case, we don't know the exact time when the actual tooth emerged. We only know that teeth have already emerged when the study begins. Hence, the observed

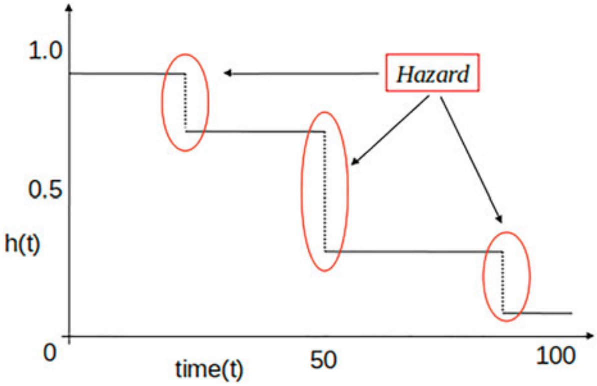


Fig. 1 Hazard function representing the probability of event

survival time is always greater than or equal to the subject's true survival time in left censoring [35]. Interval censoring gives the upper and lower bound on failure time. In this scenario, we know the time interval of the event. Right censoring and left censoring are special cases of interval censoring.

Hazard Function

The hazard function, also known as the failure rate or hazard rate, is used to define the instantaneous risk of experiencing the event of interest, within a very narrow time frame [38]. It is used to model a patient's chance of death as a function of their age.

Hazard function is calculated by dividing the number of subjects experiencing an event in the interval beginning at time t by the product of the number of subjects surviving at time t and interval width [36]. Mathematically, the hazard function h(t) is defined as $h(t) = P(t)/1 - D(t)$, where D(t) is the distribution function of the time to failure of a random variable t and P(t) is the probability density function of survival time t [39]. Figure 1 represents the hazard function h(t), showing the probability of event (death) at any particular time instant. Here X-axis represents the time, i.e., age in years from 0 to 100, and Y-axis represents probability of event ranging between 0 and 1.

Linear Regression and Its Limitations

Regression analysis is a predictive modelling technique that studies the relationship between a dependent (target) variable and independent (predictor) variables [40].

It is used for forecasting, time series modelling, and finding the causal effect relationship between the variables [41]. Linear regression is a statistical data analysis method used to determine the extent of linear relationship between one or more predictor variable(s) and one continuous target variable. Linear regression is mainly used for predictive analysis, e.g., it can be used to quantify the relative impacts of the predictor variables like age and gender on the outcome variable like height. The logistic regression is also a predictive analysis which is used to describe data and relationship between one categorical dependent binary variable and one or more nominal independent variables. The dependent variable in logistic regression is a binary variable: 1 (yes) or 0 (no). In linear regression, the relationship between dependent variable and independent variable must be linear. In logistic regression, it is not required to have the linear relationship between the dependent and independent variable.

Linear regression works only if the dependent variable is continuous and not categorical (e.g., gender= male or female) or dichotomous(0/1, yes/no, etc.). Linear regression modelling also fails to handle censoring or time-dependent covariates. That is where the logistic regression model comes into play which can well handle categorical variables. Moreover, adjusting for more than one variable while building a logistic regression model, in other words, opting for multivariate analysis over univariate regression, can be a better predictor of survival functions between two groups.

4 Popular Survival Analysis Method

Kaplan-Meier

The Kaplan-Meier estimator is a non-parametric statistic which is used for estimating the survival function using lifetime data. A Kaplan-Meier curve estimates the survival probability at each point in time [42]. Survival function is estimated using Kaplan-Meier estimator. It is used to measure the fraction of patients who stayed alive for a definite amount of time from lifetime. This method is also known as the product limit estimator [43]. It gives the average view of the population using the survival curve, which is a graph for visualization of the population at risk [7]. It unearths the possibility of surviving in a given time duration, considering small time intervals. Kaplan-Meier analyzes the time-to-event data using statistical methods. Survival time is the time starting from a default point to the occurrence of the event of interest [15]. Time to event is used to estimate the survival time, such as onset of illness. The survival time can be estimated more accurately if the examination happens frequently, that is, if the time gap between examinations is very little.

This method can be used by the researchers to analyze the participants who developed the disease, survived the disease, or dropped out of the study. This method can be used to compare two groups of subjects such as control and treatment groups [44].

The survival function estimator S(t) is given by:

$$\widehat{S(t)} = \prod_{i:t_i<t} \left(1 - \frac{e_i}{p_i}\right) \tag{1}$$

where t_i gives the estimated time when at least of the events happened, p_i gives the alive patients up to time t_i, and e_i gives the number of deaths that has occurred at time t_i.

Log-Rank Test

A direct comparison of the Kaplan-Meier curves for two or more groups can be made using the log-rank test which is similar to one-way ANOVA for survival analysis [45]. It is advantageous since it allows easy comparison with few assumptions and it is the only test needed for various settings. However, the limitation of the log-rank test is it fails to extend easily to complex settings such as continuous covariates and interactions [46].

The null hypothesis [47] states that there is no statistical significance between the populations in the probability of an event at any time point. Log-rank test is used to test this hypothesis. It is similar to the Kaplan-Meier estimator which is used to compare two or more groups of subject, e.g., treatment versus control group, in a randomized trial. Log-rank test is based on the times of events. For each time, the observed number of deaths in each group is calculated such that there is no difference between the groups. The time is divided into small time periods like days, and comparison is done between the number of actual events occurring in each time period [48]. It is effective in detecting a higher cured proportion of patients in one group than the other group. However, it is not effective if there is no significant difference between specified populations when the survival is prolonged in one group compared to another. A major drawback of log-rank test is that it evaluates the effect of only a single variable at the time of prognosis. For multiple variable analysis, methods such as the Cox proportional hazard model are used [49].

The log-rank test statistic is given by:

$$\frac{(ob_1 - e_1)^2}{e_1} + \frac{(ob_2 - e_2)^2}{e_2} \tag{2}$$

where the expected number of events for log-rank test in each group is represented as e_1 and e_2 and the total number of events observed in each group is represented as ob_1 and ob_2.

Cox Regression

Cox proportional hazard model is a statistical semi-parametric model. It is not entirely parametric or non-parametric. It is an important and flexible model for survival analysis [50]. Cox proportional hazard model is a regression model commonly used for estimating survival time (time-to-event) outcomes on one or more variables in medical research for investigating the association between the patient's survival time and covariates [51]. Cox regression allows multivariate investigation, i.e., the effect of several variables upon the time interval of a specified event. It is used for survival analysis in the context of death from a particular disease. It assumes that the effects of the predictor variables upon survival are constant over time and are additive in one scale. The hazard rate allows us to examine how specific factors influence the rate of a particular event happening at a particular time point.

Cox models work efficiently for datasets including censored data from observations where the event didn't occur, as well as data from observations where the event occurred. It can handle quantitative and categorical predictor variables. The advantage of using Cox regression over other methods such as Kaplan-Meier analysis (which can only handle one variable) is that it can analyze multiple risk factors for survival. Hence, it addresses the issue of patients' heterogeneity caused by different participants [52].

The important assumptions for the use of this method are that the survival times between different patients are independent in the sample, predictors and hazard have multiplicative relation, and hazard ratio is constant over time. The Cox proportional hazard regression model can be expressed as:

$$h(t) = h_0(t)\, exp(b_1X_1 + b_2X_2 + \ldots + b_pX_p) \tag{3}$$

where $h_0(t)$ represents the baseline hazard when all the independent variables (X_1, X_2,...,X_p) are equal to zero and h(t) is the predicted hazard at any instance t which is the product of $h_0(t)$ and the exponential function of independent variables and coefficients $b_1, b_2, ..., b_p$.

Random Forest Model

Different regression models are used in survival analysis to predict the risk of future events. Random forest approach provides an alternative way to build a risk prediction model [53]; thus, it has been extended for survival analysis. Current risk prediction models are mostly based on regression approaches or machine learning algorithms that are static [54].

Random forest is a dynamic decision tree-based non-parametric classification algorithm that can be used to build risk prediction models [55]. The decision trees grow identically based on the bootstrap aggregation samples, and majority vote for the overall prediction. It can handle multiple variables, their complex interactions, and time-varying values from large, heterogeneous datasets to support improved clinical decision-making for individual-level risk prediction [53, 56]. Random survival forests are recently used for the study of right-censored survival data. Predictor variables at each decision tree node are randomly sampled to improve the precision of the ensemble predictions by decreasing the correlation among the trees in the forest [53].

The random forest for risk prediction is formed with decision trees designed to partition the data according to the individual treatment effect. To compute the estimation of the individual treatment effect, the forest provides a set of similar subjects for a new subject. Random forest decision trees identify the importance of the response variable of overall survival time by analyzing the genes using an unbiased approach, instead of analyzing the genes with cancer associations. The individual treatment effect with censored data can be estimated from random forest data [57].

5 Result and Discussion

We have performed here basic survival analysis using Kaplan-Meier and Cox PH estimator for both breast and ovarian cancers. From both the data, we have identified the key difference between single and multivariate regression models and how it is related to expression value of different genes. We have also examined the results of the Cox regression estimator using the log-rank method.

Data Preparation

For our analysis, we have used TCGA data for breast and ovarian cancer downloaded from *The Human Protein Atlas* [58] along with fragments per kilobase per million reads mapped (FPKM) value. FPKM is a normalized gene length-based estimation of gene expression from paired-end RNA-seq data. It is calculated as the number of reads mapped to a particular gene sequence divided by gene length in bp and total number of mapped reads from the given library [59].

We have 946 breast cancer samples and 353 number of ovarian cancer samples. For breast cancer, we have used the following parameters to represent the attributes in our datasets. We have represented status as alive=1 and dead=2; ethnicity as white=0, black or African American=1, and Asian=2; and stage as stage I=0, stage Ia=0, stageIb=0, stageIIa=1, stageIIb=1, stageIIIa=2, stageIIIb=2, stageIIIc=2, and stageIV=3. The average FPKM value for BRCA1 is 2.48, BRCA2 is 1.03, and

ATM is 2.43. Genes have been represented with low FPKM=1 and high FPKM=2 for both cancers. For ovarian cancer, all the attributes are the same except stage which is not available in the ovarian cancer dataset. Here the average FPKM for BRCA1 is 1.226, BRCA2 is 0.069, and ATM is 1.447.

Survival Analysis Using Kaplan-Meier (KM)

In this section, we have analyzed the data from breast and ovarian cancer using Kaplan-Meier analysis method. We have analyzed both datasets corresponding to three major variables, i.e., stage, ethnicity, and FPKM.

Stage: Stage is an important feature for breast cancer survival analysis. We have considered the stage attribute only for the breast cancer data. In Fig. 2, we have considered the breast cancer dataset for analysis based on cancer stages I to IV using KM method. The graph is a representation of the cancer stage vs number of days the patient was alive. Here, X-axis represents the duration, and Y-axis represents the probability of survival using Kaplan-Meier method. We have observed that a group of patients over 50 years of age having stage IV cancer have the lowest survival rate, i.e., 0%, whereas those having stage I have more than 40% survival rate, and stage II and stage III have more than 20% and 40%, respectively. We have also observed that the survival days are doubled in stage I as compared to that in stage IV. The survival time is almost the same in stage II and III, i.e., greater than 7000 days.

Ethnicity: Figures 3 and 4 represent the KM plot for breast cancer and ovarian cancer, respectively, based on ethnicity categorized into white, Asian, and black. Ethnicity is used to characterize cultural expression and identification of distinct populations. In input dataset, the number of available samples for white ethnicity is

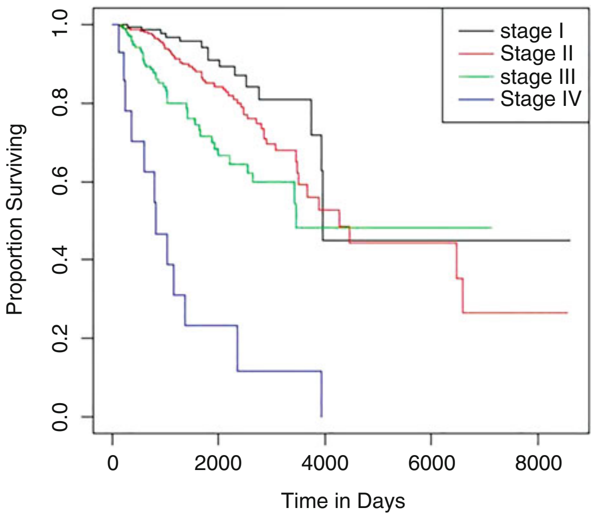


Fig. 2 KM plot for stage in breast cancer

Fig. 3 Ethnicity in breast cancer

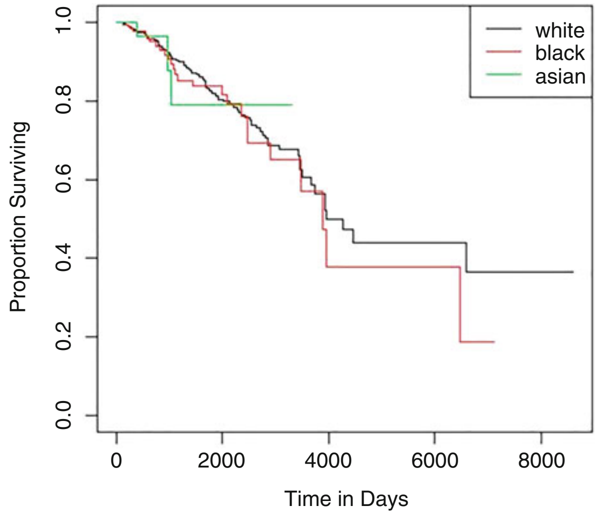
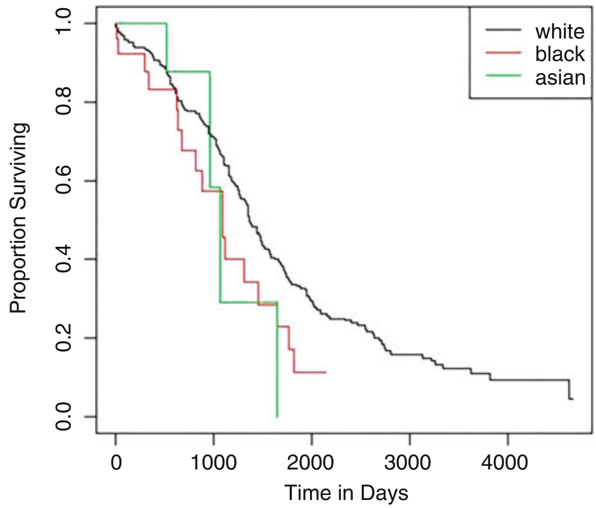


Fig. 4 Ethnicity in ovarian cancer



higher than that of Black or African American ethnicity, and very few Asian samples are available in both breast and ovarian cancer datasets.

We observe from the group of samples that white population in both breast and ovarian cancers have high survival time compared to other colors. In breast cancer, the survival rate of white is almost 40%, whereas in ovarian cancer, it is near to 0%. In the case of black population, the survival rate is 20% for breast cancer and 10% for ovarian cancer, but survival time is 2000 and 7000 days, respectively. Since Asian sample data is less, we observe a survival rate of 80% and survival time approximately 4000 days for breast cancer, and in case of ovarian cancer, survival rate is 0%, and survival time is approximately 2000 days.

FPKM: In the survivability analysis presented here, we have considered three important genes, i.e., ATM, BRCA1, and BRCA2, for both breast and ovarian cancers. For both cancers, we have three different FPKM values for these genes depending on time of study.

For **breast cancer**, the Figs. 5, 6, and 7 represent the calculated survival rate according to the FPKM values of the selected genes. For our analysis, we consider the mean FPKM value for each gene and divide the corresponding FPKM values into high FPKM and low FPKM based on mean FPKM value. For example, if the mean FPKM is 50, then genes having FPKM less than 50 are categorized as low FPKM, and genes having FPKM greater than or equal to 50 are categorized as high FPKM.

Fig. 5 KM plot for gene BRCA1

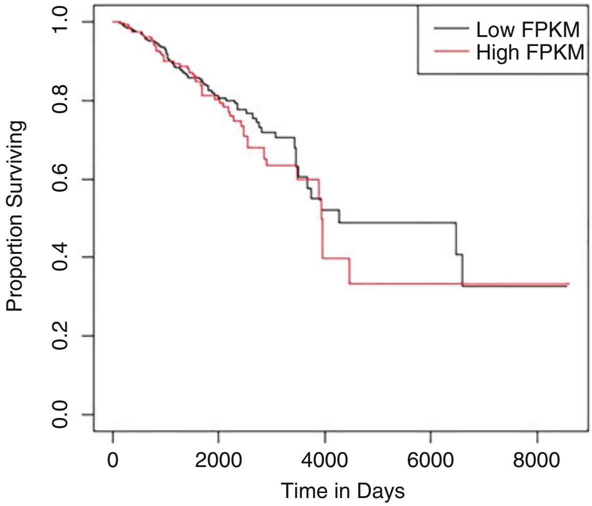
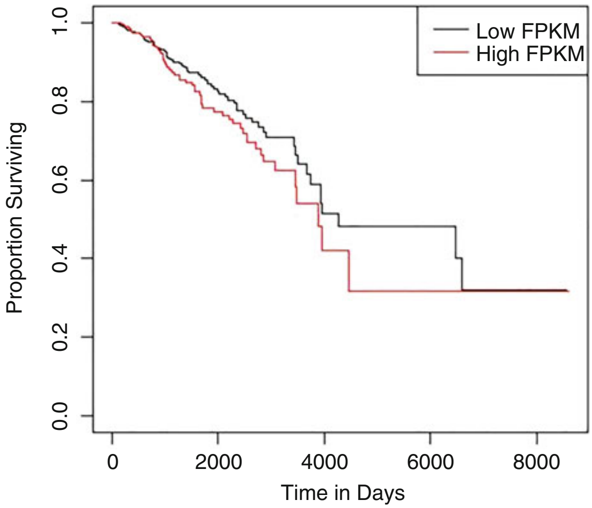


Fig. 6 KM plot for gene BRCA2



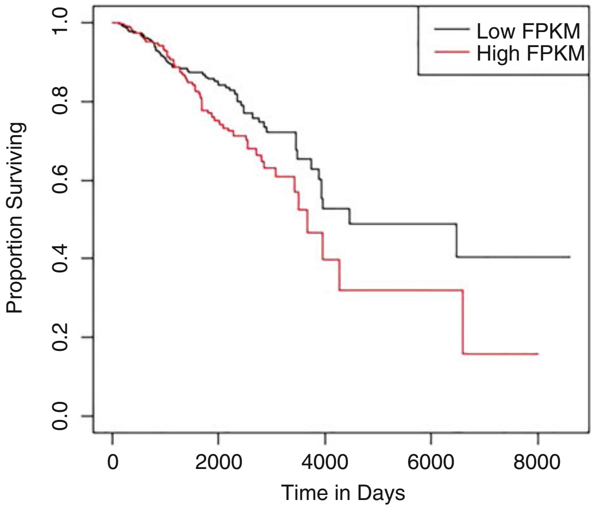


Fig. 7 KM plot for gene ATM

From Figs. 5 and 6, we observed that BRCA1 and BRCA2 have similar results in both high and low FPKM, i.e., approximately 30% survival rate, and also have the same survival time (8000 days). But in the case of the gene ATM (Fig. 7), survival time slightly varies from low to high, and there is a significant difference of 20% in survival rate. In ATM, we observe that samples with low FPKM have higher chances of being alive than those with high FPKM values since high FPKM samples have 20% survival rate, whereas in low FPKM, it is 40%. From this analysis, we can conclude that breast cancer with low expression value of the gene ATM has a significant role in survival as compared to BRCA1 and BRCA2.

In case of **ovarian cancer**, we have considered the same genes BRCA1, BRCA2, and ATM for our analysis and the same mean value for categorizing low and high FPKM values.

Figures 8 and 9 represent the KM plot of the genes BRCA1 and BRCA2. We observe that in both the graphs, survival rate is almost the same, i.e., 10% of both genes behave the same in ovarian cancer with respect to low and high FPKM value. However, in case of ATM (Fig. 10), high FPKM has 0%, and low FPKM has 10% survival rate; hence, we conclude that in gene ATM, there is a difference between low and high FPKM. The survival rate differs by 10% in low and high FPKM, and the survival time is the same in case of ATM.

Survival Analysis Using Cox Regression

The Cox proportional hazard (PH) model is used to identify prognostic factors for survival analysis. Cox PH model aids in identifying the clinical factors associated with the cure of the disease and the variables associated with the time to recurrence

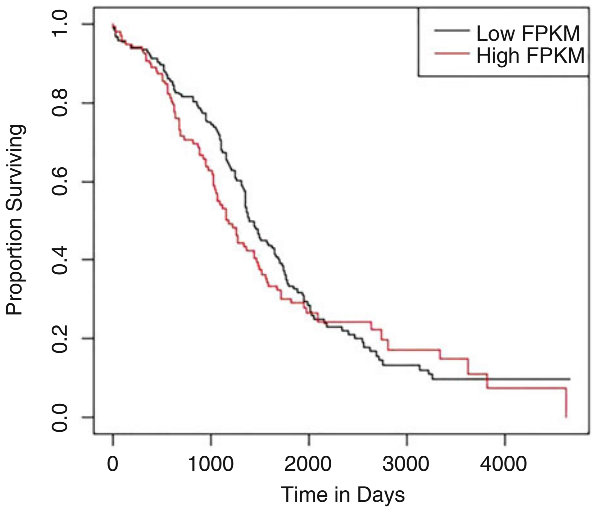


Fig. 8 KM plot for gene BRCA1

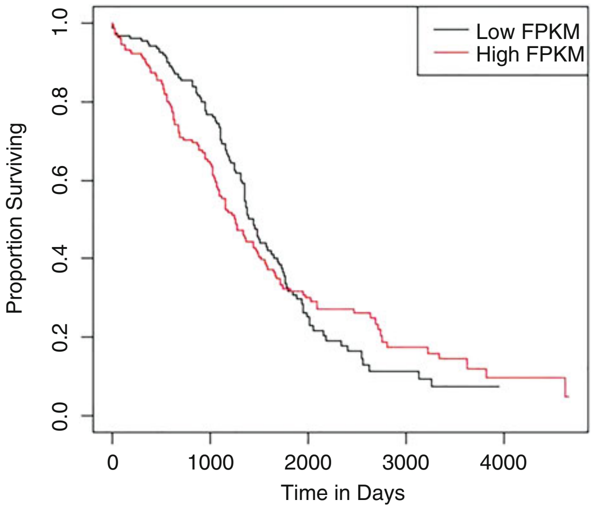


Fig. 9 KM plot for gene BRCA2

or death that cannot be identified using KM model. In our analysis, we have used breast and ovarian cancer data. Additionally we have considered three genes BRCA1, BRCA2, and ATM together for the analysis.

Stage: We have considered only breast cancer dataset for comparison of survivability with respect to stage where the stage information is available. The stage information has been categorized into four stages, i.e., stage I to stage IV. We have considered stage, age, and FPKM from the input data for the analysis. The average age is 58 in our dataset. For all the stages, i.e., stage I to stage IV, we have taken the mean age 58 and considered three genes BRCA1, BRCA2, and ATM together to plot the survival graph. Our aim is to understand how the survivability plot changes

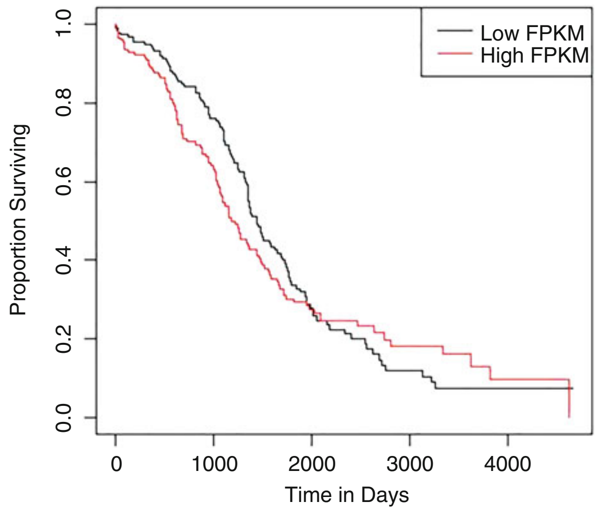


Fig. 10 KM plot for gene ATM

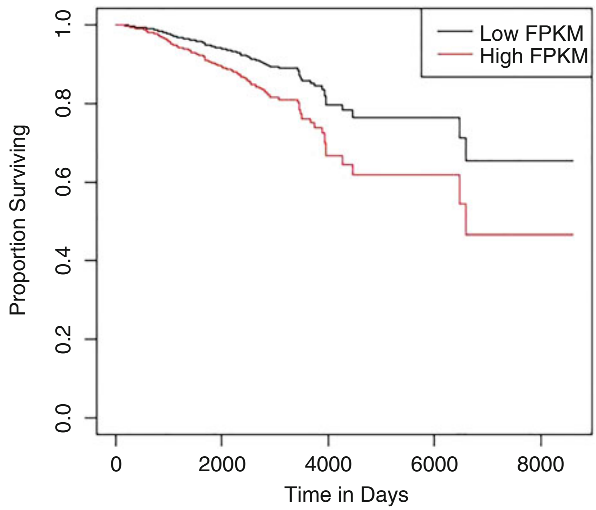


Fig. 11 Stage I Breast cancer Cox plot

with respect to lowly and highly expressed genes. Here we are considering the effect of stage on patient survival when either all the three genes of interest, i.e., BRCA1, BRCA2, and ATM, are lowly expressed or all of them are highly expressed. Figures 11 and 12 represent the Cox regression plot for stage I and stage II cancer, respectively.

We observe from stage I and stage II plot that there is a 20% difference of survival rate between low and high FPKM. In Fig. 11, we observe that for stage I, survival rate for low FPKM and high FPKM is 50% and 40%, respectively, and survival time is the same, i.e., 8000 days for both. However, in Fig. 12, we observe that in case of

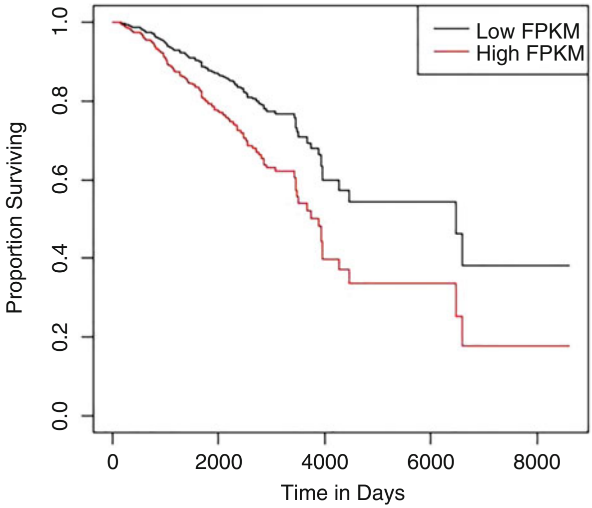


Fig. 12 Stage II Breast cancer Cox plot

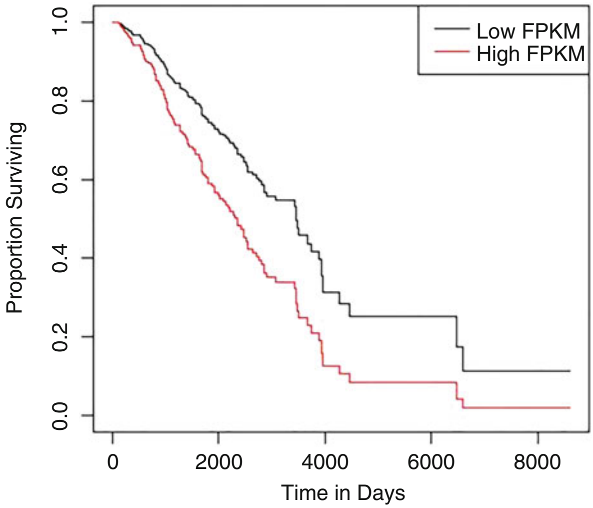


Fig. 13 Stage III Breast cancer Cox plot

stage II cancer, the survival rate in low FPKM is 40% and high FPKM is 20%. The survival time is the same as stage I, i.e., 8000 days.

Figures 13 and 14 represent the Cox regression plot for stage III and stage IV cancer. We observe that they have slightly different plots compared to stage I and stage II cancer. From both the Figs. 13 and 14, it is evident that there are no similarities between plot for stage III and stage IV. In case of stage III, there is a 10% difference between low FPKM and high FPKM with the same survival time, i.e., 8000 days. High FPKM have almost 0% survival rate and low FPKM have 10% survival rate.

Fig. 14 Stage IV Breast cancer Cox plot

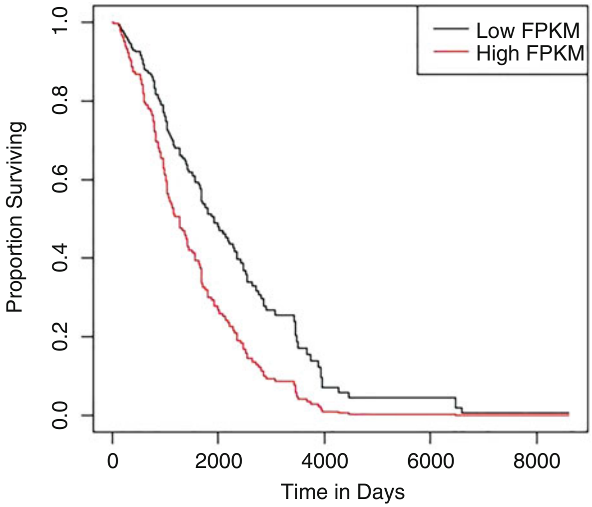
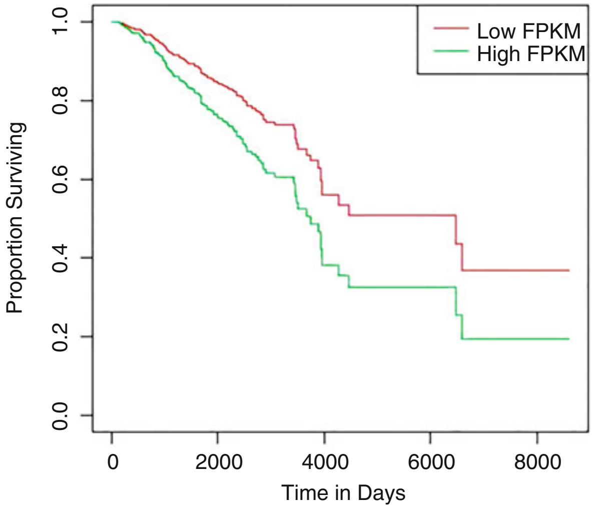


Fig. 15 Cox plot for ethnicity white for breast cancer



In case of stage IV, both high FPKM and low FPKM have approximately 0% survival rate with the same survival time. We clearly observe from the graphs for stage I to stage IV that the survival rate decreases from 40% to 30% to almost 0%. We also observe that high FPKM always has a low survival rate in all the four stages.

Ethnicity: Figure 16 represents the Cox regression plot for black or African American, and Fig. 15 represents the Cox regression plot for white population for breast cancer.

We have done analysis on both breast and ovarian cancer datasets based on ethnicity, grouped into black or African American and white. In the breast cancer dataset, we can observe from both the graphs that white population has a higher survival rate than black or African American population. In the case of black

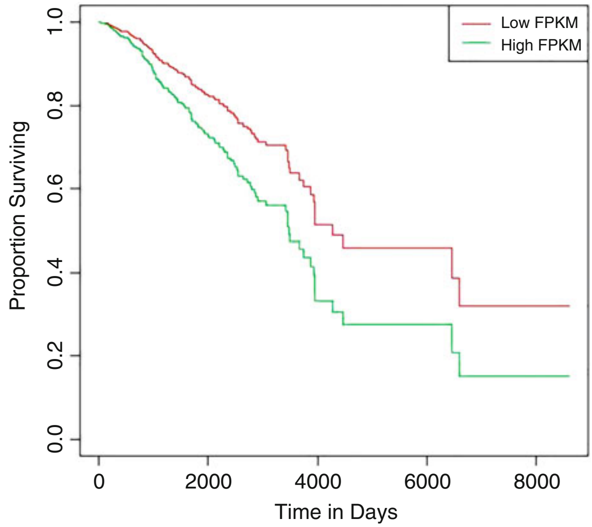


Fig. 16 Cox plot for ethnicity black or African American for breast cancer

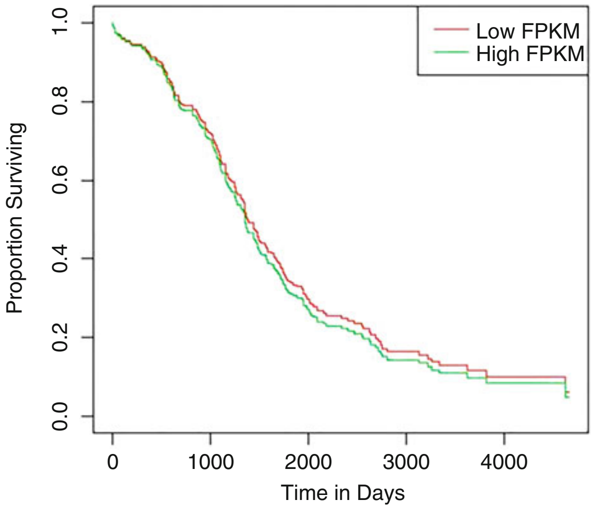


Fig. 17 Cox plot for ethnicity white for ovarian cancer

population, the survival rate of high FPKM and low FPKM is less than 20% and 40%, respectively, and survival time is the same for both expression values. But in the case of white population, the survival rate of low FPKM and high FPKM is almost 20% and 40% (Fig. 16).

Figure 17 represents the Cox regression plot for black or African American, and Fig. 18 represents the Cox regression plot for white population for ovarian cancer. In the ovarian cancer dataset, we observe that both colors have almost the same survivability.

There is no significant difference between high FPKM and low FPKM values for both black or African American and white populations. Black or African American

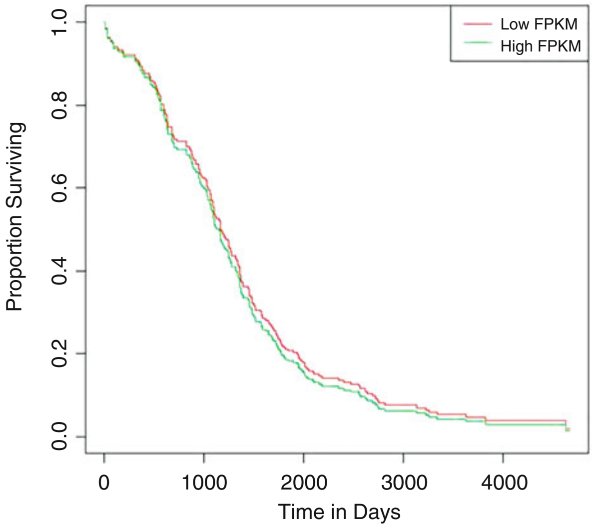


Fig. 18 Cox plot for ethnicity black or African American for ovarian cancer

have almost 0% survival rate, whereas white has 10% survival rate with the same survival time.

We have used the Cox regression model using the following covariates, age, ethnicity, and three genes BRCA1, BRCA2, ATM for both ovarian and breast cancers, where stage-specific data is available for breast cancer only. We begin by computing multivariate Cox analyses for all these variables. In Cox regression model, we have defined the number of events as the number of non-censored subjects for whom an event occurred. In our dataset, we have 946 samples (i.e., n = 946) for breast cancer, and the number of events observed is 133. For ovarian cancer, we have 353 samples (n= 353), and the number of events observed is 220. Here, we have fit multivariate Cox analysis using six variables for breast cancer and five variables for ovarian cancer to describe the factors that cohesively impact survival.

In Tables 1 and 2, we have interpreted the Cox regression results as follows:

1. The first attribute describes the variates used for our analysis.
2. The second attribute "coef" describes the regression coefficients. The positive sign of coef means that the hazard is higher and the prognosis is worse for subjects with higher values of that variable. Similarly, negative value indicates a lower risk of death.
3. The third attribute "exp(coef)" describes the hazard ratios which gives the effect size of covariates. A covariate with hazard ratio > 1 is called a bad prognostic factor, and a covariate with hazard ratio < 1 is called a good prognostic factor.
4. The fourth attribute "se(coef)" describes the standard error of the coefficient.
5. The fifth attribute marked "z" gives the Wald statistic value. It gives the statistical significance of each variable measured as the ratio of each regression coefficient to its standard error (z = coef/se(coef)).

Table 1 Breast cancer summary of multivariate Cox regression

	coef	exp(coef)	se(coef)	z	Pr(> |z|)
Age	0.038209	1.038949	0.006757	5.655	1.56e-08
Stage	0.822814	2.276897	0.118613	6.937	4.01e-12
Ethnicity	0.195266	1.215635	0.182884	1.068	0.286
BRCA1.FPKM	−0.031426	0.969063	0.208913	−0.15	0.88
BRCA2.FPKM	0.326704	1.386391	0.217864	1.5	0.134
ATM.FPKM	0.29432	1.342213	0.182291	1.615	0.106

Table 2 Ovarian cancer summary of multivariate Cox regression

	coef	exp(coef)	se(coef)	z	Pr(> |z|)
Age	0.019784	1.019981	0.006327	3.127	0.00177
Ethnicity	0.362809	1.437361	0.180482	2.01	0.04441
BRCA1.FPKM	−0.010422	0.989632	0.244623	−0.043	0.96602
BRCA2.FPKM	−0.340948	0.711096	0.336462	−1.013	0.3109
ATM.FPKM	0.427266	1.53306	0.388078	1.101	0.27091

From Table 1, we conclude that age and stage are significant for survival analysis using the multivariate Cox regression because Pr($>$ |z|) is less, i.e., the probability is within the significant cutoff ($p < 0.05$). The Cox model summary gives the hazard ratio (HR) for the second or later group relative to the first group. Although FPKM variables do not qualify the significant cutoff level for all the three genes, we observe that BRCA1 FPKM being high (BRCA1.FPKM=2, where 1 corresponds to low FPKM and 2 corresponds to high FPKM) reduces the hazard by a factor of 0.97 or 3%. Thus having high FPKM value of BRCA1 can be considered as an indicator for good prognosis factor.

Like that of breast cancer, here also we observe age and ethnicity to be a significant variable (pval = 0.00177 and pval = 0.04441 for age and ethnicity, respectively). Similar to breast cancer, FPKM variables do not qualify the significant cutoff level for these three genes, yet we observe that having high BRCA1.FPKM and high BRCA2.FPKM (i.e., FPKM = 2, where FPKM = 1 ~low expression and FPKM = 2 ~high expression) reduces the hazard by a factor of 0.98 or 2% and 0.71 or 29%, respectively. We conclude that having a high FPKM value of BRCA1 and mostly BRCA2 is associated with a good prognosis. To test the global statistical significance of the Cox model, we have generated p-values for three asymptotically equivalent alternative tests – likelihood-ratio test, Wald test, and score log-rank statistics. For breast cancer, likelihood-ratio test gives p-value = 2e-14, Wald test gives p-value = 4e-15, and score log-rank test gives p-value = 3e-15. For ovarian cancer, likelihood-ratio test, Wald test, and score log-rank test gives equivalent p-value = 0.01. Thus, we find that both the models satisfy the criteria of being statistically significant. However, the statistical significance of Cox regression model has better behavior for large sample sizes, viz., breast cancer here; hence the score is better.

6 Conclusion

In this chapter, we have weighed up the survival analysis by Kaplan-Meier method using breast and ovarian cancer data. We have found the survival rate and time with respect to a single attribute, i.e., genes (either BRCA1, BRCA2, or ATM) or stage or ethnicity. Due to the limitations of Kaplan-Meier, we have further analyzed the same cancer data using Cox regression method, where we have observed better results with respect to Kaplan-Meier. In Cox regression, we have used multivariate Cox regression for three genes (BRCA1, BRCA2, and ATM) together with age, stage, and ethnicity. Shared genetic mutation between breast and ovarian cancers is well known with the existence of a common 25-gene testing panel for both the carcinomas. We have considered three of those genes for this study to evaluate their potential as prognostic markers in terms of their expression in association with overall patient survival. The Cox regression score for these three genes is much below the significant cutoff level indicating high or low expression of these genes has no striking difference on overall patient survival. The sample size of breast cancer is larger($n = 946$), and the maximum survival time has gone up to 8000 days. On the contrary, ovarian cancer, where sample size is around one third to that of breast cancer ($n = 353$), 4000 days is the maximum survival time recorded. Thus, throughout our analysis, the ovarian model performed poorer than that of breast. Due to this limitation, we plan on accumulating more data and conduct further analysis incorporating the rest of the marker genes in the future.

We conclude that Kaplan-Meier estimation is used to create groups from the observed survival curves, while log-rank test is used to compare the curves from different groups. Survival regression models such as the Cox model are used to test continuous predictors or multiple covariates at once. We have used the R package "survival" (https://github.com/therneau/survival) for the statistical analysis of breast and ovarian cancer data for survival analysis. The code and input files are available at https://github.com/sudipmondalcse/SurvivalAnalysis. In this chapter, we have focused on female cancer, but in future, we aim to focus on popular male cancers along with cancer among children and teenagers for survival analysis. We aim to implement random forest analysis hereafter to gain further optimal results. We have used one of the major features of genomics, i.e., FPKM or expression values of wrt genes, but we will train our model with structural variations like single nucleotide polymorphisms (SNP), copy number variation (CNV), and insertion or deletion (Indel) information in future for better survival analysis.

References

1. S. Verma, D. Bakshi, V. Sharma, I. Sharma, R. Shah, A. Bhat, G.R. Bhat, B. Sharma, A. Wakhloo, S. Kaul, et al., Int. J. Gynecol. **148**(1), 118 (2020)
2. S. Mallick, C. Kumar, R. Benson, G.K. Rath, in Practical Radiation Oncology (Springer, Singapore, 2020), pp. 299–300

3. S. RL, CA Cancer J. Clin. **70**(1), 7 (2020)
4. R.B. Prasad, V.S. Rathna, C. Nithish, R. Sowmya, A. Arora, S. Buchanalli, Cancer **6**(2), 48–51 (2020)
5. M.K. Mallath, D.G. Taylor, R.A. Badwe, G.K. Rath, V. Shanta, C. Pramesh, R. Digumarti, P. Sebastian, B.B. Borthakur, A. Kalwar, et al., Lancet Oncol. **15**(6), e205 (2014)
6. S. Thorlacius, G. Olafsdottir, L. Tryggvadottir, S. Neuhausen, J. Jonasson, S. Tavtigian, Nat. Genet. **13**, 117 (1996)
7. R. Dikshit, P.C. Gupta, C. Ramasundarahettige, V. Gajalakshmi, L. Aleksandrowicz, R. Badwe, R. Kumar, S. Roy, W. Suraweera, F. Bray, et al., The Lancet **379**(9828), 1807 (2012)
8. D. Vij, Stud. Indian Place Names **40**(40), 1834 (2020)
9. R. Srinivasan, Planning Commission. Nic. In reports/genrep/—-/26_bg (2020)
10. S. Ganguly, S. Kinsey, S. Bakhshi, Cancer Epidemiol. **71** Part B, 101679 (2020)
11. P. Mobadersany, S. Yousefi, M. Amgad, D.A. Gutman, J.S. Barnholtz-Sloan, J.E.V. Vega, D.J. Brat, L.A. Cooper, Proc. Natl. Acad. Sci. **115**(13), E2970 (2018)
12. R. Singh, K. Mukhopadhyay, Perspect. Clin. Res. **2**(4), 145 (2011)
13. K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Comput. Struct. Biotechnol. J. **13**, 8 (2015)
14. J. Xie, C. Liu, Stat. Med. **24**(20), 3089 (2005)
15. V. Bewick, L. Cheek, J. Ball, Crit. care **8**(5), 389 (2004)
16. J. Liu, T. Lichtenberg, K.A. Hoadley, L.M. Poisson, A.J. Lazar, A.D. Cherniack, A.J. Kovatich, C.C. Benz, D.A. Levine, A.V. Lee, et al., Cell **173**(2), 400 (2018)
17. N.K. Mishra, S. Southekal, C. Guda, Front. Genet. **10**, 624 (2019)
18. S. Yousefi, F. Amrollahi, M. Amgad, C. Dong, J.E. Lewis, C. Song, D.A. Gutman, S.H. Halani, J.E.V. Vega, D.J. Brat, et al., Sci. Rep. **7**(1), 1 (2017)
19. B. Crawford, S.B. Adams, T. Sittler, J. van den Akker, S. Chan, O. Leitner, L. Ryan, E. Gil, L. van't Veer, Breast Cancer Res. Treat. **163**(2), 383 (2017)
20. A. Kumar, S.M. Bhagabaty, J.P. Tripathy, K. Selvaraj, J. Purkayastha, R. Singh, Asian Pac. J. Cancer Prev. **20**(12), 3711 (2019)
21. H.A. Park, S.R. Brown, Y. Kim, Int. J. Mol. Sci. **21**(14), 5040 (2020)
22. F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, CA Cancer J. Clin. **68**(6), 394 (2018)
23. S. Altekruse, C. Kosary, M. Krapcho, N. Neyman, R. Aminou, W. Waldron, J. Ruhl, N. Howlader, Z. Tatalovich, H. Cho, et al., Bethesda, MD: National Cancer Institute (2010)
24. S. Altekruse, C. Kosary, M. Krapcho, N. Neyman, R. Aminou, W. Waldron, et al., MD National Cancer Institute (2015)
25. O.L. Griffith, J.W. Gray, Breast Cancer Res. **13**(6), 230 (2011)
26. L.J. van't Veer, S. Paik, D.F. Hayes, J. Clin. Oncol. **23**(8), 1631 (2005)
27. J.S. Reis-Filho, L. Pusztai, The Lancet **378**(9805), 1812 (2011)
28. U. Singh, V. Solanki, B. Prakash, S. Mehrotra, M.L. Verma, V. Solanki, Indian J. Gynecol. Oncol. **18**, 59 (2020)
29. F. Ahmed, E. Wiltshaw, R. a'Hern, B. Nicol, J. Shepherd, P. Blake, C. Fisher, M. Gore, J. Clin. Oncol. **14**(11), 2968 (1996)
30. N. Kumar, A. Yadav, H. Kaur, Indian J. Gynecol. Oncol. **18**(1), 22 (2020)
31. M.S. Piver, J. Malfetano, T.R. Baker, R.E. Hempling, Gynecol. Oncol. **46**(3), 357 (1992)
32. J.D. Wright, M. Shah, L. Mathew, W.M. Burke, J. Culhane, N. Goldman, P.B. Schiff, T.J. Herzog, Cancer Interdiscip. Int. J. Am. Cancer Soc. **115**(18), 4118 (2009)
33. S. Pilleron, D. Sarfati, M. Janssen-Heijnen, J. Vignat, J. Ferlay, F. Bray, I. Soerjomataram, Int. J. Cancer **144**(1), 49 (2019)
34. B.B. Yeole, A.V.R. Kumar, Asian Pac. J. Cancer Prev. **5**(2), 175 (2004)
35. Z. Zhang, J. Sun, Stat. Methods Med. Res. **19**(1), 53 (2010)
36. R.G. Miller Jr, Survival Analysis, vol. 66 (Wiley, 2011)
37. J.P. Klein, H.C. Van Houwelingen, J.G. Ibrahim, T.H. Scheike, *Handbook of Survival Analysis* (CRC Press, Boca Raton, 2016)
38. B. George, S. Seals, I. Aban, J. Nucl. Cardiol. **21**(4), 686 (2014)

39. M. Mills, *Introducing Survival and Event History Analysis* (Sage, Los Angeles, 2010)
40. M.P. Robertson, C.I. Peter, M.H. Villet, B.S. Ripley, Ecol. Model. **164**(2–3), 153 (2003)
41. S. Chatterjee, A.S. Hadi, *Regression Analysis by Example* (Wiley, New York, 2015)
42. E.L. Kaplan, P. Meier, J. Am. Stat. Assoc. **53**(282), 457 (1958)
43. M.K. Goel, P. Khanna, J. Kishore, Int. J. Ayurveda Res. **1**(4), 274 (2010)
44. İ. Etikan, S. Abubakar, R. Alkassim, Biom. Biostatistics Int. J. **5**(2), 00128 (2017)
45. E. Papastavrou, A. Charalambous, H. Tsangari, Eur. J. Oncol. Nurs. **13**(2), 128 (2009)
46. I.R. Savage, Ann. Math. Stat. **27**(3), 590 (1956)
47. J.D. Storey, R. Tibshirani, Proc. Natl. Acad. Sci. **100**(16), 9440 (2003)
48. J.M. Bland, D.G. Altman, BMJ **328**(7447), 1073 (2004)
49. B. Damato, A. Taktak, in *Outcome Prediction in Cancer* (Elsevier, Amsterdam, 2007), pp. 27–41
50. D.R. Cox, J. R. Stat. Soc. Ser. B Stat Methodol. **34**(2), 187 (1972)
51. M.J. Bradburn, T.G. Clark, S.B. Love, D.G. Altman, Br. J. Cancer **89**(3), 431 (2003)
52. A. Cantor, Sas® survival analysis techniques for medical research (sas institute inc. cary, nc) (2003)
53. L. Breiman, Mach. Learn. **45**(1), 5 (2001)
54. U.B. Mogensen, H. Ishwaran, T.A. Gerds, J. Stat. Softw. **50**(11), 1 (2012)
55. H. Esmaily, M. Tayefi, H. Doosti, M. Ghayour-Mobarhan, H. Nezami, A. Amirabadizadeh, J. Res. Health Sci. **18**(2), 412 (2018)
56. H. Ishwaran, U.B. Kogalur, E.H. Blackstone, M.S. Lauer, et al., Ann. Appl. Stat. **2**(3), 841 (2008)
57. S. Wongvibulsin, K.C. Wu, S.L. Zeger, BMC Med. Res. Methodol. **20**(1), 1 (2020)
58. M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, et al., Science **347**(6220) (2015)
59. C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. Van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Nat. Biotechnol. **28**(5), 511 (2010)

Machine Learning with IoT and Big Data in Healthcare

Tasneem Jahan

1 Introduction

Healthcare experts have taken a step ahead towards personalized healthcare, which incorporates artificial intelligence techniques with web-based and social media information, along with electronic health records (EHR), wearable devices, mobile devices, sensor devices and Internet of Things (IoT). Machine learning makes data collection easier and improves the tracking of disease propagation, effectively detects patterns and levels of transition, and ultimately builds the patient management systems with ease and at fast pace. The fundamental feature of learning models is to carry out analysis over collected sensor data and generate the patterns specifying behavioural characteristics and clinical requirements for the patients or affected population. Machine learning brings an automated image of healthcare systems. The theme of this chapter explains the significance of machine learning in healthcare. Learning models integrated with artificial intelligence and cloud computing create the foundation of IoT and big data technologies in healthcare.

The ambitions of artificial intelligence and machine learning have been greatly achieved in medical science. Artificial intelligence and automation have credited many discoveries to the various sectors of the digital industry globally, and machine learning has had its sway in recent times. Computer vision and image analysis are helpful in diagnostics. Deep learning plays a crucial role in applications used for diagnostics. It efficiently explores and updates the artificial intelligence process by utilizing the data sets of medical images. Microsoft carried out an InnerEye initiative. It works with the tools of image diagnosis. Big data uses business intelligence by data analytics for data-driven decisions. Predictive analytics assists in error-free treatment and minimizes fraud.

T. Jahan (✉)
Truba Institute of Engineering and Information Technology, Bhopal, Madhya Pradesh, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_5

Deep learning is extensively used to discover hidden patterns in clinical data and to explore potential strategies for treatment of patients. Thus, lowers the risks and efficiently diagnoses the life-threatening disease beforehand. IoT and wearable devices collect voluminous sensor data. These records and patient's information, treatment history, medical reports and insurance reports are processed by deep learning algorithms, and decisions are drawn. In recent scenarios around the globe, the COVID-19 outbreak has claimed the lives of more than 20 million people. Digital technologies involving machine learning, IoT, big data have become a powerful tool to tackle COVID-19. Effective monitoring, tracking, detection and prevention strategies have been developed using the digital framework.

2 Background

Machine learning is extensively used in various business domains. The techniques combined with artificial intelligence, deep learning, big data and data mining help researchers and scientists to understand relations and detect patterns in medical data. Knowledge Discovery in Databases (KDD) helps in analysing complex medical data generated from wearables, instruments and sensors. Machine learning facilitates the expert analysis of data. It accelerates the diagnosis and treatment of diseases with reduced time and reduced errors.

Ali Hasan in his work proposed efficient quality of service (QoS) methods through intelligent learning models for medical data processing [1]. Healthcare applications have been restructured with the development of machine learning in technologies such as IoT, cloud computing and big data. Idemen B.T., Sezer E. and Unalir M.O. presented how big data creates a scalable model in cloud application for data protection [2]. It uses machine learning algorithms to identify violations in data.

In the view of Yousefi S., machine learning provides efficient models in IoT to drive data analysis, communication and security in a healthcare system [3]. They also discussed challenges faced by machine learning approaches in IoT system. In the research published by Xie, he introduced a healthcare system that focuses on data optimization to save size and time during prediction [4]. Their work is concentrated on personalized healthcare with IoT. K. R. Dalal in his report presented an analysis of machine learning and how its implementation revolutionizes the healthcare industry [5].

3 Machine Learning

Machine learning handles massive medical data in a smart, fast and cost-effective manner with its learning models. It increases efficiency of data analytics for diagnostics, symptom research, pattern identification and patient treatment. Machine learning contributes to intelligent healthcare for pattern analysis from big data and to study pathogens, viruses and symptoms of various diseases. Image-based diagnostics have gained accuracy with the help of machine learning models. Personalized healthcare and personalized treatment strategies are possible with machine learning.

Below are the machine learning algorithms widely used in Intelligent Healthcare framework.

1. Support vector machine (SVM) classifier: Support vector machine is a supervised learning model for classification and regression analysis. It builds a non-probabilistic, binary, linear classifier and separates input data by dividing it into categories as wide as possible. C. Venkatesan in their work presented how an SVM classifier is used for pre-processing and abnormality detection in ECG signal [6].
2. Naïve Bayes: It is a probabilistic classifier and assigns class labels to input data. B. A. Patel and A. Parikh developed a model by using Naïve Bayes classifier to predict anaemia [7]. They analysed different parameters of the complete blood count (CBC) test. In the study carried by N. Ravindran and O. A. Sheryl, they applied Bayesian classifiers to identify safe and unsafe scenarios when children play and generated a warning assistance [8]. The same methodology is applied to identify the normal and contraction time of a pregnant woman.
3. Decision trees: It is a predictive modelling machine learning approach. It splits input data into subsets based on an attribute value test. It does not require much computation to perform classification. Y. P. Sinha and P. Malviya in their work proposed a method for patient healthcare that acts as an automated, self-adapting and contextual care protocol [9]. T. Xie tested a research on Heartbeat Classification Algorithm Based on CART Decision Tree [10]. They proposed a new method for premature ventricular contraction (PVC) detection based on abnormal eigenvalues and decision tree.

Support vector machine algorithms have high accuracy rates compared to other algorithms of machine learning. But it also suffers with a drawback of a longer training period. It works with two classification algorithms, and SVM classifier would only classify whether the input symptoms are related to a disease or not.

Naive Bayes is extensively used for classifying reviews from patients in sentiment analysis. Naive Bayes helps in obtaining consumer and patients' thoughts as feedback on different drugs and medicines. Their sentiment helps to draw conclusions on whether a medicine is good or not effective for usage.

Application Areas of Machine Learning in Intelligent Healthcare

1. Pathogens – Machine learning has become a very efficient tool in detection of pathogens. The algorithms based on support vector machines and tools of feature selection are extremely popular to identify a set of pathogens and in discriminating their potential and impact. When a pathogen is identified early in a person's body, then it is far easier to initiate relevant treatment of the infectious pathogen. This reduces patient's panic, diagnostic, healthcare and medicine costs. The spread of microbial pathogens can also be controlled. Conventional methods of sampling a culture and tests are a week long process. The machine learning-integrated systems can perform detection and diagnosis in a time-efficient manner, with cost-effective steps. Learning models work with extremely huge and dynamic data sets. Therefore, they withdraw related information extremely fast. Artificial intelligence is a driver of machine learning. Together artificial intelligence and machine learning algorithms can determine various pathogens in effective computational time.
2. Shortage of doctors – Many countries face a serious shortage of healthcare service providers. Robots can serve the purpose and help solve this issue. Robotic process automation (RPA) is an emerging technological trend. RPA uses machine learning models, and in the healthcare sector, it provides a doctor-level medical knowledge. This is an excellent solution to overcome the shortage of qualified doctors.
3. Decision-making – Natural language processing (NLP) offers assistance to physically challenged and handicapped staff and paralytic patients. The integration of machine learning in IoT assists patients in their routine activities.
4. Clinical examination – Early diagnosis of high-risk patients is easily done in an intelligent healthcare system with machine learning. Cancer and tuberculosis can be identified in its early stage. Screen and cell mutation can be observed.
5. Diagnosis of heart diseases – Heart is the most vital organ of the human body. SVM and Naive Bayes are used to detect heart diseases. Early diagnosis is a remarkable achievement of machine learning algorithms in healthcare.
6. Diabetes prediction – Diabetic patients have drastically increased across the world in the past two decades. Diabetes damages other organs of the body and also hinders the treatment of any underlying disease. Diabetes prediction systems are created using classification algorithms, such as KNN, Decision Tree and Random Forests. The algorithms work fast and have excellent accuracy.
7. Liver diseases prediction – Liver is also a significant organ in the human body. It plays a very important role in metabolism. Liver diseases such as fatty liver, cirrhosis, liver cancer and hepatitis are prolonged and life threatening. Various classification and clustering algorithms of machine learning help in the liver disease prediction system.

8. Robot-assisted surgery – It is a modern surgical procedure done with a robotic system. A camera and surgical tools are attached to robotic arms. Machine learning allows robot-assisted surgery with minimal or no invasion from a surgeon.
9. Enhancing genomic sequencing – Deep learning algorithms could improve genomic sequencing processes, identify DNA methylation and any polymorphisms, which may later cause a disease. Machine learning algorithms can successfully perform prediction on large-scale genome data.
10. Detecting tumours and cancer prediction – Detection and classification of tumours are being successfully done by machine learning algorithms. Deep learning algorithms are capable of detecting cancer and reducing error percentage in diagnosis of cancer. CNN is widely applied in the classification of cancer by features extraction from gene sequence data.

Advantages of Machine Learning

1. Deep learning is a step-ahead technology to machine learning. Problems that could not be solved by machine learning have been proven solvable through deep learning. The driver of deep learning is neural networks. Neural networks give accurate results during illness diagnosis.
2. Deep learning is widely used in domains of NLP, speech recognition and face recognition. Recognition techniques also help to find patterns and to carry diagnostics.

4 Big Data in Healthcare

Big data has become a catalyst in the intelligent healthcare industry. The Process Insights Data (PID) approach of big data drives the transformation of conventional healthcare systems. Big data aims to re-engineer the healthcare industry to deliver automation and future-ready robust services for enhancing their business value and to gain intelligence in innovation. Healthcare data gathered from patients and healthcare professionals provide insights of analytics to give meaning of data. Big data addresses complex data and discovers patterns and meanings to help patients, researchers, health policy-makers, doctors, etc. Big data creates data reservoirs that subsequently become the source of knowledge to drive the engine of healthcare system. The analytics approaches integrated with artificial intelligence and machine learning aim to predict information and perform data discovery. Commercial enterprises utilize big data and analysis to research behaviour of consumers.

The visualization tools of big data are the potential prospects of intelligent healthcare. Data analytics with big data are giving a rapid momentum to healthcare services. Personalized medicine, EHR and clinical practices help to employ fast learning models. e-Health technology is emerging and rapidly growing with wearable sensors. They also support the promising tools of cloud computing. With deep learning models and convolutional neural network (CNN), low-level input data are extracted to gain high-level features and business insights. A knowledge-based healthcare system transforms towards online analytical processing (OLAP) with big data storage and analysis technology. There is a significant shift in the biological domain with feature extraction algorithms. Big data has potential for the future of intelligent healthcare. Its tools could detect and research environmental factors, UV radiation exposure, pandemic infections and their spread. Its decision-making and pattern recognition tools will help in assessing diagnostics and tests. A guaranteed treatment with effective and quality assured health services is delivered with big data. Big data helps in medical practices to design strategies for treatments, to deeply understand risks and complex diseases, in detecting safe drugs and medicines, and to compare diagnostic and prevention schemes. A better personalized treatment is based upon accurate information of benefits and risks. Big data generates accurate patterns and evident discoveries of predictions to help in decision-making. Big data aims to fill the void between clinical researches and medical practices. It gains insights from existing data by formulating hypotheses and deriving the cause of disease. Machine learning helps to create models that learn from data and design approaches for treatment to work with different patients in different environments. The deep learning strategies help to identify suitable drugs for any underlying cause. They also help in predicting epidemics and pandemics. For example, the COVID-19 outbreak was studied by identifying the patient clusters, by studying their symptoms taxonomies and then by defining outcomes based on biological, clinical and behavioural characteristics.

Big data methods are a powerful tool to understand huge amounts of data. It provides insights of data integrated with mathematical and computational approaches. The predictions and decision-making are facilitated by big data. Big data combined with artificial intelligence and parallel computing revolutionizes personalized healthcare. Diagnostic errors are reduced, and complex medical imaging is possible with big data [16]. The magnificence of big data lies in its power to keep the complexity of data but remove the influence of physical and biological factors on it. Medical data are voluminous and complex. Researchers face immense challenges in data exploration and data prediction. Big data efficiently handles real-time data exploration and enables data scientists and data analysts to predict future trends accurately. Thus, they make significant improvements in life-saving treatment strategies and cost of treatment gets reduced. Data analytics derived from big data tools are hugely transforming the healthcare sector. It is helping medical practitioners to create the right treatment plan for the patients at the right time.

Big Data Characteristics for Intelligent Healthcare

In data warehousing and data mining, there are three dimensions of data. The medical data emerging in huge volumes since the last few decades also exhibit these dimensions. The three dimensions that define the characteristics of medical big data are as follows:

1. Volume:
   The size of medical data is considered by its volume. Medical data are increasing with a rapid rate. EHR, medical imaging, genome sequencing, diagnosis, biometrics, sensors, wearable devices etc. are increasing the volume of medical data. Wearable devices and sensors continuously produce medical data.
2. Velocity:
   Since the inception of the client-server model in computing, data streams flow in a continuous manner. Data in intelligent healthcare is also emerging and growing rapidly. Devices in real-time applications produce data without delay.
3. Variety:
   Data in the medical domain also arrive in all shapes and sizes. Apart from a patient's physical characteristics data, there is also a variety of medical data in healthcare. Blood pressure and pulse rate monitoring, X-rays, ultrasound, CT scan, MRI, cancer and tumour biopsy, genome sequencing, RNA sequences, drugs and pharmacy are a few examples of variety in medical data. These heterogeneous data can be structured, semi-structured or unstructured.

Stages of Big Data Analytics

As discussed in the paper by M. Ambigavathi and D. Sridharan, there are 5 stages in intelligent healthcare for big data analytics [11]:

1. Data Collection and Storage:
   As discussed earlier, medical data are heterogeneous in nature. There can be various sources of medical data. Hence, the formats of data vary (variety). Data collection must take into consideration the security prospects and data privacy. Finding the correct metadata, which describes what kind of health data is stored, is also very difficult. All these issues are considered and resolved during the first phase of data analytics.
2. Data Cleaning, Extraction and Classification:
   Data from various sources is voluminous, and some information is not useful many times. If the not useful data is passed onto further phases, it only leads to computation errors and increases processing cost. Doctor's prescription, sensor's data, medical image and scan data are collected in an unstructured format. This must be transformed into the structured format before beginning the analysis. Adding and removing missing values are carried out in this phase.

3. Data Integration and Representation:

   The data obtained in previous phases are integrated to obtain detailed knowledge for effective data analysis. For example, the precise information of an electronic health record consisting of a patient's personal data, data of record creation, health status, diagnosis and treatment history, count of patient's visit, etc. are aggregated and shared among data analysts, researchers, hospitals and government agencies. Health records are sensitive data; this phase utilizes a suitable representation model for data representation in real-time scenario.
4. Data Modelling with Analysis and Query Processing:

   Diagrams, text and symbols are readable and easily understandable. Modelling helps to view interrelated data and relationships. Data analysis for the health data begins with query processing. Individuals produce a query to know the patient's health status. Big data analysts perform analysis according to the complexity of the query.
5. Data Interpretation with Feedback:

   Data interpretation is the final and important phase. It produces clear results that are later utilized in decision-making. The health reports are generated. Feedback is obtained from patients and decision-makers.

The integration of machine learning and artificial intelligence with big data has proven an innovative trend to fight the global COVID-19 pandemic. It is helpful in efficient screening, tracking of patients, contact tracing and predicting symptoms. Early detection and diagnosis of ailments is the major application of big data with artificial intelligence. During coronavirus spread, the pace of developing drugs and vaccines has accelerated, and big data assisted healthcare practitioners to combat the disease. Raju Vaishya in their paper discussed the role of big data as a decisive technology for analysing, preventing and fighting against the COVID-19 [12]. The significant application of machine learning models helped to detect the clusters of positive cases and to predict the spread of virus-predictive analytics tools. The collection of patient's data, asymptotic patients and recovery rate helps to analyse the spread using big data technology. The proficiency of learning models grows with real-life situations and data-driven tools. With machine learning and AI, big data framework mimics human intelligence.

The integration of big data with IoT is producing remarkable outcomes in intelligent healthcare [13]. Data collection, storage and analysis are done among cloud services. Patients at remote locations are assisted by doctors through various smart devices, and the emergency cases are also handled appropriately. Big data is the leader in computing technologies, leading the framework of intelligent healthcare. Information extraction, analysis and decision-making are greatly impacted with the inception of big data. Together with machine learning, big data and IoT are helping medical professionals in diagnosis and treatment of several chronic illnesses. Caregivers can now offer immediate help to patients.

5 IoT in Healthcare

Traditional healthcare has observed a digital transformation. This shift has gained momentum due to continuously emerging sensor data. Connected devices exchange seamless data. Data analytics with IoT and big data creates an intelligent healthcare framework. The decision support system created over IoT brings a technology-efficient healthcare platform, which is robust and caters preventive medical needs. The key factors of a decision support system are a data set, inference engine and communication model. With deep learning tools, the insights of patients' data are utilized for treatment of a disease and to explore the diagnosis of related risks for predictive analytics. Early detection of chronic diseases is supported by decision support systems. They are now extensively used in automated analysis systems. IoT establishes a high-end connected network with devices and shares information. It aims to provide the following:

1. Better user experience.
2. Personalized healthcare.
3. Multiple stakeholders communicate at a single platform with information on multiple technologies.
4. Secure platform for a wide range of protocols and applications.
5. It plays an efficient role in diagnostics, monitoring and assistance.

IoT (Internet of Things) is a technology that connects smart devices and wearables to network applications. They track health records, medical history, diagnosis, treatment and risks. The entire hospital management system is being revolutionized with IoT [17]. Examples of IoT devices are fitness bands, fitness trackers, pulse oximeters, smart clothing, wearables, smart glasses, etc. IoT increases physician's engagement with patients, resulting in improved treatment results and reduced costs of healthcare. IoT benefits all elements of a healthcare system, that is, patients, families, physicians and insurance business. Patients are able to access personalized treatment and attention with wearables and digital devices. Patients have handy records of their calorie intake, workout information, blood pressure and heart rate reports. Healthcare providers are also equipped with IoT and monitoring systems. They keep track of the patient's fitness and well-being. Physicians stay alert with patients and connect with them proactively. With predictive analysis, the physicians could find the best treatment for patients. Hospitals utilize IoT technology to track hospital supplies and equipment. Wheelchairs, nebulizers, monitors, oxygen pumps etc. are labelled with sensors and their real-time location can be easily monitored. Infection control is also monitored with IoT-enabled devices. Data acquired by health monitoring sensor devices are supportive to health insurance companies. Claims handling, pricing underwriting and risk assessment process are enabled through IoT-driven framework. Fraud claims can also be detected early and valid claims are adhered to treatment measures.

IoT Application Areas in Healthcare

1. Electronic Health Records

   Data management is the most attractive essence of IoT. The global medical records are increasing with each passing data and bring along significant challenges in data management. It is estimated that compound annual growth rate (CAGR) of healthcare data will rise by 36% by 2025. The fundamental focus of the IoT paradigm is healthcare records. In the last few years, nearly 90% of medical set-ups have adopted electronic health records (EHR) and electronic medical records (EMR). IoT offers an efficient platform to collect, store and process data. The electronic health records store medical history of patients, their allergies and follow-up details. EHR is the most important and highly used application of IoT in healthcare. It offers robustness in record keeping and efficiently manages consistency in creation of records. It is also a cost-effective application that eventually makes saving in millions to stakeholders. Medical devices, fitness watches and wearable devices produce data that are integrated with the cloud network of healthcare systems. This data are used at patient, hospital, robot and network levels.
2. Pharma and Drug Discovery

   IoT in pharmaceuticals is an advanced technology to discover treatment opportunities for new emerging diseases. IoT enables management of medication and tracks records of improvement in patient health conditions. Development of drugs, management of supply chain and clinical trials are now based on IoT solutions. Sensor devices and connected technologies are improving the concept of personalized healthcare. IoT boosts competency and innovation among healthcare providers. Big data and analytics play a crucial role to improve a patient's quality of life. Time and costs of research and development are reduced. Data are handled in an efficient way with great speed in the development of medicines and drugs. The process of clinical trials and experiments gets shortened to find conclusions. IoT sensors handle chemicals and biomaterials flawlessly ensuring controlled manufacturing and preventing drug fraud. For example, Itransition is an expert software development company in the healthcare industry. It delivers custom healthcare solutions. Itransition builds IoT-based health applications and utilizes cloud data to collect and monitor data analytics. Pharma industry needs 24×7 monitoring of the drug discovery process. Maintenance of equipment is also crucial. IoT continuously tracks status information in drug manufacturing units for air compressors, vacuum pumps, sterilizers, temperature, pressure, pH probes, labels, RFID tags etc. The sensor devices of IoT technology are performance oriented and optimize predictive equipment maintenance. Small IoT sensors are produced by AntTail, a Dutch start-up, and these can be directly placed on drug packages. Thus, they monitor the manufacturing process down supply chain with GPS-enabled vehicles.

3. Personalized Treatment

   IoT framework is a great advantage for healthcare professionals in treating patients. It provides medical history through sensors, wearables, wireless technology and mobile networks. Entire data are stored in a single cloud network and are easily accessible. Personalized and custom-tailored treatment plans and diagnostics can be prepared. A patient with diabetes has a health record on a secure cloud. This is readily accessed by medical staff on mobile devices or desktop computers. Test samples record, lab results, prescription details are easily accessible, and in no time the patient will be suggested what things he has to avoid, medicine suggestions and work out details if needed. Remote care is also facilitated by IoT, as it provides a virtual hospital concept to patients. The medical facilities and healthcare entities are remotely accessed by patients with connected things remotely on an intelligent platform.
4. Decision Support Systems

   Decision support systems have increasingly become popular in providing evidence-based clinical decisions to healthcare providers. They examine electronic health records and suggest reminders to healthcare staff, providing information related to treatment protocol. Decision support systems are widely used in IoT-based data analytics. It not only suggests treatment strategies but also helps in building models for identifying risks and treatments. With data analytics, it is faster and easier to recognize risks, narrowing down the diagnosis and prescribing the treatment. Remote care platform can also be built upon this. A customized and efficient strategy for treatment can be devised by healthcare personnel. Artificial intelligence and machine-learning-integrated IoT framework executes without manual intervention and reduces the cost significantly. IoT devices monitor physiological parameters. The clustering of the records provides an impactful parameter to draw inference. The action plan of early detection, diagnosis, risks identification and treatment could be smoothly executed with decision support systems. K-means clustering has been extensively used to create decision support system models.
5. Remote and Virtual Care

   It is another cost and time effective healthcare model lying on the foundations of IoT. It offers not only consultation at home scenarios but also automated test procedures through IoT devices and sensors. Wearable devices have sensors that provide intelligent analytics and assist in remote monitoring of patients. With the concept of IoT in healthcare, patients do not depend exclusively on hospitals. They do not need to visit clinical settings. Remote patient monitoring (RPM) has emerged as a technology and sub-domain of IoT infrastructure. It is a virtual healthcare model. Patients can access care and treatment at the comfort of their home or in a remote area. Thus, the delivery cost for healthcare services decreases. It aims to deliver services through telecommunications. It not only decreases the hospital costs but also signals early health deterioration. Super quality virtual healthcare is offered with continuous monitoring of patient's health progress. Many elderly patients stay remotely or away from their families. Upon any variation in their routine activities, an alert signal is sent to family members and concerned healthcare personnel. This is a benefit of IoT for remote healthcare. Physiological data obtained by wearable sensors help healthcare personnel in making medical decisions. It is also beneficial in passive medication and suggesting and monitoring healthy lifestyle. Remote healthcare has been proved accurate and efficient as conventional healthcare.

IoT-enabled healthcare platform and big data analytics establish an excellent communication collaboration link between patient and caregiver. Wearables, intelligent sensors and wireless technology promote the concept of distributed services in healthcare. Patients need not to be hospitalized for weeks or months. Patients are remotely monitored and virtual healthcare is provided with proactive treatment. Researchers have also observed that IoT and big-data-based healthcare services led to reduction in death toll of patients. Patients get freedom from being attached to heavy machines and syringe tubes. Portable devices offer ease of movement and comfort.

IoT Offers Following Advantages in Healthcare

(i) Reduced Cost

The real time monitoring is facilitated by IoT technology. This considerably reduces physical visits to clinics, OPD admissions and hospital stays.

(ii) Optimized Diagnosis and Treatment

Analytics help in pattern matching and decision-making. Real-time constant monitoring of patients helps in early diagnosis of disease, symptoms are identified and treatment plan is generated.

(iii) Proactive Treatment

IoT devices and sensors work 24×7 and offer proactive treatment in medical emergencies.

(iv) Management

Pharmacy, inventory, drugs and equipment management is challenging in the healthcare industry. IoT opens doors for effective management and efficient utilization of existing infrastructure.

(v) Reduced Errors

Healthcare operations are carried out smoothly and waste generation is reduced with IoT. This also leads to reduced errors in hospital services and creates a robust healthcare system.

(vi) Remote Access to Villages and Far Locations

Better healthcare services are in approach to remote towns and villages. IoT provides good healthcare to such areas upon integration with hospitals and governments.

(vii) Assistance in Health Insurance

Data collections, claims processing and fraud claims identification are carried out with transparency with device monitoring. IoT devices also justify whether a patient was aligned with treatment guidelines by monitoring recovery phase.

IoT methodology is proved optimal and proficient in the fight against COVID-19 pandemic. Its tools offer adequate facilities to cope with a situation like global coronavirus outbreak. The Internet of Healthcare Things (IoHT) and Internet of Medical Things (IoMT) have emerged as the latest tools to resolve the pandemic situation. IoT has helped government bodies and medical authorities to prepare guidelines for safety and to develop facilities of treatment. IoT laid a monitoring system to keep track of infected patients in quarantine. It made biometric diagnostics easier with the internet-based network. The disease is still on its growing curve, and IoT ensures effective virus control with enhanced diagnostics and proper treatment. With IoT methodology, there are error-free scenarios as machine learning models and wearable devices work with integration to generate analysis and patterns. Infected patients are assured effective treatment with low expenses and accuracy. This technology resulted in workload reduction from clinical workers with a significant improvement in their work efficiency. Remote monitoring of patients becomes easier with IoT. Lockdown in various nations across the globe has posed challenges in healthcare. Virtual follow-up and consultation of mild health issues have been encouraged. With the growing spread of COVID-19, there is increased utilization of interconnected networks. IoT deployment allows connected services to work with effective flow of data and enables continuous exchange of data. Identification of patient clusters is optimized with mobile application. Several innovative ways for better quality of care have been identified. Indian government has launched the Arogya Setu mobile centric application. It is a digital framework to capture patient information. It tracks and alerts interconnected devices for an efficient health management system.

IoT and big data together helped health systems with predictive analytics models for determining the impact of coronavirus, disease outcomes and to gain insights of COVID-19 risk. Majority of the population is still vulnerable to COVID-19 disease. There has been a strong economic impact globally due to the virus outbreak. The biological race of Coronavirus is similar to SARS-CoV, which was observed in 2002. Less social distancing and higher mobility increase the risk of virus spread. The hospitalization risk can be calculated by a predictive analytics model. Asymptomatic patients are less likely to be hospitalized. Patients with underlying lungs diseases, diabetes, hypertension, children under the age of 10 years and citizens over 60 years of age are prone to be caught by the virus. The future of the COVID-19 crisis depends on genome sequences of patients and their social interactions. Population spread over a region is explored for identifying the virus spread. The healthcare services are focussed on patient-centric, cost-saving, accessible and customized workflows.

6 Challenges

It is challenging for deep learning models to present 'how' a prediction happened. How it arrives at a recommendation is not specified by the deep learning system. Training a model for massive and heterogeneous data is challenging with sparse noisy data. There are considerations for ethical practices, legal implementation of tools, healthcare worker's understanding of machine learning tools and data security.

There are challenges associated with the use of machine learning in IoT and big data. In the view of F. Ahamed F. Farid, data transmission and data security remain the issues in personalized healthcare [14]. Clinical response, lifestyle and behaviour of patients are very sensitive data. Learning algorithms cannot judge the bias involved in data collection and interpretation. This leads to incorrect decision-making. Lack of availability of domain experts and scientists is a global issue in intelligent healthcare. The domain of many diseases and medical conditions is complex and not much explored. Strong and efficient interoperable training models for image and speech analysis are yet to be developed and learned. In the context of routine clinical research, big data poses challenges in data analysis and data security.

According to research [15], reproducibility remains an obstacle in the progress of intelligent healthcare. The interpretation of new findings and their validation validated in context of their uses and complexity is an emerging issue in personalized healthcare. Multiple hypothesis, random analysis and incomplete documentation is challenging in reproducibility. However, underlying data access is restricted, but data could be replicated to train a learning model. Connected devices brought along the advantage of real-time monitoring of patients and have saved the lives of millions of patients during medical emergencies. But it also brings huge challenges. Continuous monitoring under IoT devices collects large volumes of data. This also captures sensitive information. Data security is the major concern with IoT in healthcare. Data protocols and standards often fail to meet compliance with government policies and regulations. Data ownership is a critical issue.

Delays and bandwidth insufficiency are also faced by IoT devices. In many countries, there is a lack of backbone support infrastructure for IoT frameworks. Techniques for delay tolerance are not incorporated in the system. Untrained medical staff is also a challenge in information collection. Most of the data are collected from sensors and wearable devices. Hence, interfaces must be simple and easily accessible. If the medical staff are not properly trained, then the information cannot be efficiently utilized. The quality of service will also get compromised.

7 Solutions and Recommendations

Healthcare and medicine are a profession involving lots of information. They are an excellent field for studying and investigating data science to learn clinical research and examine findings. The programming framework of Hadoop is the first choice in the medical domain for learning, analytics, NLP, pattern finding and decision-making. Big data offers inexhaustible data sources to enterprises involved in medical research. IoT further leads to addition of massive data sets. Patients symptoms, functional behaviour, medical treatment, diagnostics, test reports, etc. emerge endlessly. Data quantity gives rise to security concerns. Safe measures need to be developed for resolving privacy issues. Data sharing runs in a continuous flow within a healthcare system. Therefore, a balanced model of data security is built to mitigate risks of cyber security and data theft.

Machine learning models in decision-making help to isolate unnecessary data and reduce study over redundant data. For example, the symptoms that adversely affect the patient are important. General symptoms that are not connected to complex health conditions are not executed during training of machine learning models. Predictive analytics plays a vital role in this scenario. It anticipates prediction for data that might raise a patient's risk by gaining insights from data. It also controls false predictions through advanced data analytics by revealing similarities and patterns. Image diagnostics and object classification is the excellent utility of machine learning models. It rapidly detects patterns through automated routines. This boosts the performance of an intelligent healthcare system through improved health services to patients and interactive automated processes.

Hospital expenses are skyrocketing, and medical professionals aim to develop innovative healthcare services that would communicate health-related issues among patients and doctors, report personalized insights by decision-makers, improve patient experiences and reduce treatment costs. The incredible potential of machine learning along with big data and IoT will harness the real value of artificial intelligence, sensor devices and predictive analytics. Deep learning, artificial intelligence and predictive analytics tools are helping early diagnosis of several infant diseases, autism, cancer, cardiac illness and heart failure diagnosis.

8 Future Work

The strength of big data will help to develop ready solutions with low costs for health practitioners in intelligent healthcare. The data obtained from electronic health records, sensors and wearable devices are the potential information in healthcare. The research practices will boost and will be carried at a fast pace. Diagnostics and predictive data analytics will meet the needs of healthcare workers, patients, clinicians, drug discovery and equipment designing in future. Machine learning in IoT and big data will contribute to build an intelligent healthcare system with continuously emerging patient records and experiences.

IoT framework and big data technologies give entries to heterogeneous data. Therefore, data with multiple dimensions arrive unexpectedly in the healthcare system. This brings challenges in data pre-processing. Advanced machine algorithms are needed to understand patient care factors. Patient's recovery after treatment, biological response and physical behaviour accelerate learning algorithms in generating predictions. Deep learning opens the door for a predictive healthcare framework and will support health practitioners in daily routine tasks.

Machine learning models can make predictions of required assistance for patients but also defer the decision of the human clinicians, based upon the human expert's experience and availability. Some situations pose obstacles with artificial intelligence tools. When an AI setting works with a new healthcare expert, it also needs to learn the expert's behaviour and other judgement parameters. Trust and biased instinct of experts become an issue to artificial tools.

Big data and its analytics tool will prove to be a powerful technology for improved population health management. The provision of comprehensive and holistic care to patients is defined under the principles of population health management. EHR, insurance claims, notes and records provide sufficient information for patient management. Big data strategies will help to build and improve the plan for population health management. Future trends are understood by analysing gender and age and living conditions, history of illness is reported and machine learning models implement such data at a wider scale. The socio-economic data are also a determinant to make predictions and define treatment plans.

9 Conclusion

Medical practitioners need to upskill and learn new tools developed from learning models and research in data analytics. It is also essential that they are trained to understand knowledge discovery and reports derived from new approaches. Nations must define a substantial budget and funding policies to enhance research in machine learning and to integrate expert systems in healthcare. Current era is the most promising time for promoting medical science to strengthen the patient care system and to gain a remarkable growth in intelligent healthcare.

The collaboration of machine learning with IoT and big data is the best opportunity to ensure qualitative patient care. Patients and healthcare providers are the major role players in a healthcare system. A technologically efficient system of healthcare ensures safety in diagnostics, drives error-free treatment plans and enhances value of the medicine science system. Speed of the automated computer-driven system is the fundamental factor to determine abilities of a healthcare framework. Efficiency is the second dimension. The accuracy of data analytics is examined during decision-making. Next important dimension is the quality of data. Big data floods healthcare systems with massive and complex data repositories. Data pre-processing models must intensify their proficiency to produce quality data and knowledge to aid prediction and pattern recognition.

Recent COVID-19 crisis has brought huge and immeasurable trouble to human society and especially to the economy. The disease was still observed on a growing curve until August 2020. There is a prediction that positive cases will further arise and the death toll might increase substantially. The pandemic caused great challenges and serious problems in healthcare services. Machine learning models have been proved trusted and reliable so far. IoT and big data technology helped the medical domain to provide personalized care to patients, and resource monitoring is carried with ease. However, many times, uncertainties also arise. The learning algorithms basically rely on sensor data. These data are used to recognize patterns in behaviour of the patient. Each and every recorded activity from a patient's routine is crucial in determining patterns. In this context, there is often a skewed decision because the patterns could judge sensitive data incorrect. Collection of data and its interpretation can be biased without human knowledge.

References

1. A.H. Sodhro, A.S. Malokani, G.H. Sodhro, An adaptive QoS computation for medical data processing in intelligent healthcare applications. Neural Comput. Applic. **32**, 723–734 (2020)
2. B.T. Idemen, E. Sezer, M.O. Unalir, A new generation architecture proposal for intelligent healthcare medical laboratories, in *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions INFUS*, (Springer, Cham, 2020), pp. 1284–1291
3. S. Yousefi, F. Derakhshan, H. Karimipour, Applications of big data analytics and machine learning in the internet of things, in *Handbook of Big Data Privacy*, (Springer, Cham, 2020)
4. R. Xie, I. Khalil, S. Badsha, M. Atiquzzaman, An intelligent healthcare system with data priority based on multi vital biosignals. Comput. Methods Prog. Biomed. **185**, 105126 (2020)
5. K.R. Dalal, Analysing the implementation of machine learning in healthcare, in *International Conference on Electronics and Sustainable Communication Systems (ICESC)*, (IEEE, Coimbatore, 2020)
6. C. Venkatesan, P. Karthigaikumar, A. Paul, S. Satheeskumaran, R. Kumar, ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications, in *IEEE Access*, (IEEE, 2018), pp. 9767–9773
7. B.A. Patel, A. Parikh, Impact analysis of the complete blood count parameter using Naive Bayes, in *International Conference on Inventive Computation Technologies (ICICT)*, (IEEE, Coimbatore, 2020)
8. N. Ravindran, O. Shery, A. Samraj, N. Maheswari, Stable and crit gesticulation recognition in children and pregnant women by Naïve Bayes classification, in *International Conference on Current Trends in Information Technology (CTIT)*, (IEEE, Dubai, 2013), pp. 259–264
9. Y.P. Sinha, P. Malviya, M. Panda, S.M. Ali, Contextual care protocol using neural networks and decision trees, in *Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, (IEEE, Bangalore, 2018)
10. T. Xie, R. Li, X. Zhang, B. Zhou, Z. Wang, Research on heartbeat classification algorithm based on CART decision tree, in *8th International Symposium on Next Generation Electronics (ISNE)*, (IEEE, Zhengzhou, 2019)
11. M. Ambigavathi, D. Sridharan, Big data analytics in healthcare, in *Tenth International Conference on Advanced Computing (ICoAC)*, (IEEE, Chennai, 2018)
12. R. Vaishya, M. Javaid, I.H. Khan, A. Haleem, *Artificial Intelligence (AI) Applications for COVID-19 Pandemic* (Elsevier, Cham, 2019)
13. M. Bansal, B. Gandhi, IoT & big data in smart healthcare (ECG monitoring), in *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects*, (IEEE, Faridabad, 2019)
14. F. Ahamed, F. Farid, Applying internet of things and machine-learning for personalized healthcare: Issues and challenges, in *International Conference on Machine Learning and Data Engineering (iCMLDE)*, (IEEE, Sydney, 2018)

15. A.L. Beam, A.K. Manrai, M. Ghassemi, Challenges to the reproducibility of machine learning models in health care. JAMA **323**(4), 305–306 (2020)
16. S.E. Dilsizian, E.L. Siegel, Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr. Cardiol. Rep. **16**(1), 441. Published online, IJAER (2018)
17. S. Anand, S.K. Rautray, Issues and challenges in healthcare narrowband IoT, in *ICICCT*, (IEEE, Coimbatore, 2017). https://doi.org/10.1109/icicct.2017.7975247

Modelling Covid-19: Transmission Dynamics Using Machine Learning Techniques

Vijay Kumar, Jyoti Chawla, Rajeev Kumar, and Arti Saxena

1 Background

Pandemic disease Covid-19 is affecting the day-to-day life of people across the globe from the past few months. Scientists across various disciplines ranging from molecular biology to applied mathematics have teamed up for the assessment and control of this rapidly spreading virus. Mathematical models play a significant role in assessing, predicting and proposing potential outbreaks [1]. Mathematical models in pandemic of diseases have been recognized as an effective tool in analysing the propagation of infectious diseases and to test the complex dynamics of diseases to propose test strategies [2–4]. Epidemic coronavirus (Covid-19) transmission rate is found to be different in different regions and that may be due to different factors such as climate change, movement of individuals from one region to another, population density, different types of immune system, population pyramid, and antibiotics resistance. Mathematical modelling may can predict the disease transmission and mortality with recognition of the possible reasons of transmission of disease. It may also analyse the effect of intervention strategies for optimal control of transmission rate [5]. Many researchers have implemented the model on the situations for control of infectious diseases [6–8]. As per the data available till date, the transmission rate of Covid-19 virus is seen to be different in different countries. It has been observed that Covid-19 outcome trends in terms of number of infected individuals depend on various factors [9, 10]. Even with all healthcare facilities on place, the challenge is to decrease the spread and doubling time of diseases. It is very important that optimal interventions should be on place as per the severity of problem.

V. Kumar (✉) · J. Chawla · R. Kumar · A. Saxena
Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_6

2 Objective

In this chapter, we have proposed a framework which employs machine learning to study the transmission of Covid-19. The objective of study is to highlight the transmission dynamics of the virus and monitor the transmission among top five countries with highest number of infected persons as on May 31, 2020 and predict the situation further. Linear regression techniques have been used for the purpose of analysis and prediction.

3 Methodology

Python has been used as the main programming language for analysis, and forecasting. Data have been taken from a reliable source (https://www.kaggle.com/imdevskp/corona-virus-report) [11]. Data have been pre-processed and recorded from 1 January 2020 to 31 May 2020. Experimental results have been illustrated by means of graphs and tables since the inception of the disease in the country. The fitting of the model is assessed by means of R^2 statistics and residual.

4 Results

The most affected 5 countries Brazil, Russia, Spain, the UK and the USA have been considered. After removing zero values, data were considered from 21 January 2020 for the purpose of analysis. The shape of data set is (5, 133), that is the profiling of number of active cases in 5 countries and 133 days. Profile of all countries and their comparison is shown in Figs. 1a–e and 2.

To calculate a good measure with 132 days, we have analysed the spread in an interval of 33 days as depicted in Figs. 3, 4, 5, 6 and 7 for all countries. Figure 3 represents the spread of Covid-19 from day 1 to 133 days in Brazil. From Fig. 3, it can be seen that from day 1 to day 45, there was no case of Covid-19 and it was increased from day 45 till the end of study (Fig. 3a–d). Figure 4 represents the spread of Covid-19 from day 1 to day 133 in Russia. From Fig. 4, it can be seen that from day 1 to day 11, there was no case of Covid-19, but after day 11, there was a drastic change in the number of cases and increased regularly till the end of the study (Fig. 4a–d). Figure 5 represents the spread of Covid-19 from day 1 to day 133 in Spain. From Fig. 5, it can be seen that from day 1 to day 11, there was no case of Covid-19, but after day 11, there was also a drastic change in the number of cases and increased regularly till the end of the study (Fig. 5a–d). Figure 6 represents the spread of Covid-19 from day 1 to day 133 in the UK. From Fig. 6, it can be seen that from day 1 to day 10, there was no case of Covid-19, but after day 10, there was a

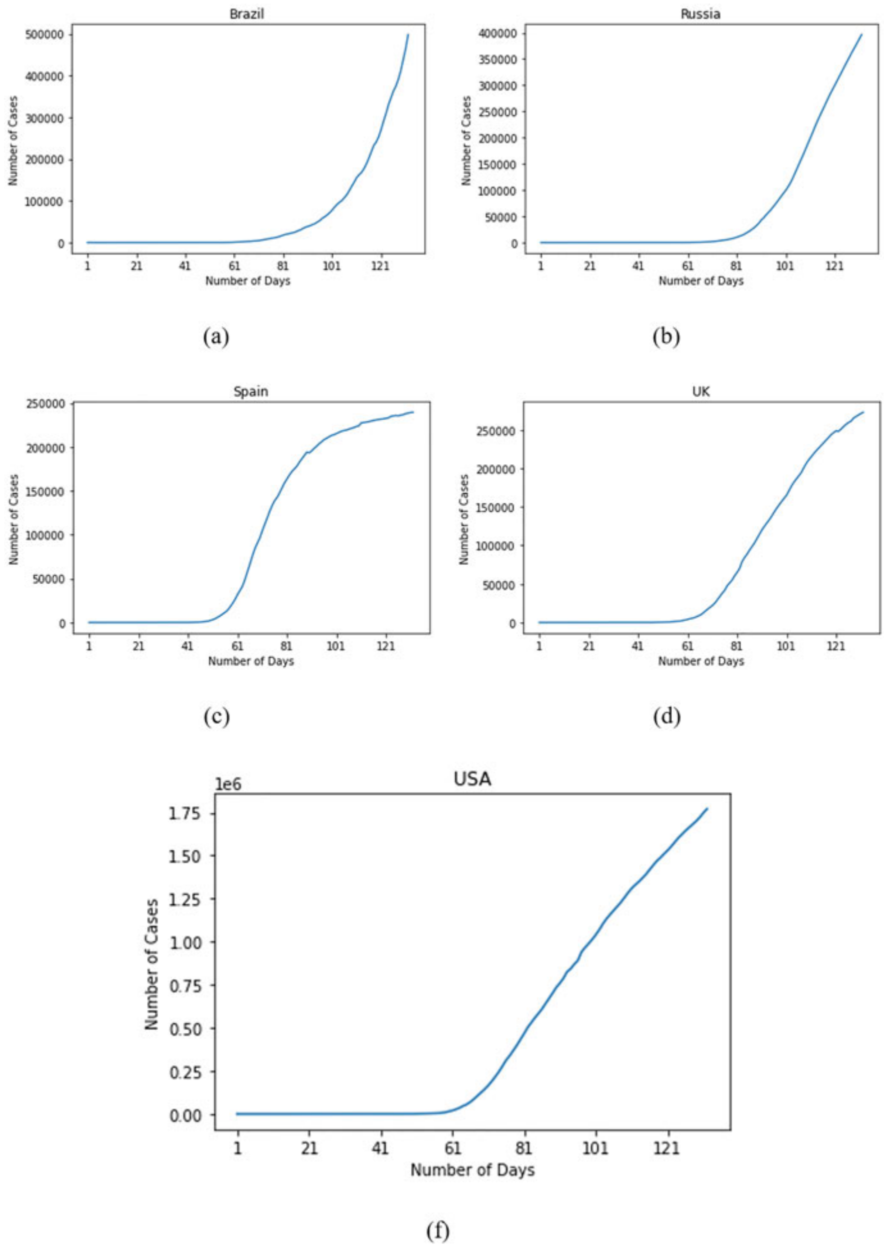


Fig. 1 Profile of different countries: (**a**) Brazil, (**b**) Russia, (**c**) Spain, (**d**) UK and (**e**) USA

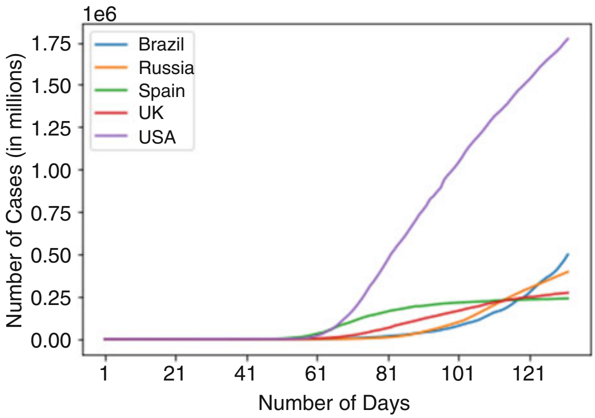


Fig. 2 Profile trends of all selected five countries from day one to last day

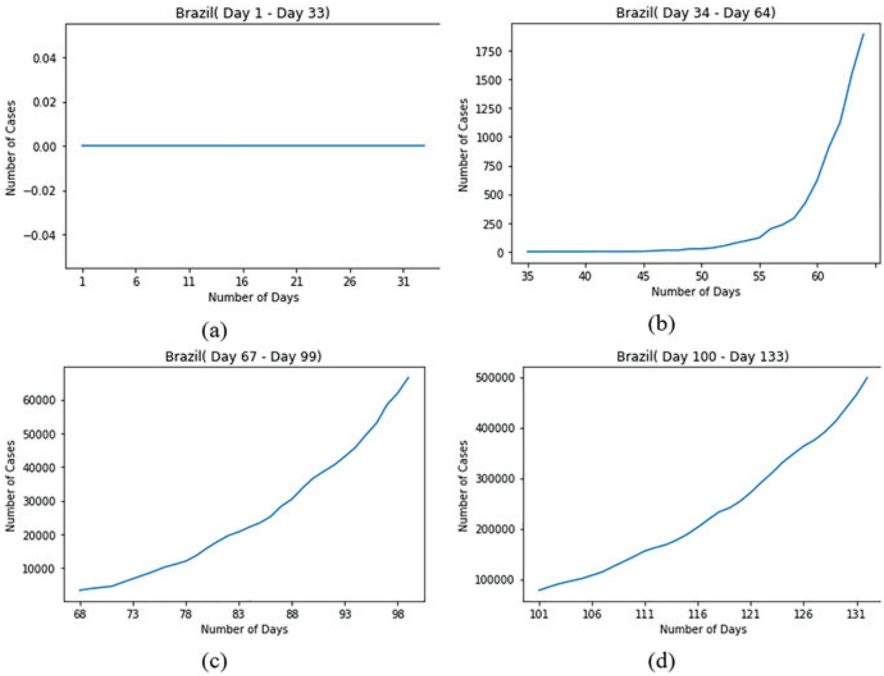


Fig. 3 Spread graph of Covid-19 from day one to last day in Brazil

change in the number of cases and increased regularly till the end of the study (Fig. 6a–d).

Out of all these countries, the number of cases in the USA was totally different. Figure 7 represents the spread of Covid-19 from day 1 to day 133 in the USA. From Fig. 7, it can be seen that from day 1, cases were spread and increased day by day (Fig. 7a–d).

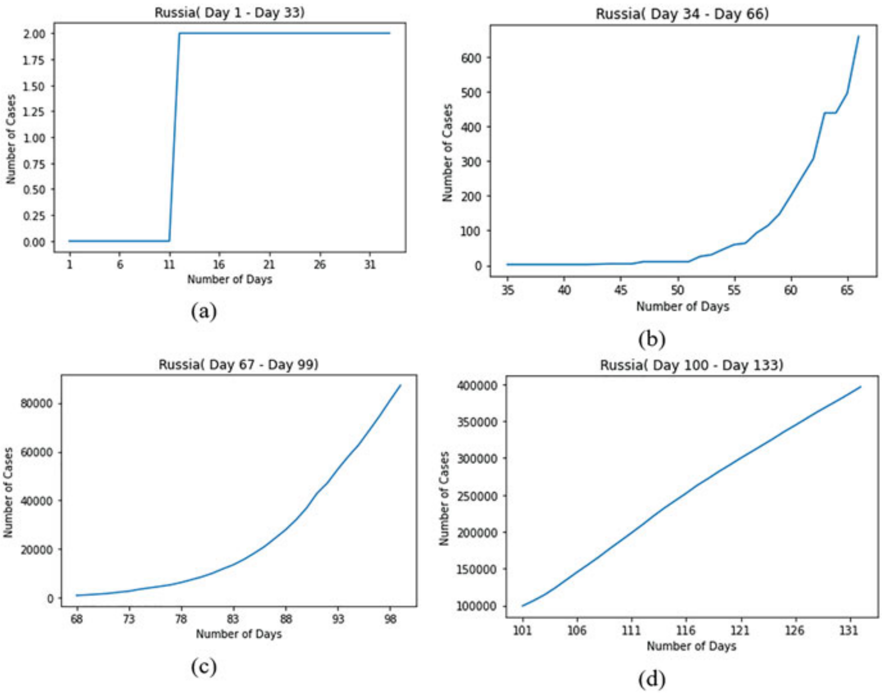


Fig. 4 Spread graph of Covid-19 from day 1 to day 133 in Russia

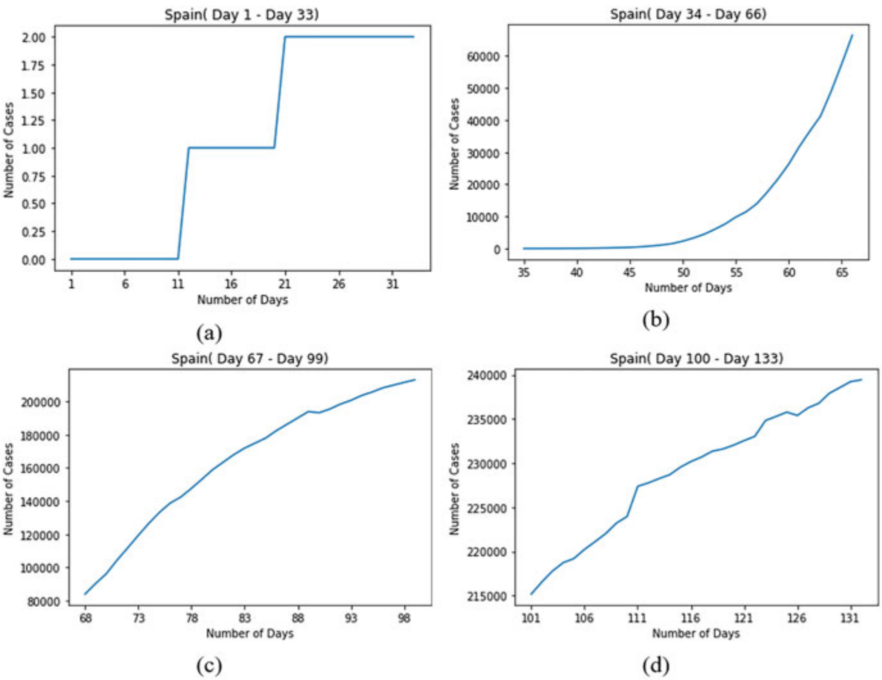


Fig. 5 Spread graph of Covid-19 from day 1 to day 133 in Spain

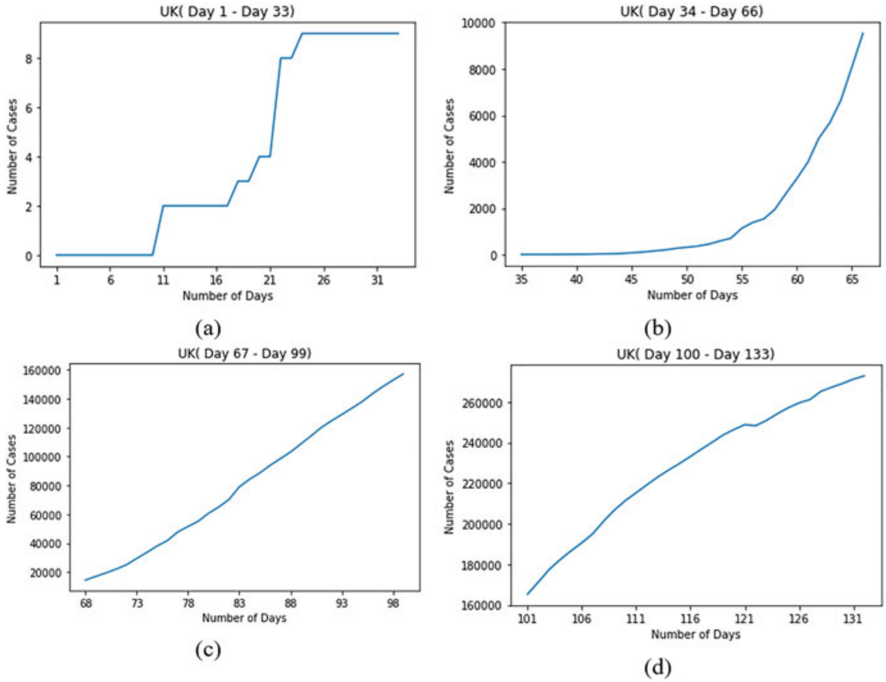


Fig. 6 Spread graph of Covid-19 from day 1 to day 133 in UK

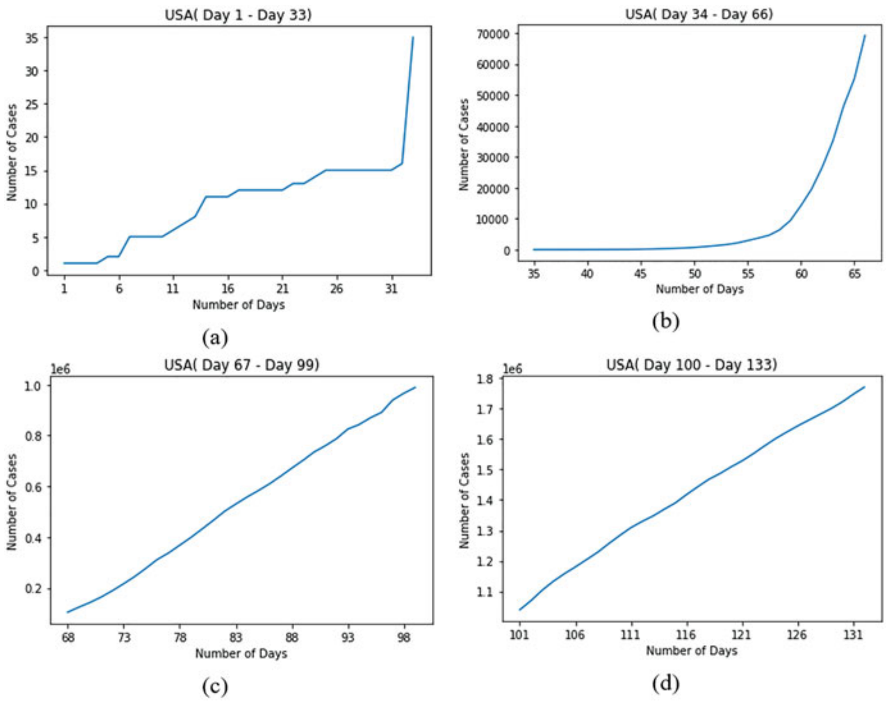
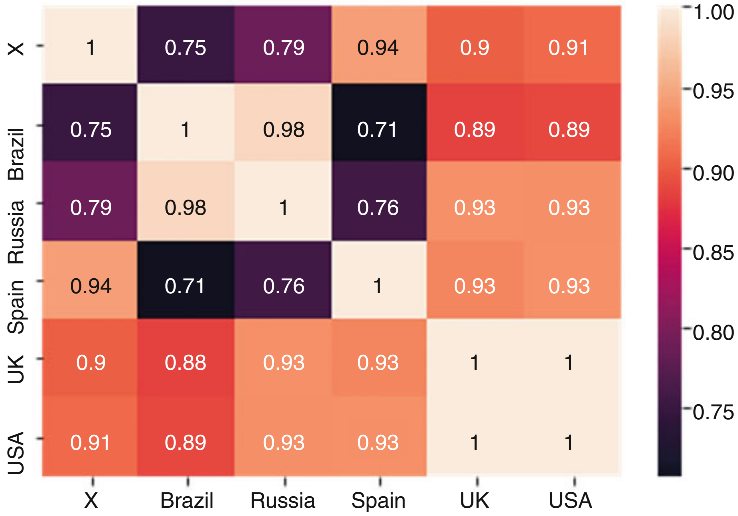


Fig. 7 Spread graph of Covid-19 from day 1 to day 133 in USA

First derivative curve was also analysed for all the countries (Fig. 8). First derivative curves of different countries show the rate changes of the number of cases with respect to the days.

Maximum infection rates calculated are 33274, 11656, 9181, 8719 and 48529 for Brazil, Russia, Spain, the UK and the USA, respectively. Scatter Plot and Correlation matrix are shown below:



R square values for linear regression are 0.680, 0.694, 0.869, 0.824 and 0.831 for Brazil, Russia, Spain, the UK and the USA, respectively. Then, a hybrid model has been proposed for the prediction of accurate daily cases. First of all, original data values in terms of model no. 1 were fitted and then re-fitting was done on the resultant values by model no. 2 to improve the R^2 value.

Mathematical models 1 and 2 were applied to get the final results.

Mathematical Model 1:
The nature of data is exponential, the proposed model$y = a\ b^x$

Its normal equations are:

$$A\ n + B\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} Y_i$$

$$A\sum_{i=1}^{n} x_i + B\sum_{i=1}^{n} {x_i}^2 = \sum_{i=1}^{n} x_i Y_i$$

Where, $Y = \log_{10}(y)$, $a = 10^A$ and $b = 10^B$

From this model, the fitted equation of the countries is shown in Table 1:

Table 1 Fitted equations from model 1

Countries	$y = a\ b^x$
Brazil	$y = (33.3892)\ (1.0908506)^x$
Russia	$y = (51.28389622)\ (1.088954485)^x$
Spain	$y = (591.0982511)\ (1.012786858)^x$
UK	$y = (98.10369544)\ (1.055804358)^x$
USA	$y = (1.662316149)\ (1.136061032)^x$

Table 2 Fitted equations from model 2

Countries	$y = A + B\ x + C\ x^2 + D\ x^3$
Brazil	$y = -3911 + 747\ x - 29.71\ x^2 + 0.333\ x^3$
Russia	$y = -2755 + 575.9x - 24.73\ x^2 + 0.307\ x^3$
Spain	$y = 590.2 + 7.692\ x + 0.038\ x^2 + 0.00\ x^3$
UK	$y = -164.7 + 58.95\ x - 2.242\ x^2 + 0.033\ x^3$
USA	$y = -6122 + 1157\ x - 45.90\ x^2 + 0.488\ x^3$

Mathematical Model 2:
The proposed model

$$y = A + B\ x + C\ x^2 + D\ x^3$$

Its normal equations are:

$$A\ n\ +\ B\sum_{i=1}^{n} x_i\ + C\sum_{i=1}^{n} {x_i}^2 + D\sum_{i=1}^{n} {x_i}^3 = \sum_{i=1}^{n} y_i$$

$$A\sum_{i=1}^{n} x_i\ + B\sum_{i=1}^{n} {x_i}^2 + C\sum_{i=1}^{n} {x_i}^3 + D\sum_{i=1}^{n} {x_i}^4 = \sum_{i=1}^{n} x_i\, y_i$$

$$A\sum_{i=1}^{n} {x_i}^2 + B\sum_{i=1}^{n} {x_i}^3 + C\sum_{i=1}^{n} {x_i}^4 + D\sum_{i=1}^{n} {x_i}^5 = \sum_{i=1}^{n} {x_i}^2 y_i$$

$$A\sum_{i=1}^{n} {x_i}^3 + B\sum_{i=1}^{n} {x_i}^4 + C\sum_{i=1}^{n} {x_i}^5 + D\sum_{i=1}^{n} {x_i}^6 = \sum_{i=1}^{n} {x_i}^3 y_i$$

From this model, the fitted equation of the countries is given in (Table 2):

After using the models, the improved R square values are 0.983, 0.988, 1.00, 0.99 and 0.947 for Brazil, Russia, Spain, the UK and the USA, respectively.

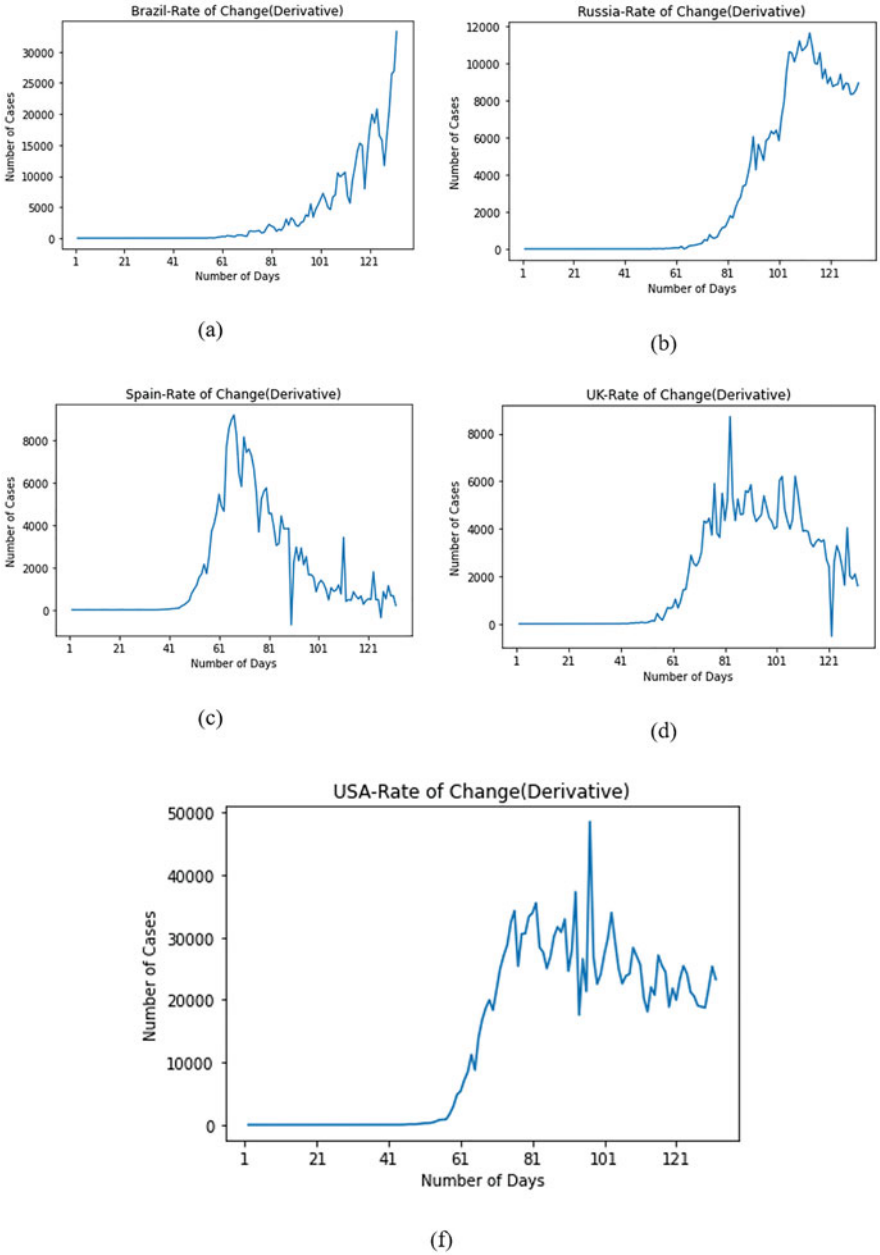


Fig. 8 First derivative curve: (**a**) Brazil, (**b**) Russia, (**c**) Spain, (**d**) UK and (**e**) USA

5 Conclusion

The prediction model obtained is based on the trend of the data with highest R^2 value and minimum residual. The model helps the authorities to make necessary arrangements during the emergency.

References

1. B. Ivorra, B. Martínez-López, J.M. Sánchez-Vizcaíno, Á.M. Ramos, Mathematical formulation and validation of the Be-FAST model for classical swine fever virus spread between and within farms. *Ann. Oper. Res.***219**(1), 25–47 (2014)
2. F. Brauer, C. Castillo-Chavez, C. Castillo-Chavez, *Mathematical models in population biology and epidemiology*, vol 2 (Springer, New York, 2012), p. 508
3. C.I. Siettos, L. Russo, Mathematical modeling of infectious disease dynamics. *Virulence***4**(4), 295–306 (2013)
4. C.E. Walters, M.M. Meslé, I.M. Hall, Modelling the global spread of diseases: A review of current practice and capability. *Epidemics***25**, 1–8 (2018)
5. M.J. Keeling, L. Danon, Mathematical modelling of infectious diseases. *Br. Med. Bull.***92**(1), 33 (2009)
6. Z. Feng, Final and peak epidemic sizes for SEIR models with quarantine and isolation. *Math. Biosci. Eng.***4**(4), 675 (2007)
7. J. Li, N. Cui, Dynamic analysis of an SEIR model with distinct incidence for exposed and infectives. *Sci. World J.***2013**, 1 (2013)
8. M.A. Safi, M. Imran, A.B. Gumel, Threshold dynamics of a non-autonomous SEIRS model with quarantine and isolation. *Theory Biosci.***131**(1), 19–30 (2012)
9. M. Coccia, Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Sci. Total Environ.***729**, 138474 (2020)
10. L. Li, Z. Yang, Z. Dang, C. Meng, J. Huang, H. Meng, et al., Propagation analysis and prediction of the COVID-19. *Infect. Dis. Model.***5**, 282–292 (2020)
11. https://www.kaggle.com/imdevskp/corona-virus-report

An Innovative Pandemic Knowledgebase Using Machine Learning

Ajanta Das, Sameya Ashraf, and Shubam Prasad

1 Introduction

In the current situation, the pandemic COVID-19 is the worst thing that can happen to humans. The long-lasting corona virus and its exponential spreading nature have brought more devastation worldwide. The long-lasting effects may vary from one country to another. However, the affected cases are increasing day by day with a series of deaths that have been observed for every country, and the pattern is the same everywhere. Obviously, people are also recovering and maybe the recovered numbers are higher than death counts, but it is very hard to predict when the spike of death and affected cases will decline sharply. Lots of precautions and prevention activities such as washing hands with soap for 20s rigorously, sanitizing all the surroundings frequently, wearing face mask, and keeping social distancing are notable. In the twenty-first century, living in a nuclear family is increasing gradually. Due to the very small size (15–100 nm) of virus, it is doubtful whether the face mask can prevent corona virus inhaling. So, people can only protect themselves by restricting their outside activities, and confine and seal themselves at home safely. Once again, it is not possible for human beings to be safe at home day after day and month after month. As a result, technology-driven and social media-based living is prioritized. Meeting people, friends, and near and dear ones on online platform gives some kind of relief from living alone or living with own family members for long within a residential flat. With the restriction on going out from home, helping hands for household jobs has also stopped as a prevention measure. Simultaneously, with equal opportunity of working and more responsibilities at home, working women are under much pressure, including those who are working for home and those working from home digitally.

A. Das (✉) · S. Ashraf · S. Prasad
Amity Institute of Information Technology, Amity University, West Bengal, Kolkata, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_7

In order to sustain the job, leading life in connectivity with various electronic devices is mandatory. As a result, exponential growth of data has arisen from various smart devices. Hence, data analysis and representation of data are necessary to prevent the spread, and are essential to get back to normal life scenarios. Data science and machine learning techniques and algorithms may need to be used to represent or interpret the data properly for further protection and awareness of general people [1]. Discovery of knowledge and decision-making information extracted from social media play a vital role. Analysis of previous epidemic and pandemic cases is presented in Sect. 2. Section 3 presents a brief study about the present pandemic situation and sustainability issues for existing healthcare policies. This chapter presents the importance of data analytics and knowledgebase in Sect. 4. Furthermore, this chapter presents two separate case studies for international and national scenarios, both with respect to the pattern of spreading of the corona virus in Sect. 5. Next, Sect. 6 proposes a knowledgebase based on the available data and then presents a rule-based learning in healthcare to handle corona virus categorically in the current pandemic situation. Finally, the chapter concludes with Sect. 7.

2 Historical Analysis of Epidemic and Pandemic

This section presents the analysis of previous epidemic or pandemic cases preceding COVID-19 cases since 1918. This historical analysis is presented based on the knowledge gathered from previous cases, such as, when it occurs, how it spreads, what is the rate of spreading, whether any symptoms are present or not, what is the lifetime of the virus, and how it is controlled or how people have protected themselves from the infection of the virus. With reference to the responses to all these abovementioned questions, analyses are presented for five virus types, namely Spanish Flu, Asian Flu, Swine Flu, West African Ebola, and Zika virus, which are as follows:

(a) S*panish Flu*: This is the oldest devastating pandemic, found from the archives or history. It is found from historical data that 500 million people were infected worldwide during 1918–1920 [2]. The severity of pandemic was highest globally, but the recovery time was 1–2 weeks. Infections spread with influenza symptoms, while very less cases were asymptotic. It was airborne and the virus life time was 2–4 weeks. People used to wear masks and sheltered themselves in their homes as a protective measure. As a consequence, offices, schools, and businesses were closed for a long time to fight the pandemic, as the fatality rate was 3%.
(b) *Asian Flu*: From east Asia, a new virus was first reported in Singapore in 1957. The pandemic was named Asian flu, and it had a high death rate. From the record, around 1.1 million deaths occurred worldwide. The virus spread both ways with and without symptoms. Compared to Spanish flu, the fatality rate

was very less, <0.2%. The lifetime of the virus depended on the severity of the infection, on average, 14–21 days and the recovery time was normally 2–4 days.

(c) *Swine Flu*: After more than 50 years, in 2009, pandemic by influenza A (H1N1) virus was reported. It started in the United States and quickly spread across the world. The virus spread asymptotically 90% with very less symptoms. However, the fatality rate was very less, <0.03%. The lifetime of the virus was a maximum of 10 days and the recovery time was normally 3–7 days.

(d) *Ebola Epidemic*: The epidemic spread during 2013–2016 in western Africa with the outbreak of Ebola virus disease (EVD). The most affected and disrupted places were Guinea, Liberia, and Sierra Leone. It was neither seasonal nor airborne. It spread through blood or body fluid of the infected person. Hence, it did not spread widely, but the fatality rate was very high, 50%. According to the report, 28,616 people were infected, while the number of deaths was 11,310 [3].

(e) *Zika Virus Epidemic*: Simultaneously with Ebola epidemic in Brazil, during 2015–2016, Zika fever spread in the United States [4]. A few cases of the same virus spread were reported in Asian and Pacific island regions. After a year of spreading cases and based on analysis, it was named Zika epidemic in November 2016 by the World Health Organization. Zika epidemic spread mostly with symptoms like normal influenza, joint pain, and conjunctivitis [5]. Although it had not taken the shape of pandemic, the rate of fatality was maximum compared to other cases, 8.3%, due to the virus life span of 2–42 days.

3 COVID – 19: Pandemic

The virus is invisible in nature but can be seen through microscope. The virus looks like a “crown” and hence the name corona, as the meaning of corona is crown. The crown is surrounded by protein which causes the speedy transmission into human beings. Till date, it is suspected that the origin of this outbreak of coronavirus is Wuhan, China, in December 2019. The infection of the virus initiated with common cold, leading to severe respiratory syndromes and mortality, and spreading throughout the globe. This most devastating disease is famous with the name COVID-19.

Often, people get confused between outbreak, epidemic, and pandemic. In case of outbreak, the spreading of any infectious disease is within a community. While this outbreak lasts for a few weeks and goes beyond control within that community or location, it is well known as epidemic. Lastly, when epidemic travels throughout the world, it is identified as pandemic. Revisiting the historical data, an example of epidemic is *Ebola Virus,* while *Spanish flu* is identified as pandemic. Since none of us are relieved from corona virus globally, COVID-19 is a pandemic affecting all countries.

The virus is usually active from 2 to 14 days. However, the removal of effect from the virus varies between 30 and 49 days, depending on the severity of the cases. These symptoms are usually mild and increase gradually. Lack of attention to these symbols may lead to severe complications followed by fatality. The virus is spreading asymptomatically also. At the outset, people need to remember that "prevention is always better than cure" and hence most effective protection and prevention ways for COVID-19 are as follows:

(a) Clean the hands frequently and thoroughly.
(b) Wear face mask whenever necessary, if possible, always.
(c) Avoid touching eyes, mouth, and nose.
(d) Cover while coughing with the bend of elbow or with a tissue.
(e) Maintain a distance of at least 1 m from others.

The next section presents the machine learning techniques for preparing the knowledgebase.

4 Data Analytics and Knowledgebase

Data analytics reviews raw data and reexamines the datasets to present the pattern of the raw dataset. Data analysis is the basis of data analytics. Analysis of the same raw datasets needs to be reiterated before presenting analytics. Now machine learning is a method of data analysis which identifies data pattern and performs decision-making process without human intervention. Well-known machine learning algorithms include supervised and unsupervised clustering or regression to automate data analytics through mechanical processes. The main objective of data analytics is to optimize the performance; more specifically, optimization of business performance is done based on data analytics.

Knowledge discovery plays an important role in extracting knowledge from fuzzy datasets and presenting the meaningful pattern from all these fuzzy datasets [1]. Various techniques like *classification*, *clustering*, *dimensionality reduction*, and *collaborative filtering* for processing these datasets are used to perform predictive analytics. Classification technique or algorithm works for well-known groups. In case of supervised technique, learning is mandatory based on training data and then predicting the model for categorization of new data. Unlike classification, clustering techniques are applicable for unknown datasets and the grouping or categorization of data is prepared based on similarity index or degree of similarity. So clustering technique is iterative. Next, dimensionality reduction is applicable for reducing various dimensions or eliminating non-desired dimensions through feature selection and feature extraction, as each feature is not important for decision-making or prediction. Last, collaborative filtering gathers preferences from users' similar interests and then generates personalized recommendations.

Hence, machine learning explores useful knowledge by applying all these above-mentioned techniques, then matches the extracted information or knowledge with historical information or existing patterns in the datasets. Usually, the existing patterns help identify the new patterns and new patterns need to be validated before integrating with existing knowledge database. Machine learning technique, specifically decision tree, can be easily applied to COVID-19 dataset for the preparation of knowledgebase and historical information. The datasets for asymptomatic or based on some symptoms are not binary, rather fuzzy. As we know, all influenza types with cough and sneezing patients are not identified as COVID-19 patients, so validation of this data is very important and specific knowledgebase can be ready for better awareness. Hence, extraction of the relevant information can be integrated with existing pandemic knowledgebase for future usage.

5 Novel Case Studies

In the twenty-first century, people are sealed inside their residential premises only as preventive strategy against COVID-19. Most of the countries around the globe somehow have been infected by this virus. We are trying to keep a record or analysis of the spread of COVID-19 around the world. Because of the ongoing pattern and multiple characteristics of the corona virus, the dataset is changing every day. The data for analysis are collected from various media sources, such as daily newspapers, bulletins, or press conferences on television, and mostly from social media [6, 7].

This section presents two different case studies for international and national scenarios to present the ratio between number of deaths and number of active cases between number of deaths and number recovered, wherever possible.

International

The virus spreading started with China, Spain, France, and Italy gradually. As the virus spread in India since late March 2020, to prepare the knowledgebase for COVID-19, it is important to study the data pattern for international countries first. This section collects data for 10 different countries like Australia, Canada, China, Italy, France, Germany, Spain, the United Kingdom (U.K.), the United States, and India for the last 6 months from February 2020 to July 2020. Table 1 presents the data for the last 6 months of the total infected cases of the above-mentioned countries.

Next Table 2 presents active cases and deaths with respect to the total infected cases during February 2020–July 2020.

Figure 1 presents the death percentages with respect to the total cases from February 2020 to July 2020. The chart also presents comparison between all the above-mentioned countries. Regarding death cases, France, Italy, and United Kingdom are showing devastation. Compared to that, results for Germany and the United States show little control over death cases. Although in case of India and

Table 1 Infected total cases during February 2020–July 2020

Country	February	March	April	May	June	July
Australia	25	4738	1991	441	641	9069
Canada	20	8592	44624	37711	13257	12108
China	68033	1730	1308	139	530	761
Italy	1128	104664	97799	29388	7599	6959
France	100	52028	77453	22172	13048	23118
Germany	79	71729	91201	20485	12338	14833
Spain	58	95865	143417	47169	9842	39251
U.K.	23	22769	132359	93774	34328	19928
U.S.A.	63	194051	905991	754843	873908	1977033
India	3	1394	33466	155746	395183	1111262

Table 2 Infected active cases and deaths during February 2020–July 2020

Country	Total cases	Active cases	Deaths
Australia	16905	6726	197
Canada	116312	6510	8935
China	72501	684	4634
Italy	247537	12422	35141
France	187919	76154	30265
Germany	210665	9141	9224
Spain	335602	Not available	28445
U.K.	303181	Not available	46119
U.S.A.	4705889	2222756	156747
India	1697054	564856	36551

Australia death cases are very less, the corona virus started spreading late compared to other countries.

It reflects from both Tables 1 and 2 that the spreading of the virus will decline only after reaching its peak and it is very hard to predict when the peak time will arrive.

Next, the chapter presents the mobility view of Corona virus throughout the world. The daily or monthly cases for various country reflect different patterns. So, it is very difficult to arrive at a conclusion as to how and when the virus will spread around the countries or stay within the same country. A study or analysis is carried out for six different countries like China, Italy, France, Germany, Australia, and United States of America. The graph represents the number of total COVID cases for the month, specifically for the week; then out of these total cases, number of deaths or fatality, number recovered, and finally, the active cases are presented for all these above-mentioned six countries in Figs. 2, 3, 4, 5, 6, and 7 respectively. The pattern reflects that the virus started spreading in China first and its peak season occurred in February; then it spread mostly to Italy, France, Germany, and in Australia during the first week of March, it started spreading, but within a span of 2–3 weeks, it rose to a peak and within one month, the fatality rate reached the highest. At the outset, in the United States, the virus started spreading during

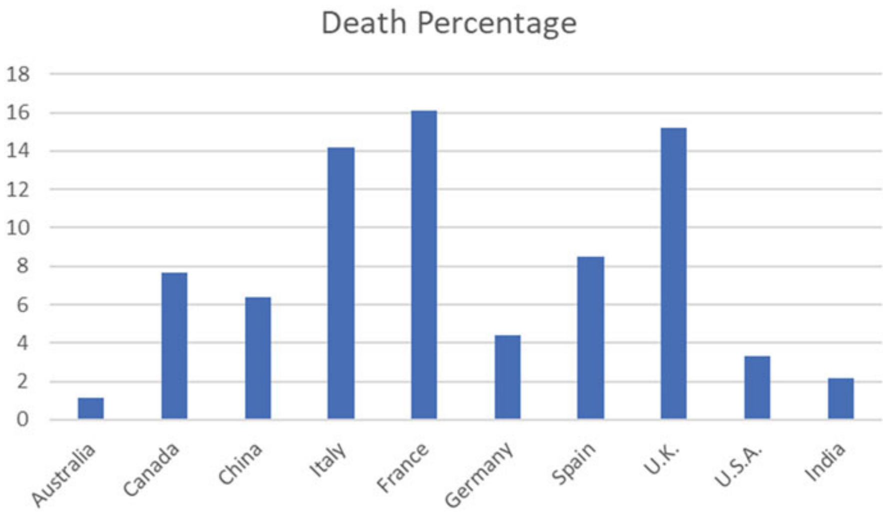


Fig. 1 International death percentage during February 2020–July 2020

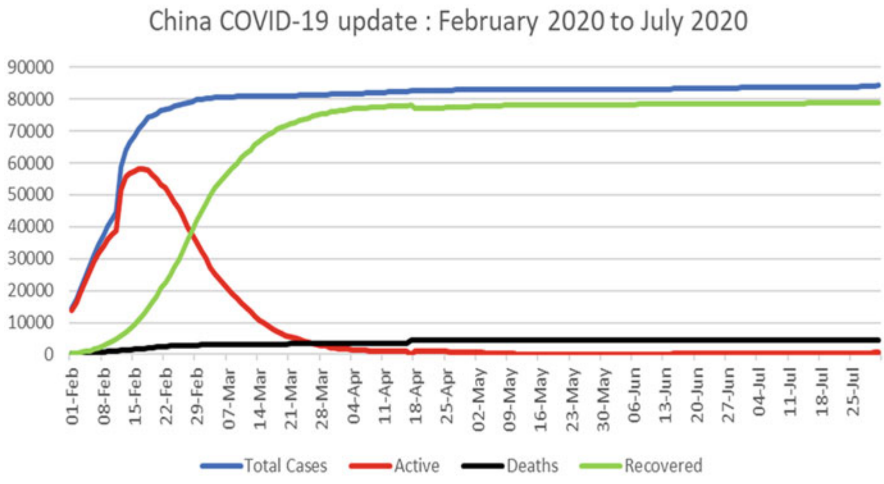


Fig. 2 COVID data analysis in China

March but it showed sharp decline during July. However, fatality was controlled. The fatality rate is highest in Italy and France among all countries. The analysis presents that while the graph for active numbers declines, the number of total cases is highest but recovered numbers are also increasing steadily and deaths are steadily declining. It provides the knowledgebase that may be almost everyone or more than 30% of population will be infected by corona virus and people will recover soon by following guidelines properly.

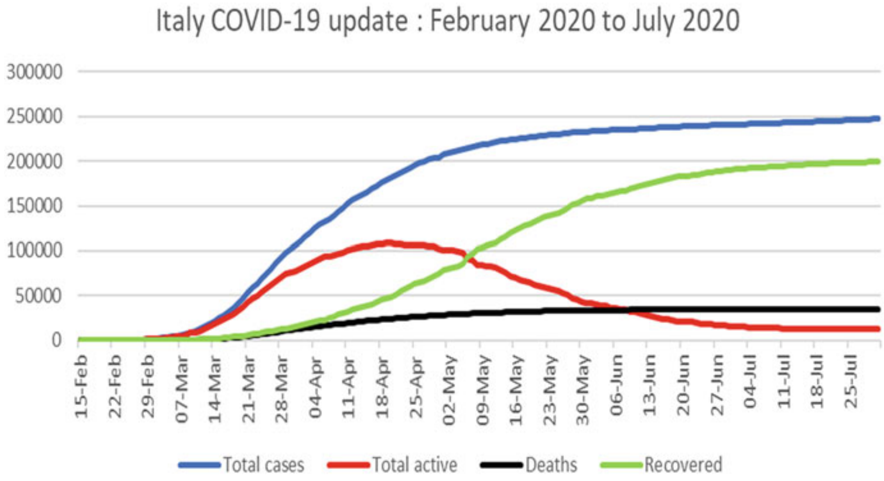


Fig. 3 COVID data analysis in Italy

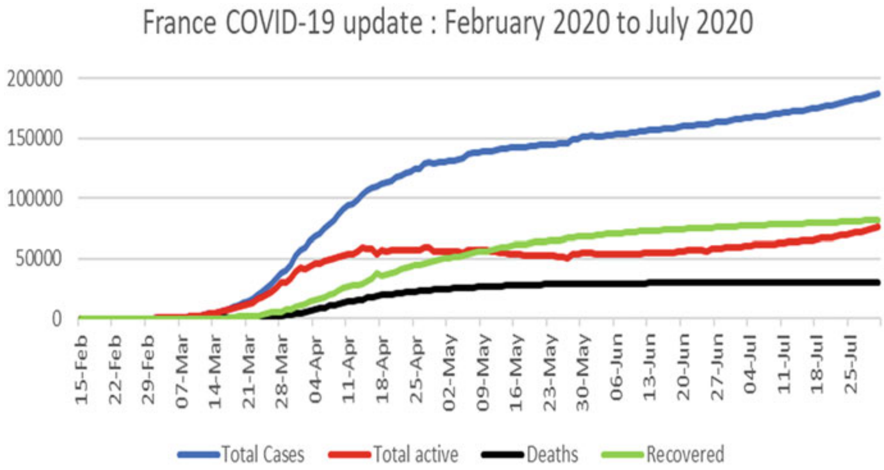


Fig. 4 COVID data analysis in France

National

The case studies for 10 different states, Bihar, Delhi, Haryana, Kerala, Maharashtra, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal, in India are presented based on observation and data collected from various social media [8, 9]. It presents the spreading pattern in various towns and villages for six consecutive months. Further, the classification of these dataset is categorized based on "A" as Active Cases, "R" as Recovered, and "D" as Deaths. It is observed that no case is reported in February 2020. So, Table 3 reflects datasets during March 2020 to July

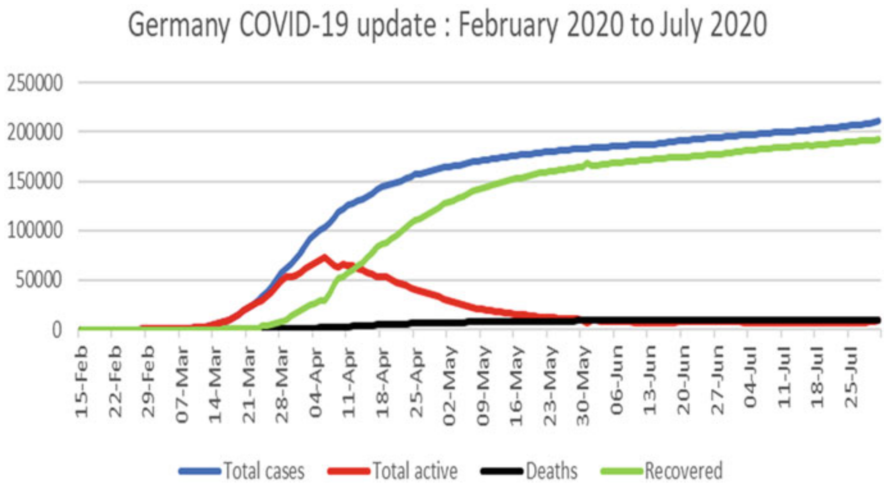


Fig. 5 COVID data analysis in Germany

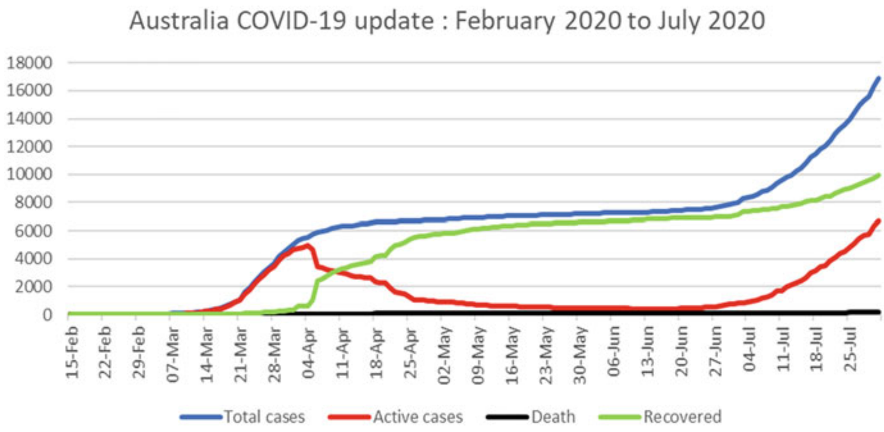


Fig. 6 COVID data analysis in Australia

2020 and Table 4 presents total active cases, recovered, and deaths for the above-mentioned states within the same timeline.

The case study presents that infected cases are increasing because the disease spreads primarily from one person to another person through the small droplets from the nose or mouth of those who have been infected with COVID-19 virus [10]. These droplets can land on objects and surfaces around the person such as tables, door-knobs, and handrails, etc., and people can become infected by touching these objects or surfaces, and then when they touch their faces.

Figure 8 presents the death percentages with respect to the total cases for all the above-mentioned states of India. The observation shows that the most infected state

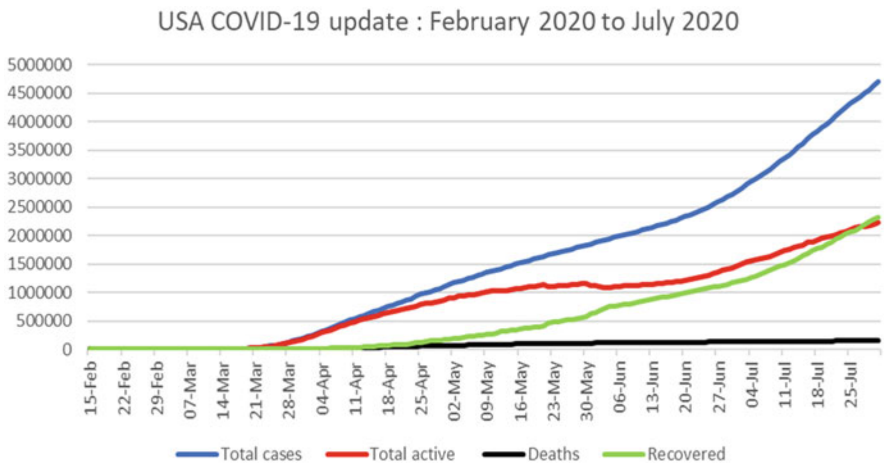


Fig. 7 COVID data analysis in United States of America

Table 3 COVID-19 datasets of various states during March 2020–July 2020

	March			April			May			June			July		
State	A	R	D	A	R	D	A	R	D	A	R	D	A	R	D
Bihar	20	0	1	339	84	1	2264	1436	21	2376	6024	45	17039	26106	230
Delhi	112	6	2	2362	1088	57	10893	7384	414	26270	49870	2269	10705	62582	1221
Haryana	19	24	0	100	211	4	1023	813	16	4340	8924	216	6317	18255	185
Kerala	215	24	2	111	359	2	670	207	6	2114	1714	15	10517	10719	49
Maharashtra	252	39	11	8266	1734	448	36040	27556	1827	75995	61582	5569	74683	150966	7139
Punjab	37	1	4	356	103	16	231	1883	25	1557	1880	99	4999	6867	242
Rajasthan	90	3	0	1633	890	58	2605	5139	136	3375	8188	219	11558	15625	267
Tamil Nadu	117	6	1	1038	1252	26	9400	11499	149	38892	37317	1025	57968	133882	2734
Uttar Pradesh	87	17	0	1620	534	40	3015	4292	177	6711	11241	480	34968	32778	933
West Bengal	31	3	3	601	121	30	3027	2033	284	5761	9973	351	20233	36244	913

Table 4 State-wise total active cases, recovered, and deaths

	Total		
State	A	R	D
Bihar	17039	33650	298
Delhi	10705	120930	3963
Haryana	6317	28227	421
Kerala	10517	13023	74
Maharashtra	150966	256158	14994
Punjab	4999	10734	386
Rajasthan	11558	29845	680
Tamil Nadu	57968	183956	3935
Uttar Pradesh	34968	48863	1630
West Bengal	20233	48374	1581

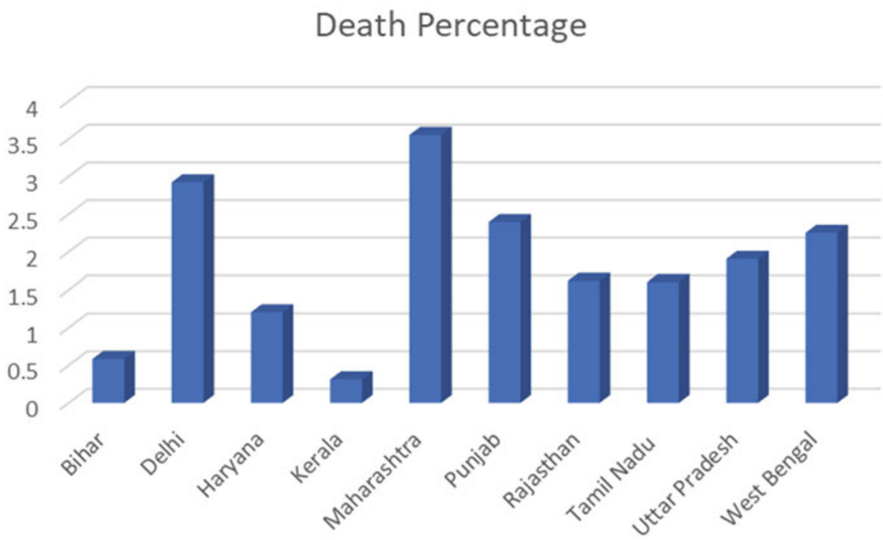


Fig. 8 National death percentage during February 2020–July 2020

is Maharashtra and the least infected state is Punjab. Similarly, in recovered cases, Punjab is also the lowest one in recovered cases among all selected 10 states and Maharashtra is the highest one in recovered cases also. And in death cases, Kerala is the lowest one among all selected 10 states and Maharashtra is the highest which proves that more devastation has occurred in all three cases.

The chapter also presents detailed analysis for six different states of India. The analysis is carried out in the same manner; data collection was from social media or from media and gathered total COVID cases were for the last 6 months from February to July 2020 [11]. However, no such cases for COVID has been reported, so the study period is considered from March and carried out till end of July. Further, total COVID cases are subdivided among the number of recovered, fatality, and finally active cases for the month. Data are presented for six states, namely, Delhi, Kerala, Maharashtra, Tamil Nadu, Uttar Pradesh, and West Bengal, in Figs. 9, 10, 11, 12, 13, and 14, respectively. Considering the virus' life time (2–3 weeks), number of active cases or recovered are counted in next month also. For example, some infected cases have been identified in last week of the month, then recovery of these cases will be floated in next month.

Hence, the analysis shows that in Delhi, recovered numbers are exceeding the total infected cases in July. It inferred that new infected cases are declining and existing infected cases have recovered. This result is very good, positive, and motivating. So, the people from other states can also believe that the fight against corona virus can overcome the situation. Analysis for rest of the states also represents that recovered numbers are higher in high magnitude compared to active and deaths.

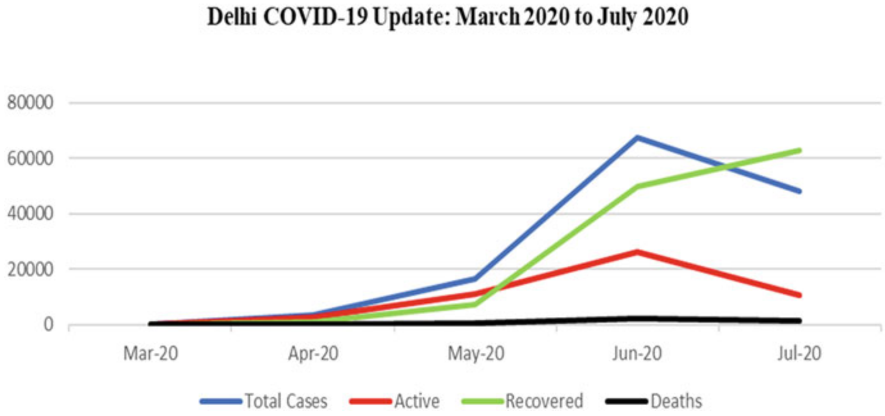


Fig. 9 COVID data analysis in Delhi

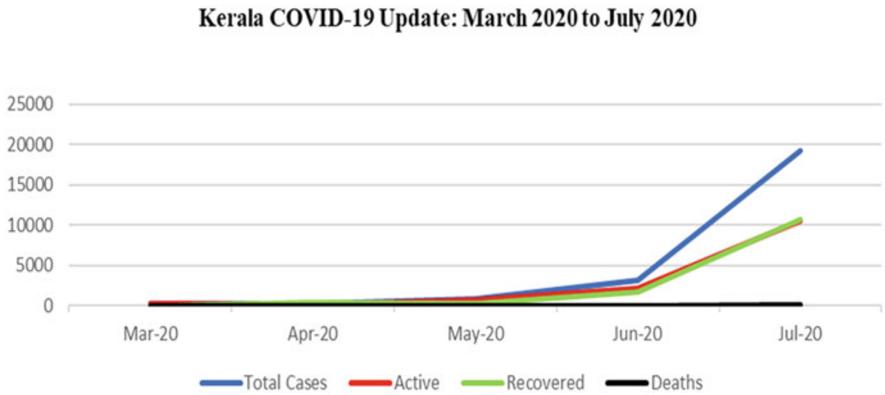


Fig. 10 COVID data analysis in Kerala

After the prolonged lockdown for consecutive 60 days, the state of West Bengal moves toward "new normal" lifestyle. In case of this "new normal" living, people are aware of corona virus. So, it is expected that even common people will also take necessary prevention while going out for work or for some other purposes. However, through "*test, track, and trace"* methodology, a few locations are identified for complete lockdown. *Complete lockdown* is enforced where the testing shows more number of "positive" results, *partial lockdown* means the locality is released from restrictions for a few hours for sustainable living, and *no lockdown* releases the location for free living. Location or region with "*complete lockdown*" is referred as "*red zone*" or high-risk zone; location or region with "*partial lockdown*" is referred as "*orange zone*" or low-risk zone, and rest of the locations are identified as "*green zone*" [12].

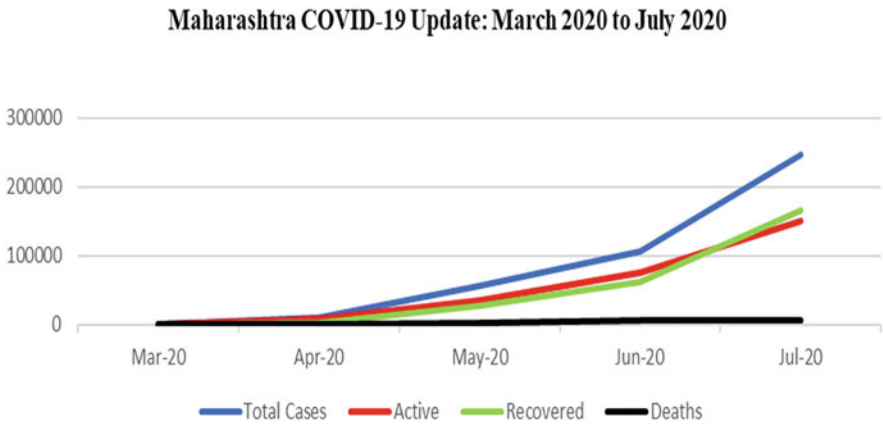


Fig. 11 COVID data analysis in Maharashtra

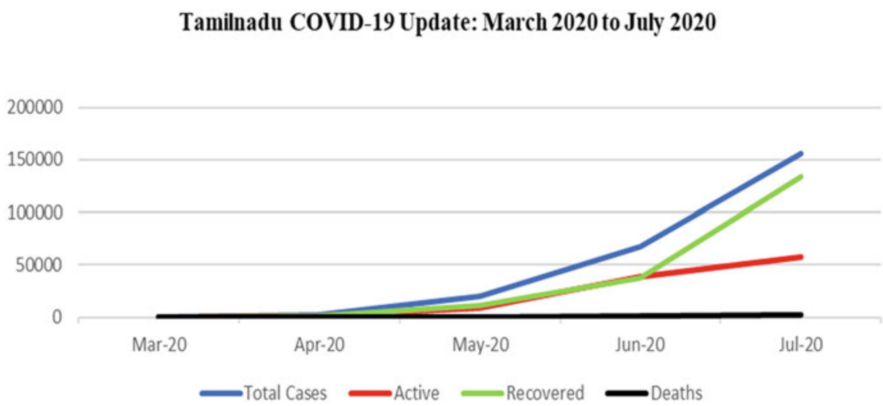


Fig. 12 COVID data analysis in Tamil Nadu

Table 5 presents the list of zones such as red, orange, and green, which specifically show that various districts of West Bengal fall in these zones.

As of July 31, according to the list, 10 districts of West Bengal are in red zones where the maximum number of coronavirus cases has emerged; five districts have been identified as orange zones in the state, while eight districts have been classified in the green zone.

Next section presents a rule-based learning for the coronavirus pandemic, applicable in healthcare.

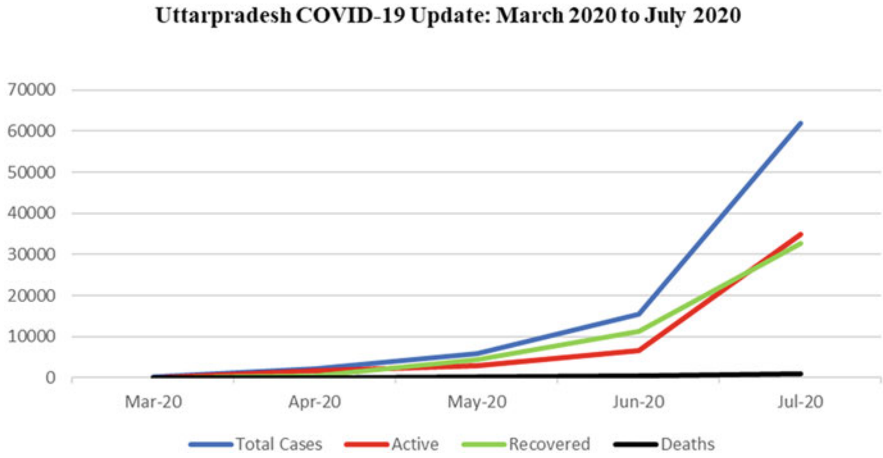


Fig. 13 COVID data analysis in Uttar Pradesh

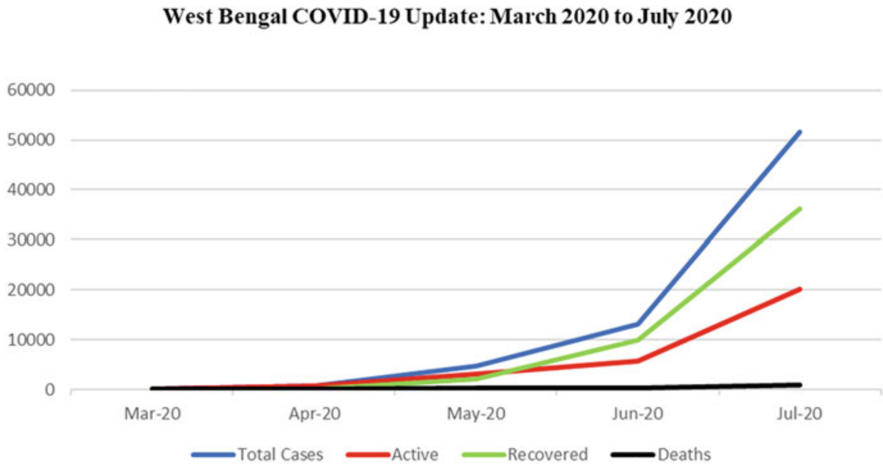


Fig. 14 COVID data analysis in West Bengal

6 Rule-Based Learning in Healthcare Using Pandemic Knowledgebase

The above case studies show that irrespective of the country, caste, gender, and age, people are devastated by the spreading of COVID-19. It is also inferred that the speed of transmission and number of deaths vary from one location to other location. Moreover, it is observed that the spreading of virus may be controlled in a more conservative manner, since the case study illustrated that death cases vary from urban to rural living style. Maharashtra being the most affected state, death

Table 5 Zone-wise district lists in West Bengal

District	Zone
Kolkata	Red Zone
Howrah	Red Zone
North 24 Parganas	Red Zone
South 24 Parganas	Red Zone
Paschim Medinipur (West Midnapore)	Red Zone
Purba Medinipur (East Midnapore)	Red Zone
Darjeeling	Red Zone
Jalpaiguri	Red Zone
Kalimpong	Red Zone
Malda	Red Zone
Hooghly	Orange Zone
Paschim Bardhhaman	Orange Zone
Nadia	Orange Zone
Purba Bardhhaman	Orange Zone
Murshidabad	Orange Zone
Uttar Dinajpur	Green Zone
Bankura	Green Zone
Birbhum	Green Zone
Cooch Behar	Green Zone
Dakshin Dinajpur	Green Zone
Purulia	Green Zone
Alipurduar	Green Zone
Jhargram	Green Zone

rates are still maximum 3.5% of its total number of cases till date. However, from the international case study, it is proved that top three death rate percentages lie between 14% and 16% of its total number of infected cases.

To control or reduce the spreading, the role of **3-T**'s is very important. These **3-T**'s are *testing*, *tracking,* and *tracing*. The first T, testing, can confirm identification of positive cases. However, India, being a country with a high population density of almost 1380 million people, cannot control the spreading only through testing. So, tracking and tracing are also very important steps for identifying the location. With the speedy transmission of virus, it is also obvious that all the positive cases are not having prominent symptoms, rather the asymptomatic cases also increase gradually [13].

The very common symptoms for COVID-19 are similar to normal influenza: cold and cough [14]. Hence, similar disease is possible through both these viruses, and due to lack of care, respiratory problems may arise. Moreover, according to the analysis of various symptoms from the COVID-19 patients, a few identified symptoms with various measures are listed below:

(a) **Fever** – below 100 °F, between 100 °F and 103 °F, and above 103 °F
(b) **Cough** – moderate to severe, more severe, leading to chest pain

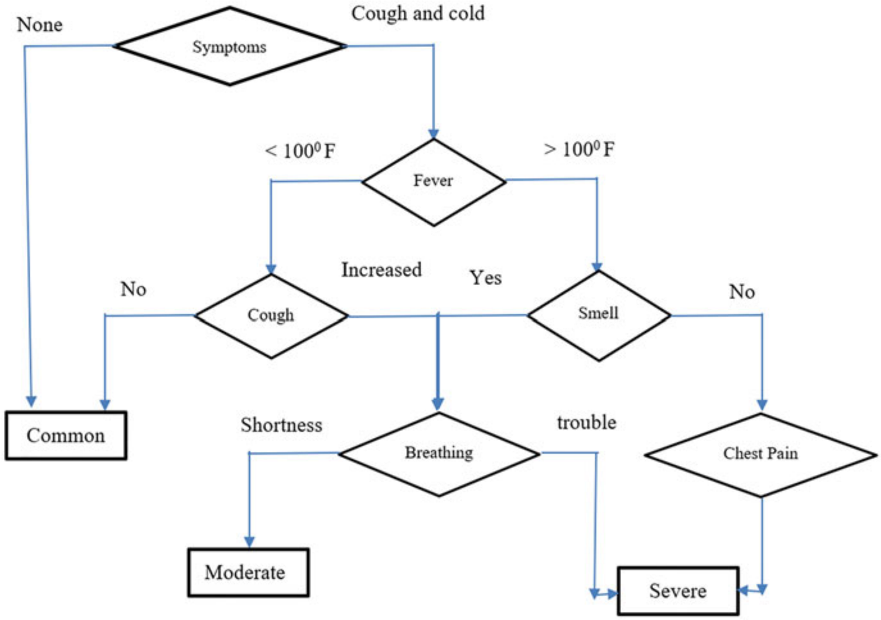


Fig. 15 Decision tree-based symptoms analysis of COVID-19

(c) **Breathing** – shortness, trouble breathing, and severe
(d) **Smell** – loss of smell, loss of taste, and both persist longer

Decision tree is an important supervised learning technique in machine learning, which is applied to both categorical and continuous variable datasets [15]. Using decision tree-based learning, classification of symptoms of COVID-19 is presented in Fig. 15 [16]. This broad classification can easily identify *common* people who have very less infection rate of the virus; Next category, *moderate* suggests chances of having the positive cases through testing, and *severe* suggests advanced stage of the infected case.

The pattern of spreading clustered in identified area or community transmission may be represented using decision tree-based analysis datasets. The rule-based learning also recommends some preventive measure to control mortality [17]. Table 6 presents a suggested rule-based preventive measure. This preventive measure will be used as knowledgebase for controlling the devastation of further spreading and lead to speedy recovery. At the outset, according to the World Health Organization, the virus is now floating in air also and most probably there is high chance of infection of everyone in every room. Hence, the campaign needs to be carried out to make people aware enough to do self-control by taking proper precaution for not to be infected. Even if they are infected, the rule base will lead them to speedy and proper recovery.

Table 6 Suggested rule-based preventive measure

Symptoms category	Suggested prevention measure
Common	Home isolation and continue immunity-boosting habits
Moderate	Medical care and quarantine in safe home with continuation of immunity-boosting habits
Severe	Mandatory hospitalization and follow quality healthcare

It needs to be mentioned that people need to increase and continue immunity-boosting habits like a few wake-up Ayurvedic starters, such as lemon water or turmeric in empty stomach and mild breathing exercises. These also prevent the spreading of virus and control the transmission speed of the corona viruses. Having major similarities to influenza virus, the major deviation is that children are main carriers of influenza virus, while in the case of COVID-19, children are less affected. The analysis presents that due to lot of vaccinations, children (age group above 10) can strongly fight against COVID-19. Hence, elderly people are at the highest risk of the coronavirus.

7 Conclusion

This chapter presents meticulous data analysis of pandemic COVID-19 to suggest rule-based knowledgebase for speedy recovery of the affected patients. The rule-based knowledgebase is prepared using machine learning technique, specifically decision tree-based algorithms. The common and severe symptoms are presented with various boundaries to make aware people of any age, color, caste, and gender. During recent years, a wide range of coronaviruses that cause a wide assortment of human and veterinary sicknesses have happened. Almost certainly, these infections will proceed to rise and cause both human and veterinary episodes inferable from their capacity to recombine, change, and contaminate various species and cell types.

Future research on coronaviruses will keep examining numerous parts of viral replication and autogenesis. In the first place, understanding the inclination of these infections to bounce between species, to build up disease in another host, and to distinguish huge repositories of coronaviruses will drastically help our capacity to foresee when and where potential waves may happen.

References

1. M. Awad, R. Khanna, *Machine learning and knowledge discovery* (Apress, Berkeley, 2015). https://doi.org/10.1007/978-1-4302-5990-9_2

2. Spanish flu. Accessible at https://en.wikipedia.org/wiki/Spanish_flu. Accessed on 1 Aug 2020 (n.d.)
3. Western African Ebola virus epidemic. Accessible at https://en.wikipedia.org/wiki/Western_African_Ebola_virus_epidemic. Accessed on 1 Aug 2020 (n.d.)
4. M.M. Haby, M. Pinart, V. Elias, L. Reveiz, Prevalence of asymptomatic Zika virus infection: A systematic review. Bull. World Health Organ. **96**(6), 402 (2018)
5. A.J.L.A. da Cunha, M.C. de Magalhães-Barbosa, F. Lima-Setta, R. de Andrade Medronho, A. Prata-Barbosa, Microcephaly case fatality rate associated with Zika virus infection in Brazil: Current estimates. Pediatr. Infect. Dis. J. **36**(5), 528–530 (2017)
6. Coronavirus COVID-19 Live tracker Johns Hopkins. https://www.grainmart.in/news/coronavirus-covd-19-live-cases-tracker-john-hopkins/ (n.d.)
7. COVID-19 World dataset, Worldometer. https://www.worldometers.info/coronavirus/. Accessed on 1 Aug 2020 (n.d.)
8. COVID-19 India dataset. https://www.covid19india.org/. Accessed on 1 Aug 2020 (n.d.)
9. WHO Coronavirus disease. https://covid19.who.int/table. Accessed on 1 Aug 2020 (n.d.)
10. India fights COVID-19. https://www.mygov.in/covid-19. Accessed on 1 Aug 2020 (n.d.)
11. COVID-19: Pandemic in India. https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_India#cite_note-4. Accessed on 1 Aug 2020 (n.d.)
12. English Jagaran: Coronavirus Hotspots in West Bengal: Red Zones in state, 516 areas identified as containment zones. https://english.jagran.com/india/coronavirus-hotspots-in-west-bengal-red-zones-green-zone-orange-zone-bengal-containment-areas-check-full-list-here-10011474. Accessed on 1 Aug 2020 (n.d.)
13. ICMR to study prevalence of asymptomatic COVID-19 infected people to check for community spread. Accessible at https://economictimes.indiatimes.com/news/politics-and-nation/icmr-to-study-prevalence-of-asymptomatic-covid-19-infected-people-to-check-for-community-spread/articleshow/75631365.cms?from=mdr. Accessed on 1 Aug 2020 (n.d.)
14. Everything you should know about coronavirus. https://www.healthline.com/health/coronavirus-covid-19#coronavirus-types. Accessed on 1 Aug 2020 (n.d.)
15. Decision Tree Algorithm. https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb. Accessed on 1 Aug 2020 (n.d.)
16. Introduction to Decision Tree. https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html. Accessed on 1 Aug 2020 (n.d.)
17. A. Das, A. Desarkar, Decision tree-based analytics for reducing air pollution. J. Inf. Knowl. Manag. **17**(02), 1850015 (2018)

Reviewing Classification Methods on Health Care

Devvrath Malik and Geetika Munjal

1 Introduction to Supervised Learning

Machine learning is used to program a computer to make predictions or decisions about a certain scenario. The computer achieves this feat by using the experiences it gained while training on a set of data known as "training data." Machine learning can be of two sorts: supervised learning and unsupervised learning. Supervised learning is a technique which maps the input for an output based on input–output pair of examples. Unsupervised learning technique is the one which learns from test data that have not been labeled. The main aim of unsupervised machine learning is to model the underlying structure or distribution in the data to get more knowledge from the data [1].

Different Supervised Learning Methods

The various classification algorithms in machine learning are divided into two broad categories: lazy learners and eager learners. (i) Lazy learner algorithm simply sets aside the training data until the test set data comes up. It classifies the instances of test data by using the stored training set data that is most related to the test set data. Hence, it has higher predicting time compared to eager learners. Two most common lazy learner algorithms are case-based reasoning and K nearest neighbor. (ii) Eager learner algorithm creates a machine learning classifier using given training set data before taking test set data for predictions. Here, a single hypothesis works for the entire dataset and hence is the reason they take more time in training the model and

D. Malik · G. Munjal (✉)
Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_8

less time in making prediction, for example, Naive Bayes classifier, artificial neural networks, and decision trees [2].

Decision tree algorithm is used in both classification and regression (cart) tasks. It builds the classifier model in the form of a tree-type structure. It works by breaking the data into minor groups of data and in the same time frame an associated decision tree is stepwise constructed [3]. An extended version of decision tree is a random forest algorithm wherein large a number of decision trees come together to work as an ensemble classifier model. Each tree in random forest algorithm performs its own operation to predict a class for a set of input attributes and the class that is predicted by majority of decision trees is set as classifier's final output prediction [4]. The low correlation in individual trees is the key to the model's prediction. Another classifier is an artificial neural networks which can be adapted in deep learning model. It is influenced by the working of a human brain as it follows the concept of a neuron. The network comprises of an input layer, successive hidden layers, and an output layer, where each node in a layer represents a single neuron that is interlinked with every single node in the next layer. It is a computational classifier with neurons acting as processing units that receive inputs and deliver outputs based on their corresponding activation functions [5].

Support vector machine (SVM) algorithm is also a supervised learner that can be used for both classification and regression tasks but is preferably used for classifying the data. In SVM, we plot each instance of a data as a single point in an N-dimensional space (where N is the number of input attributes in the given dataset). The value of each feature for a given instance corresponds to the value of the coordinate for that point. The main motive of SVM classifier is to find a suitable hyperplane that can distinctly classify these data points [6]. Another classifier is the K nearest neighbor, which is a powerful yet simple classification algorithm mostly used in pattern recognition and recommendation systems. It is built upon the assumption that similar kinds of things exist close to each other. However, the algorithm becomes significantly slow for large size datasets [7].

Some classifiers are probabilistic such as Naive Bayes classifier which is based on Bayes theorem that provides a principled way to calculate conditional probability. The Naive Bayes algorithm works upon the assumption that all independent variable input predictors are independent of each other and no two different predictor's correlates with each other. In real-life scenario, the probability for this assumption to hold true is quite small. However, even for the data where this assumption does not hold, the Naive Bayes approach works surprisingly well for that data. Mathematically, the Bayes theorem formula is written as:

$$P\ (A|X) = \frac{P\ (X|A) \times P\ (A)}{P(X)} \tag{1}$$

where A and X are the events and $P(X) \neq 0$. The Naive Bayes classification algorithm uses this Bayes theorem equation to classify the input features into different classes [8]. In terms of machine learning ideology, the above equation can be rewritten as:

$$P\ (A_i\ |X)\ \ = \frac{P\ \ (X\ |A_i) \times P\ \ (A_i)}{P(X)} \tag{2}$$

where X represents all input features or independent variables and A_i is the ith category of output class. In accordance with the Naive Bayes assumption of independent relationship between the input features of data, the probability $P\ (X|A_i)$ can be calculated as the product of each feature's X_j's probability appearing in the category A_i (X_j being the jth feature of all the input feature in the dataset). The Naive Bayes algorithm calculates the $P\ (A_i|X)$ for all the i number of categories for a single instance of data and compare their values to select the category with the highest probability value, as the output class for that instance.

Another class of supervised learners is "adaptive," such as AdaBoost or adaptive boosting, where boosting refers to an ensemble technique that combines many weak learner algorithms to create a strong classifier. AdaBoost classification algorithm can be used in conjunction with many different classification algorithms to boost their performances. However, the algorithm is best suited to boost the performance of decision tree classifiers as it is a weak learner for binary classification task. The AdaBoost algorithm makes use of the decision stumps instead of the complete decision tree as weak learners. The decision stumps refer to a decision tree with a depth of one that performs just better than the random classifier [9]. The final classification output of the AdaBoost classifier is the output predicted by that category of a stump whose net significance is higher [10].

Logistic regression is another classification tool taken by machine learning from the area of statistics. It is again a supervised linear classification algorithm that makes use of a logistic sigmoid function to transmute its output prediction into a probability value mapped between 0 and 1. A contradiction seems to occur with the term "regression" being used for classification, but that is what makes it special. The algorithm uses the linear regression equation to give discrete binary outputs. However, unlike the linear regression model that fits a straight line on the data, the logistic regression classifier fits an s-shape curve on the input data that correspond to its sigmoid function [11]. Its popularity has been increased progressively over the last two decades, particularly for binary classification tasks. A simple logistic regression model for binary classification is represented mathematically as shown below:

$$\ln\left(\frac{\tau}{1-\tau}\right) = \alpha + \beta X \tag{3}$$

where τ is the probability that the input X belongs to the default class $y = 1$. Formally, it is written as $\tau(x) = \tau(y = 1|x)$. The ratio on the left side of Eq. 3 is known as the odds ratio of the default class. Simple logistic regression formula can easily be extended to multiple input features $(x_1, x_2, \ldots x_n)$.

Comparative Summary of Supervised Methods

Based on our studies of above algorithms, we have summarized pros and cons of various classifiers in Table 1.

Table 1 Advantages and disadvantages of various classification algorithms

Classifiers	Pros	Cons
K nearest neighbor	No training period is required New training data can be added any moment Very easy to apply	Memory consuming Less efficient with large and high dimensional dataset
Logistic regression	Good accuracy Easy implementation Fast training	Not applicable to nonlinear problems No assumption about classes in feature space
Naive Bayes	Easy implementation and quick prediction [12] Requires less training dataset	Very sensitive about input data May predict wrong if the attributes are correlated
Support vector machine	Works well with clear margin of seperation Efficient with high dimensional datasets Memory efficient Effective even with unlabeled data	Does not work with large datasets Training time is more
Random forest	Very effective Provide a reliable feature importance estimate	Slow prediction Memory consuming High computational cost
Decision trees	Needs less labor in pre processing of data Normalization and scaling of data is not required Classifier model is intuitive	Small change in data leads to instability in the model Complex calculations Adding small amount of data causes big change in the structure of the tree
AdaBoost	Classification accuracy is better Versatile, can be combined with any machine learning algorithm Not much parameter tuning is required	Sensitive to uniform noise Weak learners need to perform better than the random chance If weak learners are too weak, overfitting occurs Learning time is longer
ANN	Ability to learn complex patterns and nonlinear relationships Parallel processing Has fault tolerance	Unknown duration of training the model Large number of parameters to be tuned Follows a black box approach Optimization time is longer

2 Applications of Supervised Learning in Healthcare

Classification methods have been applied in various sectors where health care has a lot of scope where supervised learning can be very beneficial in solving various problems. Some of them are to help in recognizing and tracing long-term diseases and patients with high risk, design appropriate medication, and reduce patients admitted to the hospitals, thus helping in the healthcare governance [13]. Adopting these supervised learning methods will reduce the pressure that the hospitals may face in times of epidemics and pandemics. One of the applications of supervised learning methods is in cardiovascular disease management, which occurs due to acute coronary syndrome (ACS), where a patient comes into the hospital with a chest pain that is mainly caused by benignant causes and thus enormous resources are needed for detection purposes [14]. The answer to this problem lies in using the current resources efficiently. To do this, we can take the help of supervised learning methods which involve the decision with regard to the magnitude of care, logical allotment of resources, and calculation of modifiable risk, which can make the patients better. We can tailor the treatment suitable for a particular patient based on his medical records, personal and family antecedents, electrocardiogram, biomarkers, noninvasive stratification tests, and coronary angiography. Various methods, including artificial neural networks and deep learning, decision trees, and support vector machines, are used for this problem. Machine learning is also helping the radiologists in various forms, such as (i) creating study protocols where machine learning can help radiologists create study protocols based on their priorities. This may involve assessing clinical information and commanding information stored in electronic devices; (ii) Refining image quality and reducing radiation dose in CT, in which there has always been a desire to minimize the dose of radiation during a CT scan. But if we reduce the radiation level in CT scan, it leads to noise in the image that is obtained, hence it is of poor quality. By using deep learning, we are essentially increasing the quality of images even though we are using low doses of radiation. (iii) Optimizing MR scanner utilization as an MRI scan takes a lot of time. Hence, by using machine learning, we analyze patients' clinical record and allot them a time slot accordingly to optimize MRI scan. (iv) Evaluating image quality as it is quite time-consuming. Hence, we use machine learning to automate it [15].

A comparative analysis of decision tree, random forest, multilayer perceptron, and Naive Bayes is presented [16], where decision tree along with correlation-based feature selection has performed well for detecting dementia. Classification algorithms are widely used in breast cancer categorization as well [17]. In medical datasets with large number of input features, preprocessing is required to identify relevant features, followed by classification task [18]. The deploying of classification algorithms in diagnosis has also helped in identifying the possibility of the return of the disease in patients who were cured earlier, in spotting a high-risk disease or illness [19]. It can help in identifying the transition of a patient from one disease state to another. Machine learning algorithms have been recently used in spotting the transition from prediabetes state to type-2 diabetes with the help of electronic health record data [20]. With the advancement of natural language processing (NLP) in machine learning and AI, the researchers and data enthusiast have been able to draw relevant information, insights from the unstructured data generated in the form of clinical reports, performance feedback of a doctor, and from other medical reports of patients after successful disease treatment. The use of classification algorithms in combination with NLP can help not only in drawing patterns from unstructured data into quality and performance, but also in early prediction and diagnosis of a disease. Recently, an automated speech analysis in combination with classification algorithms was performed on the free speech generated from the in-person interviews of individuals at clinical high risk for psychosis [21]. This study was able to predict transition to psychosis state with great accuracy for a group of individuals marked at high risk. With the evolution of health monitoring technology that is heavily dependent on machine learning and artificial intelligence, it is not only possible to keep track of one's health and to predict early symptoms of a disease but also to monitor slightly different aspects of health status like mental fatigue. Mental issues like depression, anxiety, addiction, and behavior disorders have become a serious health issue nowadays, and it comes at a very high public health cost. A recent study on predicting mental fatigue using eye-tracking data was successfully able to detect the problem in individuals with 91% accuracy [22]. There are enormous applications of machine learning algorithms in medicines; quite recently, Google developed a machine learning algorithm that is able to predict the cancerous tumor on mammograms. This new approach obtained an accuracy of 89% compared to 73% of a human pathologist [23].

3 Healthcare Datasets Used in the Study

Considering the importance of classification task on health care, we have tried to analyze performance of various classifiers on various medical datasets available in open platform. The results of all the classifiers are compared on various metrics based on confusion matrix. The data are preprocessed and classified using KNN, Naive Bayes, logistic regression, AdaBoost, decision tree, and artificial neural networks. We have applied all these supervised methods on various healthcare datasets, which are briefed in Table 2, followed by their detailed description.

Cleveland Heart Disease Dataset Heart disease refers to the broad number of health conditions that has a direct impact on a human heart. It is one of the leading reasons behind a large number of deaths across the world. The original source of the dataset is Cleveland Clinic Foundation, which includes about 303 observation samples, each with 13 input predictor attributes, which are gender, age, chest pain category, blood pressure, serum cholesterol, blood sugar level, ECG, maximum heart rate, induced angina, depression induced by exercise w.r.t rest, slope of peak exercise, number of vessel, and thal. All these are used to classify if the patient is suffering from a heart problem or not [24].

Table 2 Description of medical datasets used in the study

Medical datasets	No of samples	No of input attributes	Remark
PIMA Indian diabetes	768	8	Classifying if the patient has diabetes
Wisconsin breast cancer	699	30	Classifying if cancer is malign
Cleveland heart disease	303	13	Classifying if the patient has heart disease
Indian liver patient dataset (ILPD)	583	10	Classifying if the person is a liver patient
Cesarean section classification	80	5	Classifying if the patient needs c-section

PIMA Indian Diabetes Dataset This dataset is used to predict if a person with certain diagnostic measurements is at risk to develop diabetes in near future or not. The data were originally given by the National Institute of Diabetes and Digestive and Kidney Diseases [25]. It is a small dataset with 768 instances and 8 input variables including pregnancy, glucose level, blood pressure, skin thickness, insulin level, body mass index, diabetes pedigree, and age. One outcome variable is 0 if patient does not have diabetes and "1" if the patient has diabetes. All the instances in it are of females of PIMA Indian heritage with minimum age of 21 years [26].

Breast Cancer Wisconsin (Diagnostic) Dataset These data were originally generated by Dr. William H. Wolberg at the University of Wisconsin, USA and include 569 samples and 32 columns. Its input features were collected from a digital scanned image of a fine needle aspirate of a breast mass collected from patients. High-resolution graphical computer program called Xcyt was used to measure the features of a cell based on the digital scan performed. It uses the curve-fitting algorithm that computed different features for each cell in the sample including ID Number, diagnosis, cell radius, texture, cell perimeter, cell area, smoothness, compactness, concavity, concave point, symmetry, fractal dimension. It then calculates three different values for each feature of a cell image namely mean value, extreme value, and standard error, resulting in a 30 real value attribute [27]. All this information is used to check if cancer is malign or not.

Indian Liver Patient Dataset The liver is one of the primary internal organs that takes control of various critical functions happening inside the human body, such as protein production, detoxifying chemicals, filtering of blood coming from digestive tracks, and many more. Liver disease is a very broad category that includes any disturbance in the functioning of the liver causing illness. The data taken up from the UCI machine learning repository is originally collected from north eastern part of Andhra Pradesh, India. It consists of 583 instances of sample data, each having 11 columns, of which 10 input features including Age, Gender, Total Bilirubin, Direct Bilirubin, Alkaline Phosphotase, Sgpt Alamine Aminotransferase, Sgot Aspartate

Aminotransferase, Total Protiens, ALB Albumin and Albumin and Globulin Ratio and one selector field to split the data into two sets [28].

Cesarean Section Classification Dataset A cesarean section, also stated as c-section, is a surgical operation performed to deliver a baby through an incision made in the abdomen and the uterus of the mother [29]. The c-section is recommended only when the vaginal birth is too risky to perform. This is usually the case when baby is in a breech position inside the mother's womb or when the mother develops high blood pressure during her pregnancy (preeclampsia). The c-section dataset, also taken up from the UCI repository, consists of 80 observations of patient reports and 5 input attributes: age, delivery number, delivery time, blood pressure (BP), and heart problem, based on which it is predicted if the pregnant woman needs a cesarean section to give birth or not.

4 Classification Metrics

To measure the effectiveness of a classification model, we need some metrics that explain the performance of a classifier. The performance analysis of all the classifiers is done on the above-mentioned datasets based on the evaluation metrics that are derived from a confusion matrix [30]. A confusion matrix is shown in Table 3.

where true negative (TN) is represented as the amount of cases when machine classified the output class as 0 and actual output class is also 0. True positive (TP) is represented as the amount of cases when machine classified the output class as 1 and actual output class is also 1. False positive (FP) is represented as the amount of cases when machine classified the output class as 1, but actual output class was 0. False negative (FN) is represented as the amount of cases when machine classified the output class as 0, but actual output class was 1.

Accuracy It is the most common metric to measure the effectiveness of an algorithm. It is the ratio of the number of correct predictions made by the classifier model to the total number of predictions made by the model. Accuracy can also be calculated from the confusion matrix as:

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN}$$

Table 3 Confusion matrix for binary classification

Predicted / True value	Negative – 0	Positive – 1
Negative – 0	True negative (TN)	False positive (FP)
Positive – 1	False negative (FN)	True positive (TP)

However, accuracy can sometimes be misleading while evaluating the model particularly for an imbalanced dataset. If we have 100 samples with 95 samples labeled as positive and 5 as negative, then a classifier which predicts the value for the most frequent class for all predictions will have an accuracy of 95%. This is called accuracy paradox.

Precision Precision is defined as the fraction of the number of true positive results to the total positive results predicted by the classifier. The significance of precision is that it measures the quality of the classifier's prediction on account of what the classifier claims to be positive. It can be calculated using the confusion matrix as:

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

Recall It is defined as the fraction of the number of true positive results to the sum of true positive and false negative results predicted by the classifier. In simple words, it is the number of correct positive predictions made by the classifier divided by all the samples that should be positive. It represents the percentage of total relevant results correctly predicted by the classifier. It is calculated as shown below:

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

F1 Score F1 score is defined as the harmonic mean between the precision and recall. There is a trade-off that occurs between the recall and precision and F1 score represents the balance between them. It tells about the robustness of a classifier and how precise it is. The value of F1 score always lies between 0 and 1 and a higher value of F1 score represents a better classification model. It is calculated as follows:

$$f1 = 2 \times \frac{(\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})}$$

5 Methodology

The datasets mentioned in Sect 3 are analyzed using various classification algorithms mentioned in Sect. 2, where common sequential approach is followed as depicted in Fig. 1. Initially the dataset is collected and preprocessed so that all the noise in the data is removed. Data preprocessing is necessary to improve the quality of data that directly affects the performance of classifiers. It refers to the *trans* transformation of raw data into a format that is suitable for a supervised machine learning algorithm to perform its functions. Irrelevant information from the dataset is also removed as part of preprocessing. The preprocessing is different for all datasets, which varies from eliminating redundant features, categorical encoding, and standardization. As an example, the cesarean section and the PIMA Indian

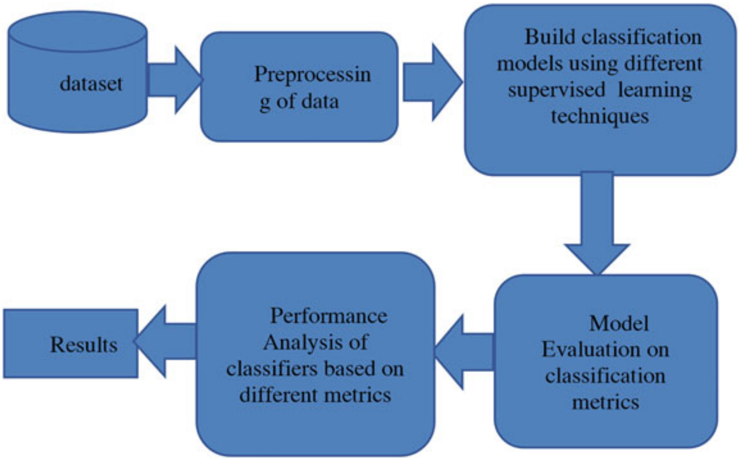


Fig. 1 Sequence of steps followed in classification task

diabetes datasets consisted of few observation samples with missing values for a primary attribute that were filled with mean imputation technique [31] to reduce the complexities in data.

Since each dataset consisted of attributes having numerical values in different range, feature scaling was also performed on all the datasets mentioned in Table 2 using standardization method [32]. Aside from PIMA Indian Diabetes dataset, each dataset consisted of one or more categorical attribute that were converted into a numerical value using label encoder and one hot encoder technique. The Wisconsin breast cancer dataset consisted of a couple of irrelevant input attributes that were removed before building the classifier models for prediction.

After the preprocessing step, the classifier models are built on each dataset using k-fold cross-validation technique. In this technique, the data sample is well shuffled and split into k number of groups. The classifier model is built/train on k-1 folds of data samples that accounts for training set and is evaluated on the *k*th fold of data sample representing a test set. The evaluation score is recorded and the process is repeated until all the k folds have been represented as test set. The mean of all the evaluation scores represents the overall performance of the classifier model [33]. After the evaluation, the parameters for the classifier model are tuned up to be fitted again on the data samples. When the performance of the classifier does not improve significantly, parameter tuning is stopped and the final performance metric values are calculated for the comparative analysis of the algorithms. Various models are compared based on their metrics values and a classifier with better performance is selected.

6 Results and Analysis

The performance analysis of all the eight classification algorithms (Table 1) is done on five different medical datasets collected from the UCI machine learning repository (Sect. 2.2). All distinguished classifier models are trained on the data samples using cross-validation technique to deal with the imbalance datasets. After the training, the models are evaluated on various performance metrics that are noted and tabulated. A perfect classifier is chosen as the one with best values for all the performance metrics.

The classification results of all the proposed supervised learning algorithms for the Cleveland heart disease dataset are reported in Table 4. Most of the models classified the data with an average accuracy of 80%. From the table it is observed that the artificial neural network (ANN) with an accuracy of 82.6% and support vector classifier with an accuracy of about 83% performed better compared to all other classification algorithms. The highest classification accuracy of 83% and F1 score of 0.856 is obtained by SVM classifier. Thus, SVM is chosen as the best algorithm to classify the Cleveland heart disease samples.

Results of PIMA Indian diabetes dataset are presented in Table 5. The diabetes dataset consisted of some missing values which were dealt with during the preprocessing stage. From the readings in Table 5, it is observed that most of the classifiers obtained an average accuracy of 75%, exception of SVM, KNN, and

Table 4 Performance metric results of algorithms for Cleveland heart disease dataset

Algorithms	Accuracy	Precision	Recall	F1 score
SVM	83.03	0.805	0.916	0.856
K nearest neighbors	81.8	0.802	0.893	0.843
Naive Bayes	81.38	0.783	0.924	0.845
Logistic regression	81.37	0.807	0.871	0.837
Random forest	80.57	0.806	0.856	0.827
AdaBoost	80.15	0.78	0.886	0.829
Decision tree	74.79	0.764	0.78	0.77
ANN	82.62	9.796	0.92	0.854

Table 5 Performance metric results of algorithms for PIMA Indian diabetes dataset

Algorithms	Accuracy	Precision	Recall	F1 score
SVM	77.45	0.758	0.916	0.615
K nearest neighbors	76.03	0.665	0.893	0.647
Naive Bayes	75.38	0.731	0.458	0.558
Logistic regression	76.24	0.70	0.562	0.618
Random forest	75.72	0.672	0.613	0.635
AdaBoost	75.7	0.712	0.527	0.599
Decision tree	68.74	0.554	0.582	0.565
ANN	73.62	0.772	0.412	0.505

Table 6 Performance metric results of algorithms for Wisconsin breast cancer dataset

Algorithms	Accuracy	Precision	Recall	F1 score
SVM	98.04	0.983	0.986	0.984
K nearest neighbors	96.94	0.966	0.986	0.975
Naive Bayes	93.42	0.947	0.95	0.947
Logistic regression	97.59	0.979	0.982	0.98
Random forest	96.92	0.969	0.982	0.975
AdaBoost	97.14	0.975	0.978	0.977
Decision tree	94.08	0.958	0.947	0.952
ANN	96.49	0.969	0.975	0.972

Table 7 Performance metric results of algorithms for Indian liver patient dataset

Algorithms	Accuracy	Precision	Recall	F1 score
SVM	71.36	0.71.35	1	0.832
K nearest neighbors	64.84	0.755	0.752	0.752
Naive Bayes	93.42	0.947	0.95	0.947
Logistic regression	71.71	0.740	0.930	0.824
Random forest	70.83	0.764	0.855	0.804
AdaBoost	71.18	0.716	0.985	0.829
Decision tree	64.29	0.749	0.749	0.745
ANN	71.35	0.713	1	0.831

logistic regression that classified the data with a little higher accuracy. Here again, it is the SVM classifier that obtained the highest accuracy of 77.45% with an F1 score of 6.1. However, it is the K nearest neighbors (KNN) classifier with accuracy of 76% that obtained the best F1 score of 0.64.

Table 6 depicts the results of breast cancer dataset. It can be observed that all algorithms classified the data with high accuracy. The best values for all the classification metrics (accuracy, precision, recall, and F1 score) were obtained by the SVM classifier that classified the data with an impressive accuracy of 98% and having F1 score of 0.98. Aside from the SVM classifier, the logistic regression and AdaBoost algorithms classified the data with an accuracy of 97.59% and 97.14%, respectively.

Performance metrics results of all the classification algorithms experimented on liver patient dataset are shown in Table 7. From the table, it is observed that five algorithms (namely logistic regression, SVM, AdaBoost, random forest, and ANN) obtained a classification accuracy between 70% and 72%. The logistic regression classifier classified the data with highest accuracy of 71.71% having F1 score of 8.2. However, the best value of F1 score (8.3) were obtained by two other classifiers, namely, SVM and ANN, having model accuracy of 71.36% and 71.35%, respectively.

Table 8 Performance metric results of algorithms for cesarean section dataset

Algorithms	Accuracy	Precision	Recall	F1 score
SVM	70.97	0.78	0.669	0.714
K nearest neighbors	63.64	0.708	0.629	0.665
Naive Bayes	73.63	0.781	0.758	0.767
Logistic regression	72.53	0.767	0.761	0.762
Random forest	57.23	0.636	0.606	0.621
AdaBoost	63.74	0.653	0.783	0.712
Decision tree	54.85	0.608	0.607	0.607
ANN	57.6	0.580	0.980	0.724

Fig. 2 Performance report of all classification algorithms in terms of F1 score metric

Performance metric results of proposed supervised learning algorithms for cesarean section dataset can be seen in Table 8. It is observed that only Gaussian Naive Bayes, logistic regression, and SVM classifiers classified the data with an accuracy of 70% or more.

The best values of classification accuracy (73.63%), precision (78.12), and F1 score (7.67) were obtained by the Gaussian Naive Bayes classifier. However, the rest of the classification algorithms failed to classify the cesarean section dataset. This failure is mainly due to the low volume of the data as it consists of only 80 observation samples. Thus, Gaussian Naive Bayes algorithm is chosen as the best algorithm to classify the cesarean section dataset.

The complete performance report of the algorithms in terms of accuracy and F1 score is much easily analyzed graphically as shown in Figs. 2 and 3, respectively.

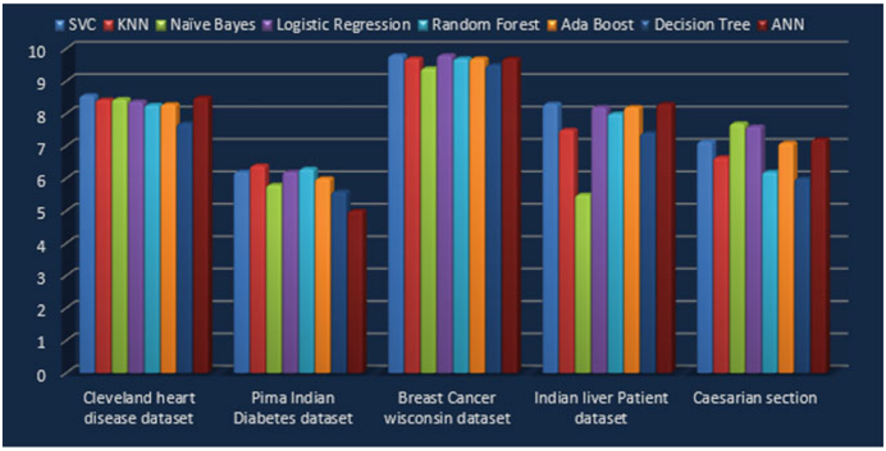


Fig. 3 Accuracy analysis of all the classifiers on five healthcare datasets

7 Conclusion and Future Work

Classifiers play a critical role in giving new insights into healthcare field from predicting prognosis, deciding treatment plan, or may be just for research purposes so that more precise and fruitful studies can be carried out.

In our study, we analyzed the performances of eight different supervised machine learning classification algorithms on five different healthcare datasets. Based on examining the classification metric results for all the algorithms on each dataset, it is observed that each classification algorithm performs differently on different kind of data, however, the SVM and logistic regression algorithm gave the best and the most consistent results for all the concerned datasets. For imbalance dataset like diabetes, F1 score is an important metric to consider. For small dataset, the Naive Bayes algorithm performs better than all other classifiers and can be preferred over SVM and logistic regression algorithms for all such cases.

The scope of supervised techniques is not limited to any extent and applying it on data can lead us to solutions to most of the problems we face today. The medical science itself is a very huge field which generates a lot of data. In the current study, we discussed techniques to limited datasets which can be extended to other different medical diagnosis. Also, with the introduction of deep learning architectures in supervised learning it can give us opportunity to explore improvement in existing results.

References

1. R. Deo, Machine learning in medicine. Am. Heart Assoc. J. **132**(20), 1920–1930 (2015)
2. D. Bhavani, A. Vasasvi, P. Keshava, Machine learning: a critical review of classification techniques. Intl. J. Adv. Res. Comp. Commun. Eng. ISSN: 2319-5940
3. O. Obaid, M. Mohammed, A. Ghani, S. Mostafa, F. Taha AL-Dhief, Evaluating the performance of machine learning techniques in the classification of Wisconsin breast cancer. Intl. J. Eng. Technol. (IJET) **7**, 160–166 (2018)
4. C. Latha, S. Jeeva, Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Inform. Med. Unlocked **16**, 100203 (2019)
5. M. Saritas, A. Yasar, Performance analysis of ANN and Naive Bayes classification algorithm for data classification. Intl. J. Intelligent Systems Appl. Eng **7**, 88–91 (2019)
6. G. Rumbe, H. Youth, Comparative study of classification techniques on breast cancer FNA biopsy data. Intl. J. Artif. Intellig. Interact. Multim. **1**, 5–12 (2010). https://doi.org/10.9781/ijimai.2010.131
7. A. Christobel, P. Sivaprakasam, Improving the performance of k-nearest neighbor algorithm for the classification of diabetes dataset with missing values. Intl. J. Comp. Eng. Technol. (IJCET) **7**(3), 155–167 (2012)
8. I. Rish, An empirical study of the Naïve Bayes classifier. IJCAI 2001 Work Empir. Methods Artif. Intell. **3**, 41–46 (2001)
9. R. Wang, AdaBoost for feature selection, classification and its relation with SVM. Intl. Conf. Solid State Devices Mat. Sci. **25**, 800–807 (2012)
10. A. Wyner, M. Maolson, J. Bleich, Explaining the success of AdaBoost and random forests as interpolating classifiers. J. Mach. Learn. Res. **18** (2017)
11. J. Peng, C.-Y.J. Peng, K.L. Lee, G.M. Ingersoll, An introduction to logistic regression analysis and reporting. J. Educ. Res. **96**, 3–14 (2002)
12. M. Gladence, M. Karthi, A. Maria, A statistical comparison of logistic regression and different Bayes classification methods for machine learning. ARPN J. Eng. Appl. Sci. **10**, 5947–5953 (2015)
13. M. Amin, A. Ali, *Performance Evaluation of Supervised Machine Learning Classifiers for Predicting Healthcare Operational Decisions* (Department of Computer Science & Engineering University of Engineering & Technology, Lahore, 2019)
14. E. Zriqat, A. Altamimi, M. Azzeh, A comparative study for predicting heart diseases using data mining classification methods. Intl. J. Comp. Sci. Inf. Security (IJCSIS) **14**(12), 869–879 (2017)
15. W. Shijun, R. Summers, Machine learning and radiology. Med. Image Anal. **16**, 933–951 (2012). https://doi.org/10.1016/j.media.2012.02.005
16. D. Bansal, R. Chhikara, K. Khanna, P. Gupta, Comparative analysis of various machine learning algorithms for detecting dementia. Procedia Comput. Sci. **132**, 1497–1502 (2018). https://doi.org/10.1016/j.procs.2018.05.102
17. G. Munjal, M. Hanmandlu, S. Srivastava, Novel gene selection method for breast cancer classification. J. Biochem. Technol. **8**(4), 1116–1120
18. M. Lamba, G. Munjal, Y. Gigras, Feature selection of micro-array expression data (FSM) – a review. Proc. Comput. Sci. **132**, 1619–1625 (2018)
19. C. Sidey-Gibbons, J. Sidey-Gibbons, Machine learning in medicine: a practical introduction. BMC Med. Res. Methodol. **19** (2019). https://doi.org/10.1186/s12874-019-0681-4
20. J. Anderson, J. Parikh, D. Shenfeld, Reverse engineering and evaluation of prediction models for progression to Type 2 diabetes: application of machine learning using electronic health records. J. Diabetes Sci. Technol. **10** (2016). https://doi.org/10.1177/1932296815620200
21. G. Bedi, F. Carrillo, G.A. Cecchi, D.F. Slezak, M. Sigman, N.B. Mota, S. Ribeiro, D.C. Javitt, M. Copelli, C.M. Corcoran, Automated analysis of free speech predicts psychosis onset in high-risk youths. Nat. Partner J. (NPJ) Schizophrenia **1**(1), 15030 (2015)

22. Y. Yamada, M. Kobayashi, Detecting mental fatigue from eye-tracking data gathered while watching video: evaluation in younger and older adults. Artif. Intell. Med. **91** (2018). https://doi.org/10.1016/j.artmed.2018.06.005
23. Y. Liu, K. Gadepalli, M. Norouzi, G. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Nelson, G. Corrado, J. Hipp, L. Peng, M. Stumpe. *Detecting Cancer Metastases on Gigapixel Pathology Images* (2017). arxiv:1703.02442.
24. N. Lutimath, C. Chethan, B.S. Pol, Prediction of heart disease using machine learning. Intl. J. Recent Technol. Eng. (IJRTE). ISSN: 2277-3878 **8**(2S10) (2019)
25. S. Vanaja, K. Rameshkumar, Performance analysis of classification algorithms on medical diagnoses-a survey. J. Comput. Sci. **11**, 30–52 (2015)
26. G. Munjal, S. Kaur. Comparative study of ANN for pattern classification WSEAS. *International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering, Bucharest, Romania* (2006)
27. R.L. Borges, Analysis of the Wisconsin Breast cancer dataset and machine learning for breast cancer detection. *Proceedings of XI Workshop de Visao Computacional* (October 5–7th, 2015)
28. M. Singaravelu, S. Rajapraksh, S. Krishnan, K. Karthik, Classification of liver patient dataset using machine learning algorithms. Intl. J. Eng. Technol. **7**, 323–326 (2018). https://doi.org/10.14419/ijet.v7i3.34.19217
29. S. Abbas, R. Riaz, S. Kazmi, S. Rizvi, S.J. Kwon, Cause analysis of caesarian sections and application of machine learning methods for classification of birth data. IEEE **6**, 67555–67561 (2018)
30. F. Sardouk, A. Duru, O. Bayat, Classification of breast cancer using data mining. Am. Sci. Res. J. Eng. Technol. Sci. (ASRJETS) **51**(1), 38–46 (2019)
31. M. Rahman, N. Davis Darryl, Machine learning based missing value imputation method for clinical datasets. IAENG Trans. Eng. Technol. **229** (2012). https://doi.org/10.1007/978-94-007-6190-2_19
32. S. Kotsiantis, D. Kanellopoulos, P. Pintelas, Data preprocessing for supervised learning. Int. J. Comput. Sci. **1**, 111–117 (2006)
33. F. Ahmed, Y. Ali, S. Shamsuddin, Using K-fold cross validation proposed models for Spikeprop learning enhancements. Intl. J. Eng. Technol. (UAE) **7**, 145–151 (2018). https://doi.org/10.14419/ijet.v7i4.11.20790

Predicting and Managing Glycemia Levels Using Advanced Time Series Forecasting Methods

Sahil Malhotra and Rita Chhikara

1 Introduction

Diabetes is a chronic disease in which either pancreas does not produce sufficient amount of insulin or body is not capable of effectively using the insulin produced by pancreas. This leads to high level of glucose in the blood and urine, which can cause damage to various parts of the body like liver, kidneys, eyes, heart, and other body organs [1]. It can be classified into two types, namely, type 1 and type 2. Type 1 diabetes is caused due to deficit production of insulin by pancreas. This type requires regular administration of insulin. The various symptoms of type 1 diabetes are weight loss, too much excretion of urine, fatigue, vision changes, thirst, and constant hunger. This is also referred as insulin dependent. The second category of diabetes (type 2) is caused as a result of body not making effective use of insulin. It is also referred as noninsulin dependent. The main causes of this type are excess body weight and sedentary life style. The symptoms are similar to type 1 but not that noticeable; hence this gets diagnosed at later stages, which could cause serious complications [2].

As per the World Health Organization (WHO) [3] statistics, the number of patients with diabetes has increased drastically from 108 million in 1980 to 422 million in 2014. Between 2000 and 2016, there was a 5% increase in premature death rate due to diabetes. It is also surveyed and found that low- and middle-income countries are more affected by diabetes than high-income countries.

S. Malhotra (✉)
Koverhoop Technologies India, Noida, India
e-mail: sahil@koverhoop.com

R. Chhikara
Department of CSE, The NorthCap University, Gurugram, India
e-mail: ritachhikara@ncuindia.edu

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_9

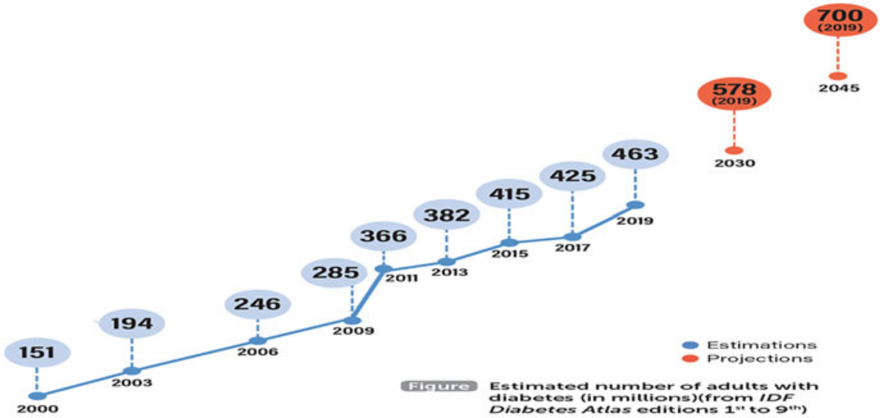


Fig. 1 Estimated number of adults with diabetes

The current scenario of diabetes as given in the International Diabetes Federation Atlas, 9th Edition [4] is that around 463 million people between 20 and 79 years were living with diabetes in 2019, and by 2045 it is likely to rise to 700 million, as shown in Fig. 1. In India, 69.2 million people were having diabetes in 2015, according to WHO. This is presumed to reach 79.4 million by 2030 [5]. The statistics of growth in the cases of diabetes are cause of alarm, and hence there is a dire need to diagnose diabetes at an early stage and thereafter self-monitor to live a healthy life.

The extreme impact of diabetes mellitus on society makes it prime field of research in medical sciences. It also generates huge amount of data which makes it apt disease for applying the machine learning techniques to predict future scenario.

Plis et al. [6] designed a general feature selection model of blood glucose dynamics to identify informative attributes for a support vector machine model. The precision is around 42%, most false alarms lying in the near-hypoglycemic events. Liszka-Hackzell described an approach to predict glucose rates using a hybrid AI approach that blends principal components method with neural networks. Results derived from this relatively basic model suggest a correlation coefficient of 0.76 in the first 15 days of analysis between the observed and the predicted values [7]. Rodríguez-Rodríguez et al. used SISAL to analyze variables like sleep, heart rate, insulin, meals, and exercise and their effect on blood glucose. The most influential variable came out to be insulin, followed by food intake. Other important features in order were exercise, heart rate, sleep, and schedule [8]. Pérez-Gandía et al. used ANN models and assessed the glucose forecasting over the prediction horizons of 15, 30, and 45 min. The RMSE for 15, 30, and 45 min of PH is about 10, 18, and 27 mg/dL, respectively [9]. Pappada et al. created an electronic diary describing factors like dosage of insulin, food consumption, everyday tasks and habits, and emotional states and trained various neural network models on this constructed dataset to forecast glucose on a window of 50–180 min prediction horizons. The models performed adequately well in the normal and hyperglycemic glucose range, whereas it overestimated the values in the hypoglycemic region mostly due to less data available for this region [10]. Amir Hayeri designed a system to forecast glucose fluctuations using heart rate, step count, and blood glucose/insulin dynamics. The system was then trained to make predictions with the data of 30 days. On average, for 60 min ahead of time, the algorithm was able to estimate glucose levels with an accuracy rate of 93% [11]. Various authors have applied different feature selection and classification techniques to create predictive models for clinical dataset of diabetes [12–15].

Researchers have also worked with deep neural networks and compared them with shallow networks. Albers et al. used three forecasting techniques, namely, "data assimilation," "averaging model" of the assimilation, and "Gaussian process with dynamic property" model regression, to produce real-time, personalized glucose forecasts. Their assessment proposes that the data assimilation forecasts contrast well with explicit glucose estimations and match or surpass the accuracy of expert forecasts [16]. Faruqui et al. proposed a technique to dynamically forecast glucose based on various parameters like physical activity, weight, diet, and glucose level taken from the prior day. The performance of the model "LSTM-NN-TF-DTW" is best in zone "A" (84.12%) of the Clark error grid [17]. Contreras et al. constructed prediction models using data for 100 online patients created by the Padova or UVA T1D simulator, obtained over 14 days. For the 100 simulated patients, midterm blood glucose was predicted using customized models and different scenarios. Around 98.31% of the predictions occurred in the Clarke error grid zones "A" and "B" [18]. Martinsson et al. presented a method for estimating diabetic blood glucose levels up to 1 hour in the future. The method is based on recurrent neural networks and uses only glucose data for a patient. The RMSE of results varies between 15 and 21 [19]. Mhaskar et al. demonstrated how deep learning models compared with shallow learning models for 30 min predict glucose levels [20].

Recently researchers have shifted their focus to time series analysis of diabetes data. Various research works have been found in the area of medical sciences which are applying time series mining for forecasting various parameters specific to that application. The ARIMA model has been applied for forecasting hemorrhagic fever in China. The best model is selected by Ljung–Box test and Akaike information criterion (AIC) [21]. The ARIMA model with Box–Jenkins methodology was used by Dan et al. del to forecast malaria mortality rate [22]. The ARIMA model has also been applied for forecasting the number of patients of epidemic diseases [23]. The ARIMA model has been found to be useful to predict the spread of COVID 2019 [24].

Most of the work in the area of diabetes focuses on techniques like feature selection, classification, and clustering. However, not much work has been done using time series analysis. It has been observed that various models of time series, namely, ARIMA and SARIMA, have been successfully applied for medical areas other than diabetes. However, there is still scope for analyzing and forecasting predictions for diabetes.

The objectives of this study are as follows:

1. To state and explain time series
2. To discover the gaps in the existing literature
3. To elucidate the need of time series in diabetes
4. To propose the model using time series for forecasting diabetes

This study has been planned as follows. Section 2 provides the background study which consists of definition, related terms of time series, and state of art specific to time series in diabetes. Section 3 explains the methodology in forecasting diabetes and also provides details of the models applied in achieving it. The experimental results and interpretation of the applied work is presented in Sect. 5. The conclusion and future scope is discussed in Sect. 5.

2 Background Study

With the advent of Internet of things (IOT) and digitization of health care, the time series analysis has shown a tremendous growth in research and is likely to grow rapidly in the years to follow. In the previous section, the scope of time series in various applications has been presented; however, a thorough study is required to give an overview of work accomplished in the area of diabetes with time series techniques. The main purpose of this assessment is to report the research done in the field of diabetes using time series techniques and its application on real-world diabetes dataset.

This section gives a brief overview of time series data and various terms related to it, like trend, cycle, stationary, etc. It highlights the studies of various researchers in the healthcare sector specific to diabetes and the applications of time series on it.

Defining Time Series and Related Terms

Time series data are data that are a collection of numeric values which are collected at varying points in time. On the other hand, a cross-sectional data would be collected over a single point in time. There is a possibility of time series data showing autocorrelation as it is assimilated over adjacent time periods. This is one property that differentiates time series data from cross-sectional data [25].

Time series visualization is a time series graph plots in which the time is represented on *x*-axis and observations are on *y*-axis. The visualization helps to identify the behavior and patterns in data which further helps to develop a robust model [26].

The time series visualization supports in identifying some of the components mentioned below which in return provides guidance in choosing the right methods for modeling the time series. It helps answer the following questions.

- Does the series show trend?
- Does it exhibit seasonality?
- Does it show explosive behavior?
- Are there any structural breaks?

Trend, Cycle, and Seasons [27]
Trend Component: A trend is either an upward movement, downward movement, or no movement (stable) over long period of time.

Cyclical Component: This time series shows recurring behavior that are of variable period. For example, eruption of volcanoes, which are not certain and vary over time. The mean length of cycles is longer than the length of a seasonal pattern. The trend element also includes the cyclical component and referred as trend cycle.

Seasonal Component: This time series displays regular fluctuations which occur over a fixed period. It could be based on 24 hours day, a month, a quarter, or a year, etc.

Stationarity and Differencing [28] The literal meaning of stationary is "not moving or not intended to be moved." Based on this definition, a stationary time series can be defined as one which is independent of time at which the data are observed. The time series showing trends or seasonality are nonstationary. On the other hand, a white noise series is stationary as it appears to be same irrespective of at what point in time it is observed. One of the methods of making a data stationary is to evaluate the differences between consecutive observations. This is referred as differencing. This process stabilizes the mean thereby removing trend and seasonality.

Various Researchers' Study on Time Series Analysis for Diabetes

This section provides a brief introduction to study of various researchers. The study of time series analysis specific to diabetes is mentioned in this section.

Rodríguez-Rodríguez et al. developed machine learning univariate time series models. They applied the model over a 15-minute window on 24 historical values taken over a span of 6 hours. The model has been able to estimate glucose levels with a mean error of 15.43 mg/dL. The mean error has further been reduced to 10.15 mg/dL by raising the sampling frequency to 72 values [29]. Frandes et al. applied individual AR models to predict over 30 min and 60 min horizons. The results for 30 min and 60 min horizons are RMSE 5.83, MAE 5.18 and RMSE 7.43 and MAE 6.54, respectively [30].

Authors applied a time series analysis on data between 2009 and 2015 (Ambulance Victoria electronic database) to predict cases of hypoglycemia and hyperglycemia by the end of 2017. They applied the SARIMA models. The best results were with SARIMA (0,1,0,12) model with a mean absolute percentage error of 4.2%. A caseload of approximately 740 was predicted on monthly basis by the end of 2017 [31].

Researchers [32] have developed a decision support system on electronic health record (EHR) for the purpose of using the American Diabetes Association guidelines for dysglycemia screening. The data comprised adults visiting five urban clinics from 2011 (March) to 2013 (December). Various use cases like percentage of patients receiving testing in a month, undiagnosed dysglcemia in patients, etc. were observed for trends. It was found that before and after the intervention, 59% of test results in eligible patients showed dysglycemia.

Jun Yang et al. has proposed an ARIMA model with a variation of adaptive orders to predict blood glucose concentrations. They have made use of CGM (continuous glucose monitoring) data to forecast blood glucose levels in future. In their work, the nonstationarity is verified with augmented Dickey–Fuller (ADF) test. They claim that the designed model outperforms the ARIMA model.

All the above studies have been performed using ARIMA models and its variants. In this study, we have applied Prophet model developed by Facebook and conducted a comparative analysis with the ARIMAX (ARIMA model with exogenous variables) model. The models have been tested on blood glucose levels of two patients taken at duration of 5 min. Data for 36 hours are taken into consideration [33]. A continuous glucose monitoring (CGM) data provided by Nightscout Foundation have been taken to predict future BG levels in order to take appropriate actions in advance to prevent hyperglycemia or hypoglycemia. The methodology is given in subsequent section and diagrammatically represented in Fig. 2.

3 Methodology

The diabetes dataset is cleaned and preprocessed in data imputation step followed by test for stationarity using augmented Dickey–Fuller test. The ACF and PACF

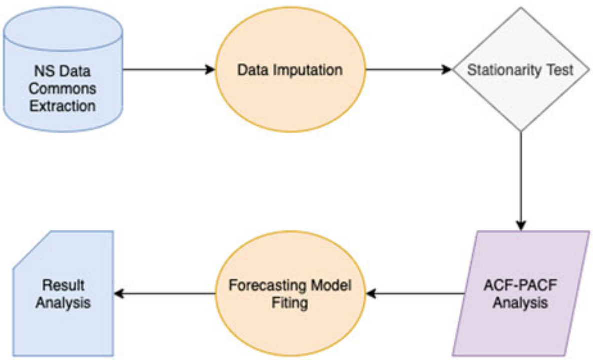


Fig. 2 Framework for time series forecasting

plots help in deciding the parameters of ARIMAX models. After performing these tests, forecasting model fitting is performed using fbProphet and ARIMAX models.

Augmented Dickey–Fuller Test

An augmented Dickey–Fuller test (ADF) checks the null hypothesis in statistics and econometrics that a unit root is present in a sample of a time series. Depending on which version of the test is used, the alternative hypothesis is different, but is typically stationary or trend stationary.

ACF and PACF Plots

Autocorrelation Function Plots A correlogram (also called the autocorrelation plot) is a graphical means of displaying serial correlation in the data that differs over time (i.e., data from time series). Serial correlation (also called autocorrelation) is when an error at a one point in time travels to a subsequent point in time.

Partial Autocorrelation Function Plots The PACF plot is a partial correlations plot of coefficients between the sequence and the lags themselves. The “partial” correlation between two variables is the amount of correlation between them that is not described with a given set of other variables due to their mutual correlations.

Techniques Applied for Time Series Analysis

Different models of time series used in this study are explained below.

ARIMAX

This algorithm was first introduced in the year 1976 by Box and Jenkins in 1976 [34]. This approach evaluates error term of univariate stochastic time series. In order to make this possible it has to be ensured that the time series to be analyzed must be stationary. All the results of the classical regression analysis become invalid for a nonstationary series. ARIMA [35] is an abbreviation that is expanded as “autoregressive integrated moving average.” The various components of ARIMA are as follows:

– **AR (autoregression).** The literal meaning of word autoregression is that it is a regression of the variable against itself. Hence, it is a model that uses the dependent relationship (y) between an observation and past values (lagged) of observations (ϕ). An autoregressive model with an order p can be presented with Eq. 1.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

where

- ε_t is white noise
- y_t is the predictor at time t
- ϕ_p represents parameters/observations
- p is the order
- c is constant values
- **I (integrated).** This is needed to make the time series stationary by using the differencing (d) method. As explained in section "Defining time series and related terms", differencing can be achieved by subtracting an observation from its value in previous time step.
- **MA (moving average).** This model uses past forecast residual errors (θ) unlike AR, where the past values of the forecast variables are considered. It is represented by the following equation:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (2)$$

This is referred as MA(q) model, which means a **M**oving **A**verage model of order q.

ARIMA models can have three parameters in python and can be represented as ARIMA (p, d, q). Each parameter is explained below.

p **–** the lag order which is the number of lag observations considered in the model.
d **–** the degree of differencing required to achieve stationarity.
q **–** the order of the moving average.

ARIMAX is an ARIMA model with exogenous variables. In our study, two exogenous variables being considered are carbohydrates and insulin.

fbProphet

Prophet [36] is open source time series forecasting software released by Facebook research team. In this model, trends which are not linear are fitted with daily seasonality, weekly, yearly plus holiday effects in an additive manner. It handles missing data and outliers effectively. fbProphet is a decomposable regression model that does not require much of hyperparameter tuning. It uses following main model components: trend, seasonality, and holidays which are mathematically represented as follows.

$$y_t = m_t + s_t + h_t + \varepsilon_t \quad (3)$$

m_t: piecewise linear or logistic growth curve for modeling nonperiodic changes in time series
s_t: periodic changes (e.g., weekly or yearly seasonality)
h_t: effects of holidays which are provided by users.
ε_t: error term takes into account the unusual changes which are not adjusted by the model

In this model, time is taken as a regressor and fbProphet works toward fitting various linear and nonlinear functions of time as components. It addresses the forecasting as a curve-fitting problem instead of looking at the dependence of each observation on time in the time series.

Trend Trend is handled by fitting a piecewise linear curve over it or the nonperiodic part of the time series. With the help of domain knowledge an analyst can then define a varying capacity for the intended time series.

Changepoints The changepoints in fbProphet can be automatically selected or manually tuned in the eventuality of the time series encountering any underlying changes in the phenomena.

If the number of changepoints allowed is further increased, then the fit becomes more flexible. A basic problem faced in any machine learning model is over-fitting (models performance is good on training set but poor on test set) and underfitting (models performance on training and test set is poor). A parameter (changepoint_prior_scale) can be used to adjust the flexibility in trend and tackle the above two problems. Higher value will fit a more flexible curve to the time series.

Seasonality The fbProphet uses Fourier series to fit and forecast the effects of seasonality. The seasonal effects s(t) can be approximated with the Eq. 4.

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{p}\right) + b_n \sin\left(\frac{2\pi nt}{p}\right)\right) \tag{4}$$

where

P is the period (7 for weekly data, 365.25 for yearly data)

Holidays and Events Holidays and events can cause foreseeable shocks to a time series. Prophet has the capability of considering a window around customized list of holidays and events separately. Additional parameters are fitted to model such effects.

Forecasting

The dataset is divided into train and test data. For this case we take the last 10 timestamps as the test data, which is forecasting for the next 50 min, and calculate the performance of the model over it.

Evaluation Using Mean Absolute Error (MAE)

Mean absolute error (MAE) in statistics is a measure of errors between paired observations which represent the same phenomenon.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{5}$$

4 Experiment Results and Analysis

Data Extraction

The dataset has been taken from Nightscout Foundation (Nightscout Data Commons) for training the models [37]. The Open Humans helped to develop a tool which allowed individuals to share their CGM and data related to with the Nightscout.

The data repository includes profile, daily glucose readings, treatment logs, like insulin, carbohydrate intake, and the device status of 58 patients. The dataset being available in json format, a toolkit has been designed first to utilize the data and load it in appropriate form. A subset of dataset of two patients (patient 1 and patient 2) has been used for training different models. The subset comprised 233 values (worth ~19 hours) and 826 values (worth ~68 hours) for patients 1 and 2, respectively.

Data Preparation

The dataset holds data of patients with diabetes in the form of a time series, with glucose values logged approximately every 5 min. Data preprocessing is performed in order to handle the missing values. These values were imputed by taking a tolerance of 15 min worth of missing glucose values and imputed it by taking the previous value.

The subset of continuous time series data which had the maximum values was then selected.

The dataset was thereafter divided into training and testing data to apply machine learning time series models (fbProphet and ARIMAX). Around 10 values (i.e., 50 min) were used for forecasting and validating the time series models and remaining were considered in training dataset.

Also, these data were merged with treatment data, so that the insulin and carbohydrates parameters can also be taken into consideration for forecasting purpose. The treatment data are merged based on the closest timestamp match.

Also note that the input to Prophet library of Facebook for time series analysis is a data frame consisting of two columns, namely, "*ds*" and "*y*." The "*ds*" (date stamp) column must be in this format YYYY-MM-DD and YYYY-MM-DD HH:MM:SS for date and timestamp, respectively. The "*y*" column should be a number and should represent the attribute to be forecasted like blood glucose in this study.

ARIMAX Model Training

In order to check whether data are stationary or nonstationary, an augmented Dickey–Fuller (ADF) test is performed.

ADF

At the onset it is ensured if the data are stationary or not. This is done using augmented Dickey–Fuller (ADF) test. The null hypothesis of the test is that the time series is not stationary (has some time-dependent structure). The alternate hypothesis is that the time series is stationary. This is applied on each patient's data and p values are analyzed. If $p < 0.05$, the null hypothesis is rejected and it indicates that the series is stationary otherwise it is nonstationary. A differencing of 1 is applied on the dataset to make the data stationary. The p values of the two patients data are given in Table 1.

For patient 2, the p value of 0.008727 indicates stationary series hence no differencing is performed and the null hypothesis is rejected. However, for patient 1, the p values using ADF test is greater than 0.05 hence differencing is done. After applying the differencing of value 1, the p value of 0.00002 is obtained and hence the series is made stationary.

Table 1 *p* Values from ADF test

ID	*p* value	*p* value (after diff.)
Patient 1	0.143	0.00002
Patient 2	0.008727	Not required

ACF and PACF

The next step after making the data stationary is to check whether AR or MA terms are needed in order to correct any autocorrelation that is still remaining in the series that was differenced.

As observed from Figs. 3 and 4, the model would perform well with autoregressive value taken as 1 and moving average as 0. The differencing of 1 is considered. Based on this observation, the ARIMAX (1, 1, 0) model has been applied on the dataset and performance is measured in terms of MAE.

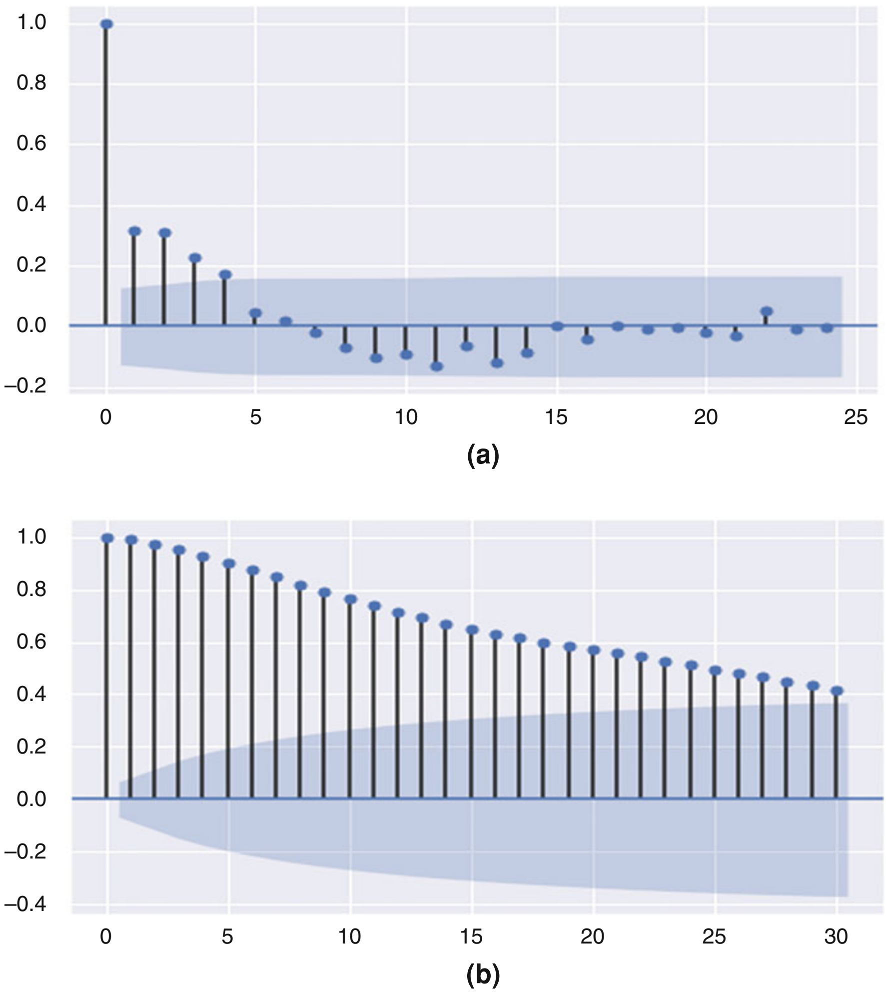
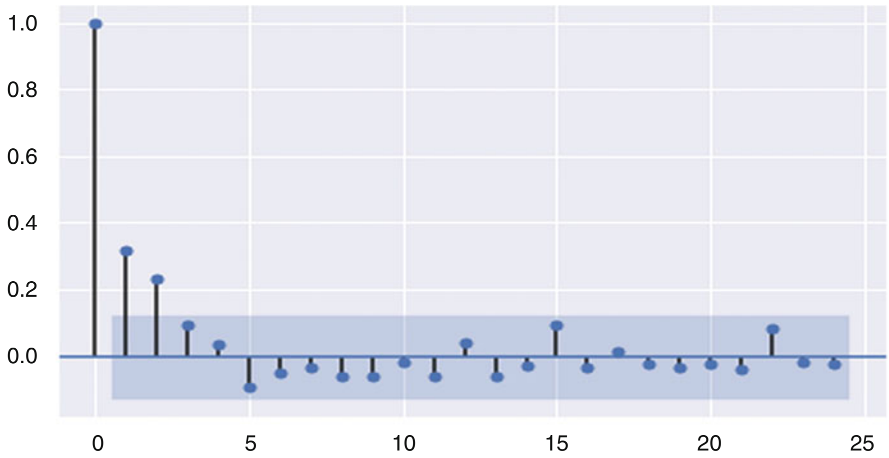
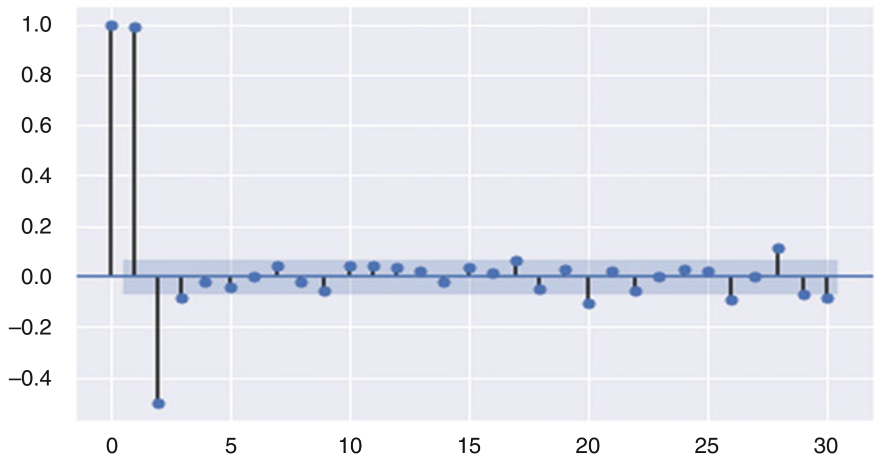


Fig. 3 Autocorrelation function for (**a**) patient 1 and (**b**) patient 2



(a)



(b)

Fig. 4 Partial autocorrelation function for **(a)** patient 1 and **(b)** patient 2

Prophet Model

The Facebook Prophet time series framework automatically handles the issues related to stationarity and autocorrelation. The visualization of data of two patients is given in Fig. 5.

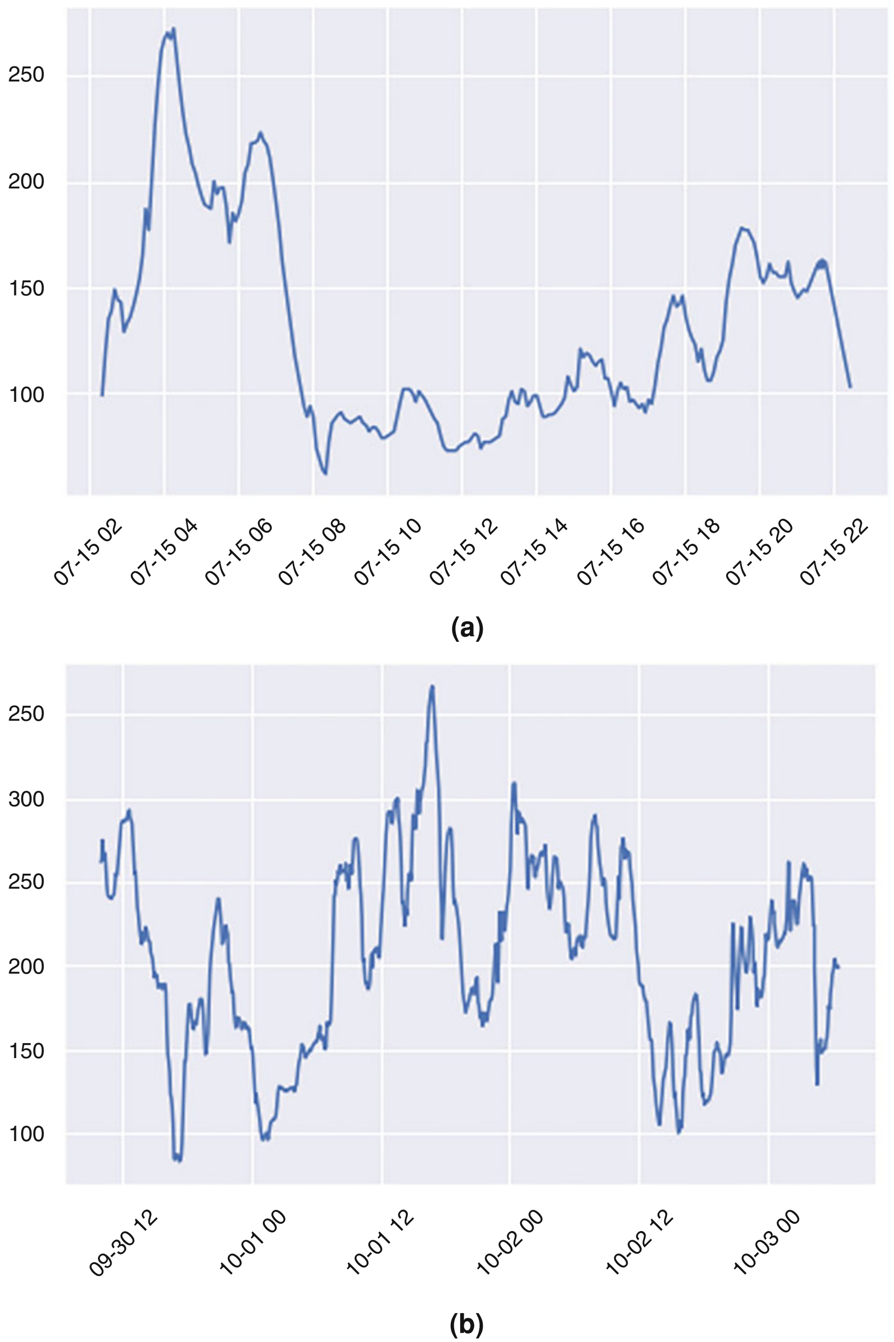


Fig. 5 Visualizing time series data for (**a**) patient 1 and (**b**) patient 2

Forecasting Model

The results are interpreted using mean absolute error (MAE) on train and test sets for Prophet and ARIMA models are given in Table 2.

The forecasting of last 10 values (50 min) is depicted in Fig. 6 for patient 1 and patient 2. The visualizations, trend, and extraregressor plots of patient 1 and patient 2 is shown in Fig. 7. The "ds" and "y" represent the timestamp and target (glucose values), respectively.

Table 2 MAE for both models

ID	Mean absolute error	
	Prophet	ARIMAX
Patient 1	5.98	15.32
Patient 2	12.82	25.07

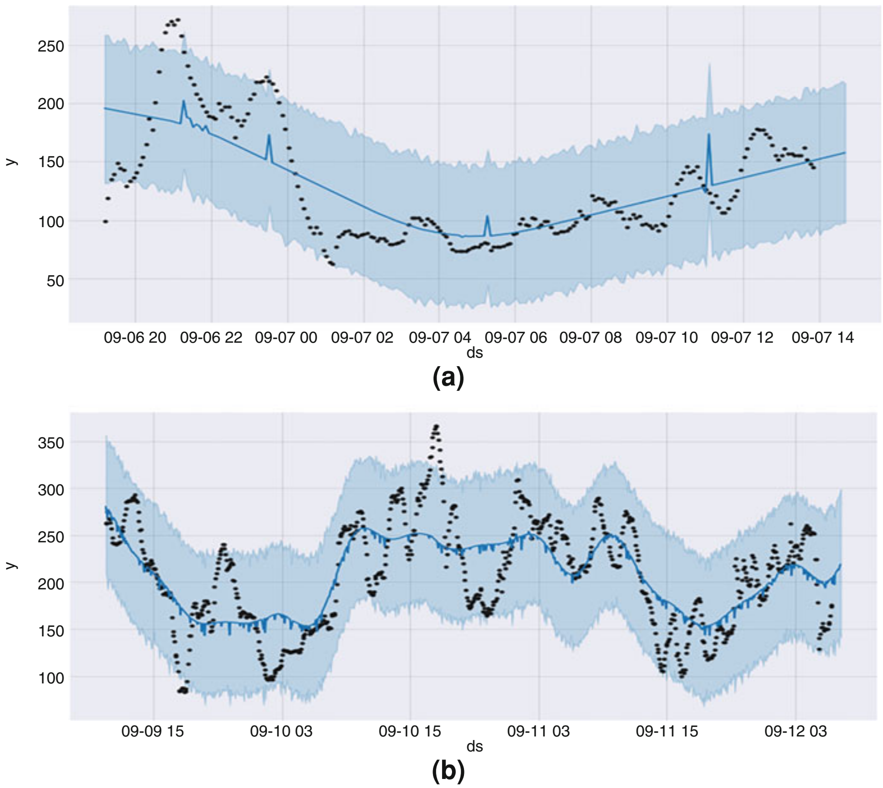


Fig. 6 Forecast graph for (**a**) patient 1 and (**b**) patient 2

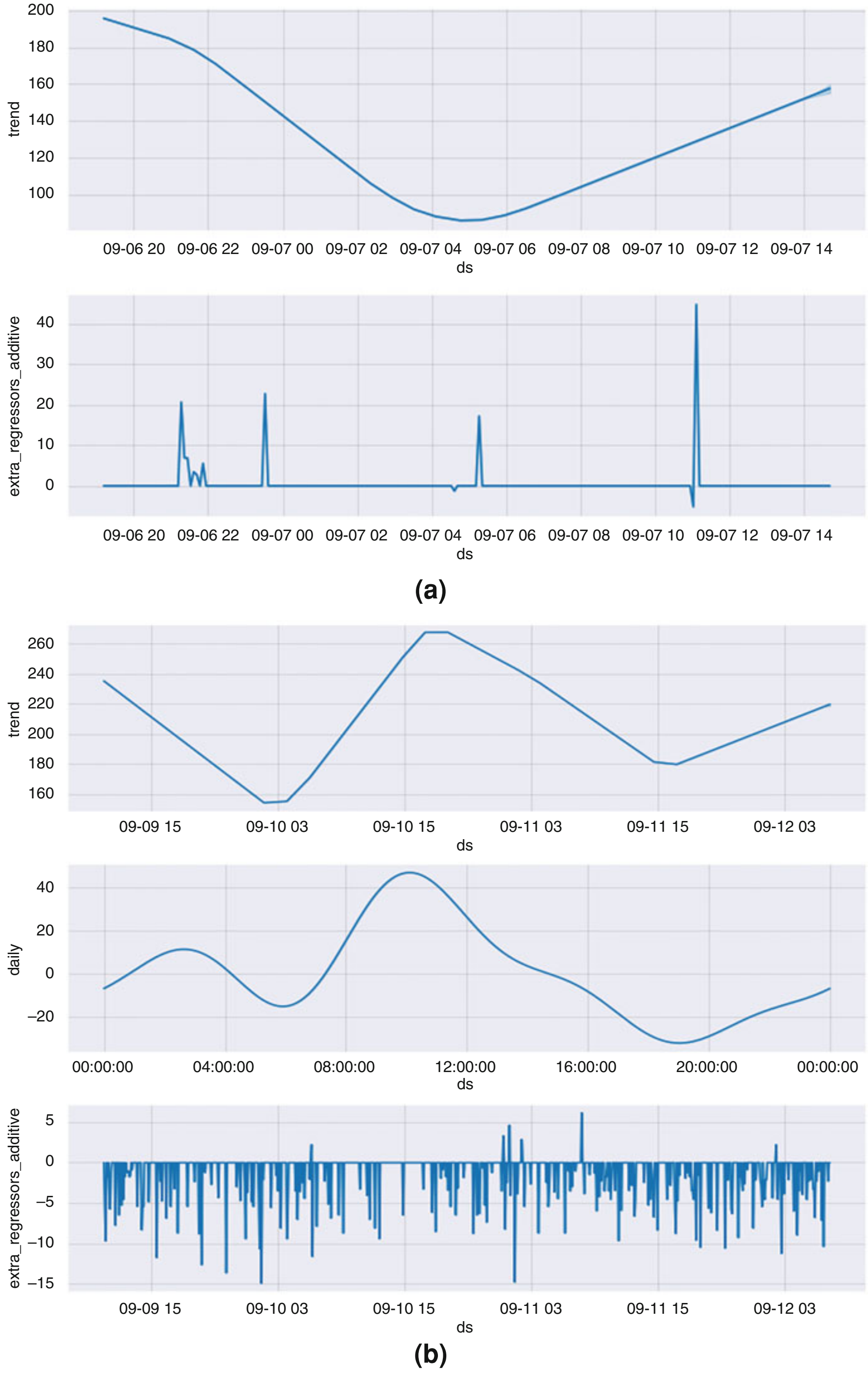


Fig. 7 Trend and additive regressor plots for (**a**) patient 1 and (**b**) patient 2

Interpreting the Results

After fitting and evaluating the forecasting model, we observe the following results:

– The Prophet model has achieved a lower mean absolute error for both the patients as compared to ARIMA model.
– The trend plot indicates that the glucose level for both the patients significantly reduces within sleep hours and starts to rise around the usual waking up hours.
– The presence of carbohydrate intake suggests a level increase in glucose level based on the amount of carbohydrates taken.
– Beta, that is, the weights array for insulin and carbohydrates for patient 1 are [−0.00472645, 0.0025285], respectively.
– Beta, the coefficients array for the regressors insulin, and carbohydrates for patient 2 are [0.04330589, −0.04525362, −0.00358408, 0.04263081, 0.02553079, 0.00023219, −0.00989248, −0.01624133, −0.00959426, 0.000421].

5 Conclusion and Future Trends

Diabetes is a disease which is caused due to high/low level of glucose in blood leading to high blood pressure, inability to heal properly, and various other problems. It would be advantageous if the future trends of hyperglycemia and hypoglycemia be predicted. In this study, focus has been on forecasting the diabetic condition of the patient within the next 50 min by creating a model with past data of the patient. Two time series models FbProphet and ARIMAX have been applied on the dataset of patients procured from Nightscout foundation. The results indicate that the models have been able to fit the minute-wise glycemic values taking into consideration the meal and insulin values as extraregressors. The performance of the model has been performed using mean absolute error. The FbProphet significantly outperforms the ARIMAX model. Various required tests like ADF, ACF, and PACF have been done as part of data preprocessing prior to applying the model.

In future, the model can be improved by incorporating activity related data and heart rate using wearable sensors and smartwatches. A much more user-friendly solution can be implemented to take the nutritional values of food (not just the carbohydrates) to gain a better insight over the fluctuations of glucose levels. There is also the scope of using training and forecasting on much more advanced time series models like recurrent neural network-based LSTMs.

References

1. S. Oviedo, J. Vehí, R. Calm, J. Armengol, A review of personalized blood glucose prediction strategies for T1DM patients. Intl. J. Num.Methods Biomed. Eng. **33**(6) (2017)
2. S.R. Chandran, R.A. Vigersky, A. Thomas, L.L. Lim, J. Ratnasingam, A. Tan, D.S.L. Gardner, Role of Composite Glycemic Indices: A Comparison of the Comprehensive Glucose Pentagon Across Diabetes Types and HbA1c Levels. Diabetes Technology & Therapeutics **22**(2) (2020)
3. WHO Diabetes (2020, June). Retrieved from https://www.who.int/news-room/fact-sheets/detail/diabetes
4. IDF Diabetes Atlas (2019). Retrieved from https://diabetesatlas.org/en/sections/worldwide-toll-of-diabetes.html
5. S.A. Kaveeshwar, J. Cornwall, The current state of diabetes mellitus in India. Australasian. Med. J. **7**(1), 45–48 (2014)
6. K. Plis, R. Bunescu, C. Marling, J. Shubrook, F. Schwartz, A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management. AAAI Workshops, North America. (June 2014)
7. J. Liszka-Hackzell, Prediction of blood glucose levels in diabetic patients using a hybrid AI technique. Comput. Biomed. Res. **32**(2), 132–144 (1999)
8. I. Rodríguez-Rodríguez, J.-V. Rodríguez, A. González-Vidal, M.-Á. Zamora, Feature selection for blood glucose level prediction in Type 1 diabetes mellitus by using the sequential input selection algorithm (Sisal). Symmetry **11**(9), 1164–1175 (2019)
9. C. Pérez-Gandía, A. Facchinetti, G. Sparacino, C. Cobelli, E.J. Gómez, M. Rigla, Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. Diabet. Technol. Therapeut. **12**(1), 81–90 (2010)
10. S.M. Pappada, B.D. Cameron, P.M. Rosman, Development of a neural network for prediction of glucose concentration in Type 1 diabetes patients. J. Diabet. Sci. Technol. **2**(5), 792–801 (2008)
11. A. Hayeri, Predicting Future Glucose Fluctuations Using Machine Learning and Wearable Sensor Data. Diabetes **67**(Suppl 1), 738-P (2018)
12. F. Bagherzadeh-Khiabani, A. Ramezankhani, F. Azizi, F. Hadaegh, E.W. Steyerberg, D. Khalili, A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results. J. Clin. Epidemiol. **71**, 76–85 (2016)
13. M.W. Aslam, Z. Zhu, A.K. Nandi, Feature generation using genetic programming with comparative partner selection for diabetes classification. Expert Syst. Appl. **40**(13), 5402–5412 (2013)
14. S. Habibi, M. Ahmadi, S. Alizadeh, Type 2 diabetes mellitus screening and risk factors using decision tree: results of data mining. Global J. Health Sci. **75**(5), 304–310 (2015)
15. A. Ramezankhani, O. Pournik, J. Shahrabi, F. Azizi, F. Hadaegh, D. Khalili, The impact of oversampling with SMOTE on the performance of 3 classifiers in prediction of Type 2 diabetes. Med. Decis. Mak. **36**(1), 137–144 (2016)
16. L.M. Albers Dj, B. Gluckman, H. Ginsberg, G. Hripcsak, Personalized glucose forecasting for Type 2 diabetes using data assimilation. Plos Computat. Biolog. **13**(4), 1–10 (2017)
17. D.Y. Faruqui Sha, R. Meka, A. Alaeddini, C. Li, S. Shirinkam, J. Wang, Development of a deep learning model for dynamic forecasting of blood glucose level for Type 2 diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR mHealth uHealth **7**(11), e14452 (2019)
18. I. Contreras, S. Oviedo, M. Vettoretti, R. Visentin, J. Vehí, Personalized blood glucose prediction: a hybrid approach using grammatical evolution and physiological models. Plos One **12**(11), e0187754 (2017)
19. J. Martinsson, A. Schliep, B. Eliasson, Blood glucose prediction with variance estimation using recurrent neural network. J. Healthcare Inform. Res **4**, 1–18 (2020)

20. H.N. Mhaskar, S.V. Pereverzyev, M.D. Van Der Walt, A Deep Learning Approach to Diabetic Blood Glucose Prediction. Frontiers in Applied Mathematics and Statistics (2017). https://doi.org/10.3389/fams.2017.00014
21. Q. Liu, X. Liu, B. Jiang, Y.W. And, Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC Infect. Dis. Article number: **218** (2011)
22. E. Dan, Modelling and Forecasting Malaria Mortality Rate using SARIMA Models (A Case Study of Aboh Mbaise General Hospital, Imo State Nigeria). Science Journal of Applied Mathematics and Statistics. **2**, 31. (2014). 10.11648/j.sjams.20140201.15
23. Pan, Y., Zhang, M., Chen, Z., Zhou, M., Zhang, Z., An ARIMA based model for forecasting the patient number of epidemic disease. International Conference on Services Systems and Services Management, ICSSSM (2016), p. 78–81
24. A. Domenico Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, M. Ciccozzi, Application of the ARIMA model on the COVID2019 epidemic dataset. Data Brief **29**, 2352–3409 (2020)
25. F. Tak-chung, A review on time series data mining. Eng. Appl. Artif. Intell. **24**(1), 164–181 (2011)
26. J.J. Van Wijk, E.R. Van Selow, Cluster and calendar based visualization of time series data trend, cycle and seasons. Proc. 1999 IEEE Symp. Inform. Visual., 1–6 (1999)
27. D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root. J. Econom. **54**(3), 159–178 (1992)
28. I. Rodríguez-Rodríguez, I. Chatzigiannakis, J.V. Rodríguez, M. Maranghi, M. Gentili, Z.-I. Má, Utility of big data in predicting short-term blood glucose levels in Type 1 diabetes mellitus through machine learning techniques. Sensors (Basel) **19**(20), 4482 (2019)
29. M. Frandes, B. Timar, R. Timar, Chaotic time series prediction for glucose dynamics in Type 1 diabetes mellitus using regime-switching models. Scientific Reports **7**, 6232 (2017, July 24)
30. M. Villani, A. Earnest, N. Nanayakkara, K. Smith, B. de Courten, S. Zoungas, Time series modelling to forecast prehospital EMS demand for diabetic emergencies. BMC Health Ser. Res. **17**(332), 1–9 (2017)
31. J.B. Albu, N. Sohler, R. Li, X. Li, E. Young, E.W. Gregg, D. Ross-Degnan, An interrupted time series analysis to determine the effect of an electronic health record-based intervention on appropriate screening for Type 2 diabetes in urban primary care clinics in New York City. Diabetes Care **40**, 1058–1064 (2017)
32. J. Yang, L. Li, Y. Shi, X. Xie, An ARIMA model with adaptive orders for predicting blood glucose concentrations and hypoglycemia. IEEE J. Biomed. Health Inform. **23**(3), 1251–1260 (2018)
33. Nightscout Diabetes Dataset (2018). Retrieved from https://github.com/nightscout
34. G.E.P. Box, G.M. Jenkins, *Time Series Analysis, Forecasting and Control* (Holden-Day, San Francisco, 1976)
35. G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis Forecasting and Control*, 4th edn. (Wiley, Hoboken, 2011)
36. S.J. Taylor, B. Letham, Forecasting at scale. Am. Statistician **72**(1), 37–45 (2018)
37. B.G. Tzovaras, M. Angrist, K. Arvai, M. Dulaney, V. Estrada-Galiñanes, B. Gunderson, T. Head, D. Lewis, O. Nov, O. Shaer, Open humans: a platform for participant-centered research and personal data exploration. GigaScience **8**(6) (2019)

Machine Learning Applications in Anti-cancer Drug Discovery

Aman Sharma and Rinkle Rani

1 Introduction

Traditional drug discovery pipelines are complex and inefficient. Worldwide various researchers are trying to shorten the drug discovery cycle to fight against many stringent diseases. Traditionally researchers were only using statistical and clinical methods for drug discovery [1]. Recently, researchers are using computational approaches to reduce drug discovery time [2, 3]. Machine learning approaches provide a powerful set of tools that can help in developing decision support systems. Such systems help in early prognosis and diagnosis of diseases such as cancer, diabetes, Parkinson's, etc. Their applications include drug response prediction, drug-target identification, biomarker identification, and drug synergy prediction. Cancer is considered a very stringent and complex disease worldwide. Scientists/researchers are trying their hard to find potential drugs or drug combinations that could help to fight against diseases such as cancer which is the leading cause of death worldwide [2, 3]. Figure 1 shows the central dogma of biology.

According to the study [4] the new drugs are produced at a constant rate during the past 60 years. Moreover, the Tufts Centre for the Study of Drug Development (CSDD) reported that the overall cost involved in developing a newly approved drug is about $2,558 million and time is about one decade. Such study focuses on the attention of researchers to develop computationally efficient drug discovery pipelines. Machine learning and deep learning methods came up as a breakthrough

A. Sharma (✉)
JUIT, Wakhnaghat, Solan, India

R. Rani
T.I.E.T, Patiala, India
e-mail: raggarwal@thapar.edu

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_10

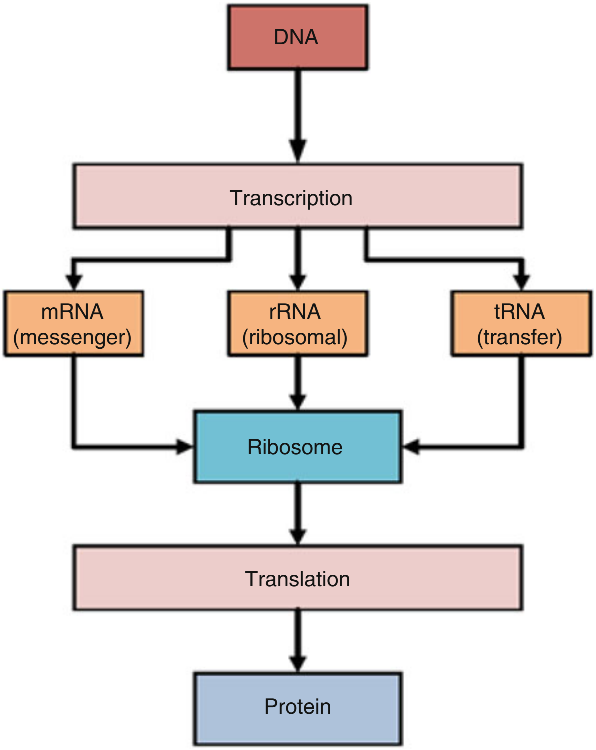


Fig. 1 Central dogma of biology [82]

in drug discovery research [5, 6]. Deep learning applications involve various hidden layers which provide data abstraction and help in data preprocessing and feature extraction [4].

All the machine learning approaches are data-driven and hence help in developing predictive modeling by utilizing the hidden correlations and patterns in data. Figure 2 shows the genome data representation for machine learning approaches. Machine learning approaches can be categorized as reinforcement, supervised, unsupervised, and semi-supervised learning. The major difference in these approaches is the quantity of information that is fed into the model which lays the basis for model training. Chemical researchers have extensively utilized machine learning capabilities especially supervised learning in anti-cancer research [7]. Supervised machine learning algorithms use target labels for training the input data and approximately predicting the output. Artificial neural networks (ANNs) and kernel methods are popular supervised learning algorithms used to transform the input space into a new feature space [8]. One of the most important features of ANNs is feature transformation using various input layers. On the other, hand kernel methods help in identifying non-linear relationships present within the data. Kernel methods utilize kernel function to perform non-linear data transformations

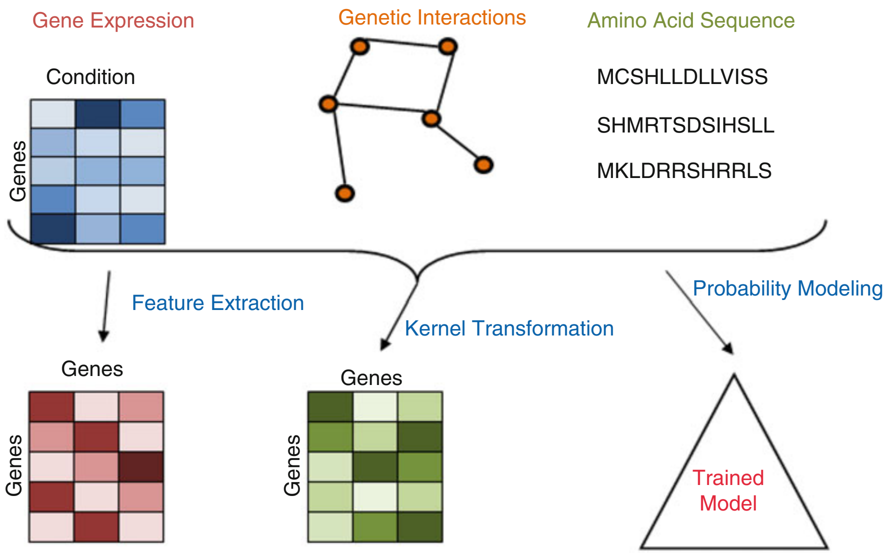


Fig. 2 Genome data representation for machine learning approaches

to use with linear algorithms. There are widespread applications of ANNs and deep learning (DL) algorithms in biomedical and drug discovery. Gene expression data and microarray data are used for developing anti-cancer drugs and biomarker prediction machine learning models.

The influence of genes in different types of cancer promoted the genetic data-driven research. Complex microenvironment results in difficult diagnosis and treatment of various cancer types. Even patients with similar type of tumor show varying responses toward the same type of drug treatment [9]. Although traditional machine learning algorithms [10] are quite helpful in developing biomedical computational models, recently we have seen a rise in deep learning algorithms. The major reason for such a sudden drift is because of the large availability of biomedical and pharmacogenomic datasets [11, 12] and high computing machines for parallel processing such as GPUs.

2 Background

In this section, we discussed the machine learning-based cancer applications. The essence of all the machine learning models is the high-quality data that we feed for training. From the last two decades, many researchers and consortiums have contributed to the field of drug discovery by providing high-quality chemical and biological data. PubChem [13] is one of the largest open-source chemical repositories. It provides the facility to search chemicals by name or structure. We

Table 1 Selected startup companies in the field of drug discovery

S. no.	Company name	Website
1.	Amplion	https://bit.ly/2PjOY94
2.	BioSymetrics	https://bit.ly/2Xk14U5
3.	Biorelate	https://bit.ly/3fpvdY2
4.	Causaly	https://bit.ly/2Xle1gf
5.	Data2Discovery	https://bit.ly/31aEkGX
6.	Data4Cure	https://bit.ly/3fgJ5Uw
7.	Elucidata Corporation	https://bit.ly/3gpqzdS
8.	Evid Science	https://bit.ly/3i0gCnC
9.	Genialis	https://bit.ly/39SnqB8
10.	HelixAI	https://bit.ly/3glrz2V
11.	Innoplexus	https://bit.ly/3fjG8CM
12.	Intellegens	https://bit.ly/3k48rZq
13.	InveniAI	https://bit.ly/30pX9XP
14.	Mozi	https://bit.ly/3190p8Z
15.	PatSnap	https://bit.ly/33dLEV4

can get the physical, biological, chemical, and toxicity data of various chemical compounds. ChEMBL [14] is a collection of bioactive drug-like small molecules.

It contains data corresponding to 2D structure and properties of bioactive drugs. The database is majorly curated and abstracted from the literature of modern drug discovery. The data for bioactivity of the drug molecule is provided in normalized form. Further, web links for the research studies corresponding to drugs are included in the database. The DrugBank [15] is the freely available database consisting of data about a wide range of drugs and their corresponding targets. The DrugBank combines data for two research domains: bioinformatics and chemoinformatics. It is like an encyclopedia for getting information and detailed description regarding various chemical compounds and their corresponding targets. It is a widely adopted resource by various pharmacists, physicians, researchers, and the drug industry. Table 1 contains selected startup companies in the field of drug discovery.

DrugCentral [16] is an online repository for information on various drugs. It contains information such as mode of action for drugs and active ingredients in chemical products. It also contains information regarding discontinued and approved drugs outside the USA. SuperDRUG2 [17] is one of the largest databases consisting of approved/marketed drugs and chemical ingredients. 2D and 3D structures, physicochemical properties, and pharmacokinetic data of drugs are also provided in the database. Along with this, it contains data for drug-drug and drug-target interactions. The GDSC [18] database is developed to improve cancer biomarker prediction and drug-target prediction. Informative data is provided corresponding to genomic variations when different cell lines are perturbed with drugs. The CCLE [12] database is a result of a collaborative effort by various drug discovery research labs. It contains 1870 RNA sequencing, 654 whole exome sequences, and 46 whole genome sequence files. Various researchers are trying their

Table 2 Selected publicly available online databases for drug discovery

S. no.	Database	Online access
1.	PubChem [13]	https://bit.ly/39MxdZc
2.	ChEMBL [14]	https://bit.ly/3k6kvcs
3.	DrugBank [15]	https://bit.ly/3hRKPFw
4.	DrugCentral [16]	https://bit.ly/2PfMTL0
5.	SuperDRUG2 [17]	https://bit.ly/2EJHdXV
6.	GDSC [18]	https://bit.ly/31bxIbp
7.	CCLE[12]	https://bit.ly/3fkagOf
8.	repoDB [19]	https://bit.ly/33oSHdL

hard to extract meaningful insights from CCLE using microarray data analysis. Table 2 contains selected publicly available online databases for drug discovery.

Drug Repurposing

The availability of freely downloadable healthcare datasets motivated the researchers to apply machine learning algorithms for predicting drug responses, biomarkers, signaling pathways, drug synergism, etc. Figure 3 describes the Gaussian kernel and multi-task learning used for anti-cancer drug response prediction. The Bayesian model has been used by Sean Ekins et al. [20] for compound selection. In the proposed technique, they have used bioinformatics as well as chemoinformatics data. Various researchers have also used machine learning models for ligand-based virtual screening (LBVS) [21]. Naive Bayes algorithm is also prominently used for predicting toxicity and biological pathways for anti-cancer drug prediction [22]. Kuang Z et al. [23] have presented a regularization-based technique for drug repurposing. Their statistical analysis suggests various drugs that can be repurposed in varying biological situations. Patrick MT et al. [24] have implemented an approach for summarizing drug information from 20 million research articles. They trained their model on various stringent diseases such as psoriasis, alopecia areata, and immune-mediated diseases to obtain the drug repurposing opportunities. Zeng X et al. [25] have proposed a deep learning approach for computational drug repurposing. Their proposed approach exploits data from various networks such as drug-disease, drug-target, and drug-drug networks. The proposed model was trained in Alzheimer's and Parkinson's disease. Kim E et al. [26] have developed the machine learning-based approach for predicting the hidden pharmacological benefits of herbal compounds. The common assumption that all the researchers assume while developing computational approaches for drug repurposing is that similar diseases can be treated with similar drugs. However, similarity can be defined in terms of drug-drug, tissue-tissue, and disease-disease similarity. Table 3 contains selected publicly available online databases for drug repurposing.

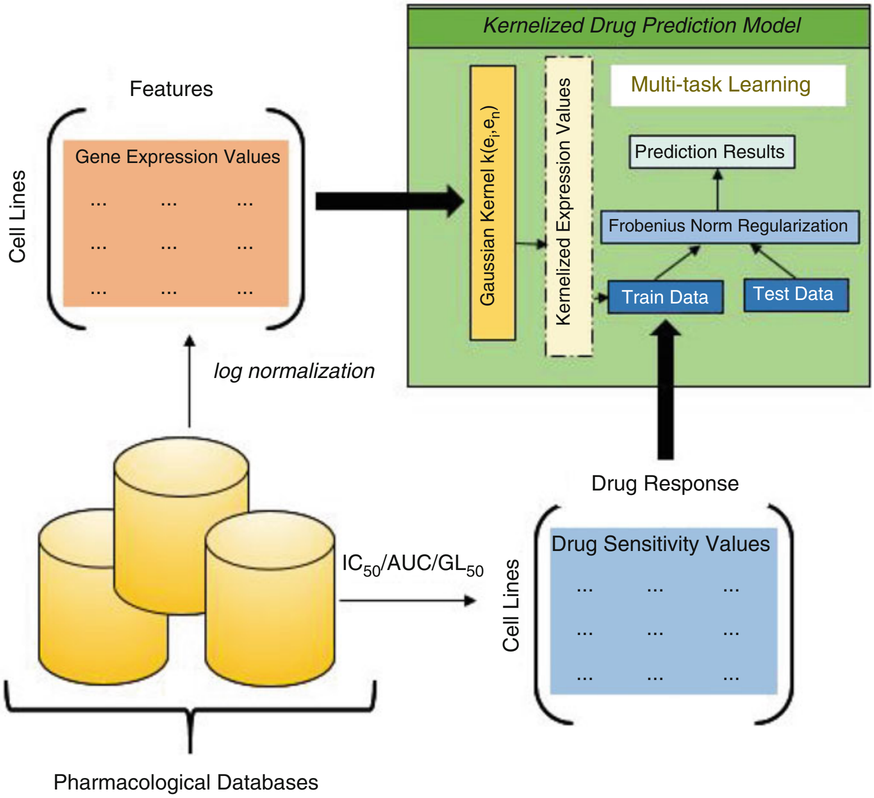


Fig. 3 Gaussian kernel and multi-task learning used for anti-cancer drug response prediction

Table 3 Selected publicly available online databases for drug repurposing

S. no.	Database	Online access
1.	NCI-DREAM 7	https://bit.ly/2Pfj4Kz
2.	NCI-60 [27]	https://bit.ly/33lCxBE
3.	TCGA [28]	https://bit.ly/39QPapq
4.	TCPA [29]	https://bit.ly/33hI4cs
5.	GDSC[18]	https://bit.ly/33jtyAU
6.	CCLE [12]	https://bit.ly/2DrT1NI

Cancer Classification

Gene selection is a challenging process in microarray data analysis. Although a lot of research has been done on identifying genomic biomarkers for different types of cancer, still no generic pipeline has been designed for cancer classification. Various algorithms/approaches have been proposed in the literature to identify relevant and potential genomic biomarkers. These algorithms can be broadly classified as wrapper, filter, and hybrid methods for gene selection [30, 31]. The filter method is defined by statistical analysis and properties of the dataset for obtaining the best optimal gene subset. Ranking of genes is performed using different types of statistical methods [32, 33]. Genes that score relatively higher rank are considered for further analysis. The methods included in this category are T-test [33], max-min correntropy [34], and information gain [35]. For detailed information on such methods, one can consider a survey on filter methods for gene selection [36]. The wrapper method relies on some kind of evolutionary technique to optimally search the relevant feature subset. In wrapper methods, random initialization of the population is done consisting of the subset of features. The fitness of each subset is obtained using an appropriate fitness evaluator. Iteratively the whole process is repeated several times to fetch the optimal solution. Such methods include the use of a genetic algorithm [37], artificial bee colony algorithm [38], bat algorithm [39], and swarm optimization [40] for gene selection. Table 4 contains selected publicly available online datasets for cancer classification (Fig. 4).

Hybrid methods are also evolutionary-based methods, but they use filter methods in the initial phase for the screening of the most promising genes from the

Table 4 Selected publicly available online datasets for cancer classification

S. no.	Datasets	Online access
1.	SRBCT cancer	Khan et al. [41]
2.	Leukemia cancer	Golub et al. [42]
3.	Prostate cancer	Singh et al. [43]
4.	Breast cancer	Hedenfalk et al. [44]
5.	Breast cancer	https://bit.ly/2XmvQLM
6.	Central nervous system cancer	https://bit.ly/33lCPse
7.	GSE25136	https://bit.ly/2PjA2HL

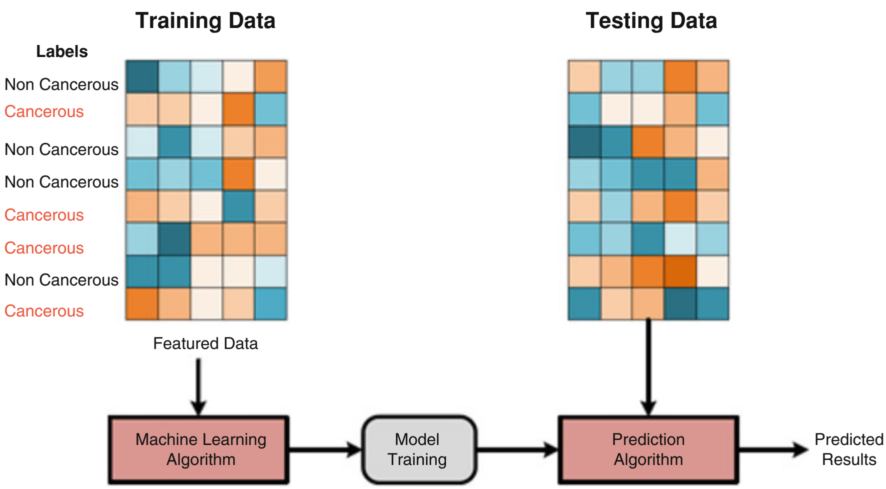


Fig. 4 Example of cancer classification using machine learning application [82]

microarray dataset. These methods use filter methods to reduce the running time of an algorithm. Some of these approaches include the chi-square test with GA [45], mRMR with GA [46], and similarity scheme with ABC [47]. Apart from these methods, some approaches integrate the feature selection task along with classification. It selects the feature subset, builds a classifier, and then checks the classifier accuracy. If performance is not appropriate, then it removes the poor genes and builds the classifier again iteratively. Such methods are known as embedded methods [48]. Most of these methods are difficult to replicate and are computationally expensive. Most of these methods are not able to optimally utilize high-dimensional gene expression data and suffer from overfitting.

Drug Synergy Prediction

Targeted drug therapy is the most commonly used treatment given to cancer patients. These drugs are specially designed based on their targets which help to suppress cancer. These targets are known as anti-oncogene which is responsible for tumor suppression by suppressing mitosis (cell division) [49]. Any alteration and changes in these genes lead to uncontrollable cell growth. Unlike these genes, some oncogenes promote tumor growth. Most of the targeted drug therapies are designed considering oncogenes as anti-oncogenes are hard to target. Various studies revealed the resistance of targeted drug therapies, which hence results in nonresponsive drug behavior [50, 51]. This resistance may have occurred because of many reasons such as cell death inhibition, change in drug targets, etc. Heterogeneous tumor microenvironment can also result in drug resistance [42]. Combination drug therapy is a good option to avoid drug resistance. It helps in overcoming the drug resistance by delaying tumor growth. It includes the usage of multiple drugs in fixed dose proportion and as a single-dose formulation. Combination drug therapy is showing excellent results in tumor suppression by reducing the chances of multiple mutations [52] and a single mutation [53] that can escape all the drugs. Additionally, combination therapy helps in lowering drug dosage and side effects [52]. A combination of two or more drugs is considered effective if the tumor suppression rate of combination is higher than individual drugs. Such a combination of drugs is known as synergistic drugs otherwise antagonistic. The proposition of dose also matters in drug synergy; we cannot mix them in any random proportions.

Combination drug therapy is widely used in treatment of various diseases such as HIV and cancer and many more diseases [54, 55]. Combination drug therapy becomes more essential for complex diseases such as cancer because of the involvement of multiple growth pathways in such diseases [54]. However, there is risk of toxicity with combination therapies, which can be handled with appropriate quantity of dosage. Many combination drugs are already approved by FDA for treating stringent diseases. For example, aspirin and dipyridamole are used in combination to reduce the risk of stroke [56], and sabarubicin and cisplatin are used in lung cancer [57]. The Drug Combination Database (DCDB) provides information about 330 FDA-approved and 1033 investigational drug combinations [58]. Drug combination works on the principle of synergy effect, which means that the overall effect of combination drug therapy is more as compared to individual drugs. Although it is observed that there is benefit of drug combination therapy, quantification of drug synergy effect is still one of the challenging tasks. Many researchers have proposed different methods such as Chou-Talalay method [59], Loewe additivity [60], and Bliss independence [61] to calculate dose-response effect of combination drugs. These techniques are based on comparison between expected and observed combination drug responses [62, 63]. Till now, for most of the diseases, drug combinations are identified based on clinical trials. But clinical trials using "trial and error" is a labor- and cost-intensive and time-consuming task [64]. Another disadvantage of clinical trials is unwanted exposure of harmful chemicals to patients [65]. Apart from clinical trials, high-throughput screening (HTS) is also used in identifying potential drug combinations [66]. In such screenings, different concentrations of drugs are used to identify potential drug combinations, but still they are not accurate enough to capture the real microenvironment [67].

Although all these methods play a crucial role in quantifying drug synergy, still there are many issues with these methods such as no method can quantify drug synergism in different feasible situations [62] and different dose-response methods can even produce different results [68]. Moreover, these methods are based on screening of all the possible drug combinations, which is impractical and time-consuming. In such a situation, identification of effective drug combinations is a challenging task. Many researchers are using machine learning and computational models to predict potential drug combinations for various diseases [69, 70].

3 Research Gaps in Computational Drug Discovery

The last two decades has seen a tremendous awareness and growth toward cancer research. Many researchers, clinicians, and academicians are trying their hard to fight against cancer. We have discussed the various already proposed computational drug discovery approaches and applications in Sect. 2. But still, there are many issues/research gaps left that need to be worked upon. Most of the techniques are crudely based on statistics which limit the utility of the applications only to statisticians. Hence, there is a need to develop user-friendly applications, which are provided by machine learning algorithms. The following research gaps still exist:

1. High dimensionality of data is one of the major issues while developing applications from genomic data. Moreover, there is an issue of imbalance class; there is a majority of one class due to lack of samples of other class. Although various techniques have been developed, still no approach has been developed which covers a wide range of applications. Some techniques are good in one situation and others in different situations. So it is more or like hit and trial method.

2. Existing approaches are developed using binary imbalanced datasets. There is a need to test those applications on multi-class imbalanced datasets. With the increase in the severity of the genetic disease, their subtypes also increase. Hence multi-class classification is required to predict the correct subtype of the disease.
3. Heterogeneity in the genetic structure of cancer patients results in heterogeneity in their drug responses. Earlier drugs were discovered based on the anatomical region of the disease, but cancer is a genetic disease, so any anti-cancer drug discovery needs to consider genetic influence while developing new drugs. Moreover, if any computational method is proposed for drug discovery or anti-cancer drug prediction, then it should strictly consider the genetic variations that are responsible for cancer.
4. Feature selection is one of the primary tasks in cancer classification approaches. But existing feature selection approaches are not scalable enough to handle maximum genetic aberrations simultaneously.
5. Cancer is a complex disease; we cannot generalize the drug therapies for different patients. There is a need to provide personalized therapies corresponding to an individual patient's drug sensitivity.
6. Machine learning capabilities for predicting drug synergism are unexplored. Predicting drug synergy will boost the anti-cancer drug discovery process. There is a need to extract better features for predicting drug synergy.

4 Future of Computational Drug Discovery

Deep Learning for Drug Discovery

"Artificial intelligence" as the name itself states that it is a kind of intelligence which is incorporated artificially in a system. There are various definitions of artificial intelligence, but broadly they are categorized as thinking humanly or rationally and acting humanly or rationally. It is an interdisciplinary branch of science which can be applied into molecular biology and genomics and in various other disciplines. Researchers are actively using artificial intelligence in bioinformatics for analyzing large amount of data and DNA sequencing [71]. From the last two decades, there is an enormous increase in healthcare data from researchers, academicians, and industry. This data holds enough potential to explore and fetch meaningful insights. There is a huge possibility of exploring hidden patterns and knowledge from this healthcare data using computational approaches. Although many researchers are using statistical approaches and machine learning algorithms to get meaningful insights from data, it is impossible to process healthcare datasets using conventional machine learning algorithms because of their huge volume, velocity, and variety. In such a scenario, deep learning algorithms play a very important role. Nowadays, researchers are using deep learning algorithms such as CNN [72] and RNN [73] for dealing with healthcare big datasets.

Role of Deep Learning in Cancer Classification

Existing literature on cancer classification has used traditional machine learning algorithms. However, very few approaches have been proposed using deep learning. Wang et al. [74] developed a deep learning-based technique to identify metastatic breast cancer using an image dataset. Ahmed et al. [75] have proposed a deep belief network-based for breast cancer classification. Skin cancer is very common nowadays, and it's hard to diagnose and predict their targets. To classify and identify the most promising biomarkers for skin cancer, Haofu et al. [76] proposed a classification approach using deep learning. Fakoor et al. [77] have developed an unsupervised feature selection technique for identifying and diagnosing cancer types. Arunkumar and Ramakrishnan [78] have developed the hybrid approach for feature selection. All these techniques have focused on reducing the dimensions of input dataset using feature selection approaches.

Role of Deep Learning in MicroRNA Analysis in NGS

"Big Data" has been a buzz topic in recent years, and it has gained huge interest from academics as well as industry. The rate at which data is being produced has increased to many folds and so is the research in this field. Data related to bioinformatics has also evolved over many years. An increase in computational capabilities and the emergence of HTS technology have led to the sudden outburst of biomedical data. This data serves a great potential in identifying disease biomarkers and discovering new drugs, but unfortunately, it is not effectively utilized. NGS technologies have created a serious need for new technologies and algorithms. Figure 5 shows biogenesis of microRNA. In such a scenario, deep learning using neural networks is considered an effective choice. Although ML approaches have been used for many years, they have the limitation of processing raw data. Deep learning is a new version of ML algorithms that incorporate artificial intelligence using multilayer neural networks. In contrast to traditional ML approaches, deep learning can extract features from data itself. In efforts to apply deep learning algorithms to microRNA prediction, researchers have proposed various deep learning algorithms. Seunghyun Park et al. [79] have proposed deepMiRGene, an algorithm used to predict microRNA precursor. They used RNN, because there is no need to input features manually, and the algorithm automatically identifies features from input data. This approach leads to the discovery of various new features too which can be used in future research. Similarly Cheng S. et al. developed MiRTDL [80], an algorithm for microRNA target prediction using CNN. It automatically extracts desired information from the data itself rather than relying on information fed manually. These algorithms have shown efficient results and have improved prediction results. The use of deep learning techniques in microRNA and their target prediction can help in novel microRNA predictions, and one can investigate better

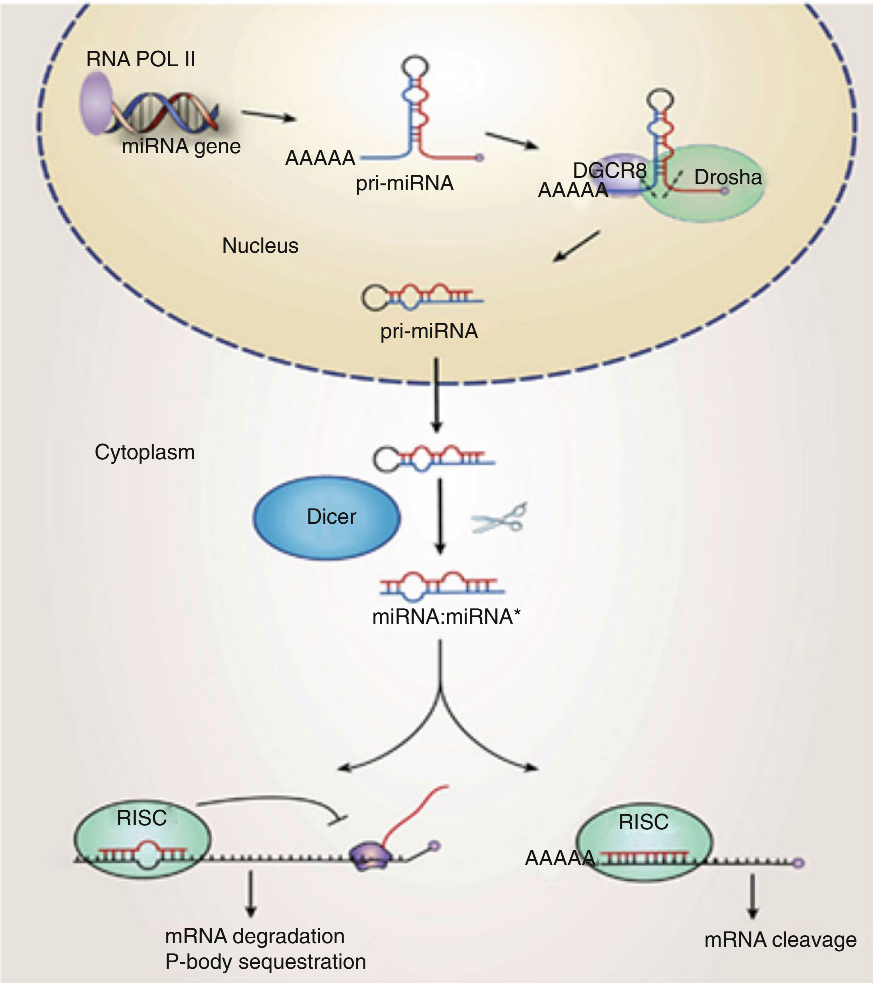


Fig. 5 Biogenesis of microRNA (Image from [81])

knowledge about the underlying mechanism. Table 5 contains selected microRNAs as potential cancer diagnostic biomarkers in blood.

5 Conclusion and Future Directions

We have seen in previous sections that various machine learning applications have been developed in the literature for anti-cancer drug discovery. But still it is a challenge to predict drugs using computational techniques which are also clinically

Table 5 Selected microRNAs as potential cancer diagnostic biomarkers in blood

S. no.	Tool/pipeline	Features	URL
1.	BioVLAB-MMIA-NGS	To find DE microRNAs and their target genes (DEGs)	https://bit.ly/3gzgN9r
2.	CAP-miRSeq	Supports sequential and parallel processing of deep sequencing microRNA data	https://mayocl.in/2DcM3fF
3.	iMir	Provides automated pipeline for microRNA data analysis	https://bit.ly/30iOHJM
4.	CPSS	Standalone tool with single data submission	https://bit.ly/30lOG7H
5.	MAGI	MicroRNA-Seq analysis using GPU technology	https://bit.ly/39LByfm
6.	miRSeqNovel	R/bioconductor pipeline package to predict novel microRNA for plant and animal microRNA	https://bit.ly/2DbmQ5s
7.	mirTools 2.0	Performs comparative analysis of experimental samples and identifies the DE microRNAs among experimental group	https://bit.ly/3flotKV
8.	MMIA	Integrates microRNA and mRNA expression data for detailed analysis	https://bit.ly/39VcJh1

efficient. Cancer is a very stringent and complex disease which needs multi-focused approach for treatment. We can't treat a patient by focusing on a single aspect of genetic behavior of individual. Multiple pathways need to be considered while developing potential treatment drugs. Drug resistance can also be targeted while designing drugs. In such cases, multiple drug combinations can be selected as a treatment option. Enormous increase in oncological datasets is also a boom to cancer research. But this data can't be mined using traditional/conventional machine learning algorithms. Deep learning algorithms should be extensively studied to utilize such big healthcare datasets.

References

1. S. Wold, L. Eriksson, S. Clementi, Statistical validation of QSAR results. Chemometric Methods in Molecular Design (Wiley-VCH, Weinheim, 1995), pp. 309–338
2. A. Sharma, R. Rani, Drug sensitivity prediction framework using ensemble and multi-task learning. Int. J. Mach. Learn. Cybern. **11**, 1231–1240 (2019)
3. A. Sharma, R. Rani, An optimized framework for cancer classification using deep learning and genetic algorithm. J. Med. Imaging Health Inform. **7**(8), 1851–1856 (2017)

4. B. Munos, Lessons from 60 years of pharmaceutical innovation. Nat. Rev. Drug Discov. **8**(12), 959–968 (2009)
5. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. nature, 521(7553), 436–444 (2015)
6. N. Stephenson, E. Shane, J. Chase, J. Rowland, D. Ries, N. Justice, . . . R. Cao, Survey of machine learning techniques in drug discovery. Curr. Drug Metab. **20**(3), 185–193 (2019)
7. K. Alberi, M.B. Nardelli, A. Zakutayev, L. Mitas, S. Curtarolo, A. Jain, . . . M. Kanatzidis, The 2019 materials by design roadmap. J. Phys. D. Appl. Phys. **52**(1), 013001 (2018)
8. M. Rupp, Von O.A. Lilienfeld, K. Burke, Guest editorial: Special topic on data-enabled theoretical chemistry (2018)
9. J. Sheng, F. Li, S.T. Wong, Optimal drug prediction from personal genomics profiles. IEEE J. Biomed. Health Inform. **19**(4), 1264–1270 (2015)
10. M.W. Libbrecht, W.S. Noble, Machine learning applications in genetics and genomics. Nat. Rev. Genet. **16**(6), 321–332 (2015)
11. W. Yang, J. Soares, P. Greninger, E.J. Edelman, H. Lightfoot, S. Forbes, . . . S. Ramaswamy, Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. **41**(D1), D955D961 (2012)
12. J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A.A. Margolin, S. Kim, . . . A. Reddy, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature **483**(7391), 603–607 (2012)
13. E.E. Bolton, Y. Wang, P.A. Thiessen, S.H. Bryant, PubChem: integrated platform of small molecules and biological activities. In Annual Reports In Computational Chemistry, vol. 4 (Elsevier, San Diego, 2008), pp. 217–241
14. A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, . . . J.P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. **40**(D1), D1100–D1107 (2012)
15. D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, . . . M. Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. **36**(suppl_1), D901–D906 (2008)
16. O. Ursu, J. Holmes, J. Knockel, C.G. Bologa, J.J. Yang, S.L. Mathias, . . . T.I. Oprea, DrugCentral: online drug compendium. Nucleic Acids Res. **45**, D932–D939 (2016)
17. A. Goede, M. Dunkel, N. Mester, C. Frommel, R. Preissner, SuperDrug: a conformational drug database. Bioinformatics **21**(9), 1751–1753 (2005)
18. M.J. Garnett, E.J. Edelman, S.J. Heidorn, C.D. Greenman, A. Dastur, K.W. Lau, . . . Q. Liu, Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature **483**(7391), 570–575 (2012)
19. A.S. Brown, C.J. Patel, A standard database for drug repositioning. Sci. Data **4**(1), 1–7 (2017)
20. S. Ekins, P.B. Madrid, M. Sarker, S.G. Li, N. Mittal, P. Kumar, . . . P. Bourbon, Combining metabolite-based pharmacophores with bayesian machine learning models for Mycobacterium tuberculosis drug discovery. PloS one **10**(10), e0141076 (2015)
21. A. Lavecchia, Machine-learning approaches in drug discovery: methods and applications. Drug Discov. Today **20**(3), 318–331 (2015)
22. H. Sun, A naive Bayes classifier for prediction of multidrug resistance reversal activity on the basis of atom typing. J. Med. Chem. **48**(12), 4031–4039 (2005)
23. Z. Kuang, Y. Bao, J. Thomson, M. Caldwell, P. Peissig, R. Stewart, . . . D. Page, A machine-learning-based drug repurposing approach using baseline regularization, in *Computational Methods for Drug Repurposing* (Humana Press, New York, 2019), pp. 255–267
24. M.T. Patrick, K. Raja, K. Miller, J. Sotzen, J.E. Gudjonsson, J.T. Elder, L.C. Tsoi, Drug repurposing prediction for immune-mediated cutaneous diseases using a word-embedding–based machine learning approach. J. Investig. Dermatol. **139**(3), 683–691 (2019)
25. X. Zeng, S. Zhu, X. Liu, Y. Zhou, R. Nussinov, F. Cheng, deepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics **35**(24), 5191–5198 (2019)
26. E. Kim, A.S. Choi, H. Nam, Drug repositioning of herbal compounds via a machine-learning approach. BMC Bioinf. **20**(10), 33–43 (2019)

27. R.H. Shoemaker, The NCI60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer **6**(10), 813–823 (2006)
28. K. Tomczak, P. Czerwińska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp. Oncol. **19**(1A), A68 (2015)
29. J. Li, Y. Lu, R. Akbani, Z. Ju, P.L. Roebuck, W. Liu, ... C. Wakefield, TCPA: a resource for cancer functional proteomics data. Nat. Methods **10**(11), 1046–1047 (2013)
30. M.W. Khan, M. Alam, A survey of application: Genomics and genetic programming, a new frontier. Genomics **100**(2), 65–71 (2012)
31. Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics. Bioinformatics **23**(19), 2507–2517 (2007)
32. V. Pihur, S. Datta, S. Datta, Finding common genes in multiple cancer types through meta–analysis of microarray experiments: a rank aggregation approach. Genomics **92**(6), 400–403 (2008)
33. Y. Qi, H. Sun, Q. Sun, L. Pan, Ranking analysis for identifying differentially expressed genes. Genomics **97**(5), 326–329 (2011)
34. N. Zhou, L. Wang, A modified T-test feature selection method and its application on the HapMap genotype data. Genomics Proteomics Bioinf. **5**(3–4), 242–249 (2007)
35. M. Mohammadi, H.S. Noghabi, G.A. Hodtani, H.R. Mashhadi, Robust and stable gene selection via maximum–minimum correntropy criterion. Genomics **107**(2–3), 83–87 (2016)
36. C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, ... A. Nowe, A survey on filter techniques for feature selection in gene expression microarray analysis. IEEE/ACM Trans. Comput. Biol. Bioinform. **9**(4), 1106–1119 (2012)
37. Y. Wang, X. Chen, W. Jiang, L. Li, W. Li, L. Yang, ... S. Wang, Predicting human microRNA precursors based on an optimized feature subset generated by GA–SVM. Genomics **98**(2), 73–78 (2011)
38. B.A. Garro, K. Rodríguez, R.A. Vázquez, Classification of DNA microarrays using artificial neural networks and ABC algorithm. Appl. Soft Comput. **38**, 548–560 (2016)
39. M. Dashtban, M. Balafar, P. Suravajhala, Gene selection for tumor classification using a novel bio-inspired multi-objective approach. Genomics **110**(1), 10–17 (2018)
40. S. Kar, K.D. Sharma, M. Maitra, Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique. Expert Syst. Appl. **42**(1), 612–627 (2015)
41. J. Khan, J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, ... P.S. Meltzer, Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat. Med. **7**(6), 673–679 (2001)
42. T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, ... C.D. Bloomfield, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286**(5439), 531–537 (1999)
43. D. Singh, P.G. Febbo, K. Ross, D.G. Jackson, J. Manola, C. Ladd, ... E.S. Lander, Gene expression correlates of clinical prostate cancer behavior. Cancer Cell **1**(2), 203–209 (2002)
44. I. Hedenfalk, D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon, ... Z. Yakhini, Gene-expression profiles in hereditary breast cancer. N. Engl. J. Med. **344**(8), 539–548 (2001)
45. C.P. Lee, Y. Leu, A novel hybrid feature selection method for microarray data analysis. Appl. Soft Comput. **11**(1), 208–213 (2011)
46. El A. Akadi, A. Amine, A. El Ouardighi, D. Aboutajdine, A two-stage gene selection scheme utilizing MRMR filter and GA wrapper. Knowl. Inf. Syst. **26**(3), 487–500 (2011)
47. E. Hancer, B. Xue, D. Karaboga, M. Zhang, A binary ABC algorithm based on advanced similarity scheme for feature selection. Appl. Soft Comput. **36**, 334–348 (2015)
48. X. Chen, H. Ishwaran, Random forests for genomic data analysis. Genomics **99**(6), 323–329 (2012)
49. Y. Lu, J. Han, Cancer classification using gene expression data. Inf. Syst. **28**(4), 243–268 (2003)

50. O. Okun, Survey of novel feature selection methods for cancer classification, in *Biological Knowledge Discovery Handbook: Preprocessing, Mining, and Postprocessing of Biological Data*, vol. 379, ed. by M. Elloumi, A.Y. Zomaya (Wiley-Blackwell, Chichester, 2013)
51. Y. Bengio, *Learning Deep Architectures for AI* (Now Publishers Inc, 2009)
52. U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, A.J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. **96**(12), 6745–6750 (1999)
53. L.J. Van't Veer, H. Dai, M.J. Van De Vijver, Y.D. He, A.A. Hart, M. Mao, ... G.J. Schreiber, Gene expression profiling predicts clinical outcome of breast cancer. Nature **415**(6871), 530–536 (2002)
54. A.W. Tolcher, L.D. Mayer, Improving combination cancer therapy: the CombiPlex® development platform. Future Oncol. **14**(13), 1317–1332 (2018)
55. T.A. Yap, A. Omlin, J.S. De Bono, Development of therapeutic combinations targeting major cancer signaling pathways. J. Clin. Oncol. **31**(12), 1592–1605 (2013)
56. T.L. Lenz, D.E. Hilleman, Aggrenox: a fixed-dose combination of aspirin and dipyridamole. Ann. Pharmacother. **34**(11), 1283–1290 (2000)
57. M. Bigioni, A. Benzo, C. Irrissuto, G. Lopez, B. Curatella, C.A. Maggi, ... M. Binaschi, Antitumour effect of combination treatment with Sabarubicin (MEN 10755) and cis-platin (DDP) in human lung tumour xenograft. Cancer Chemother. Pharmacol. **62**(4), 621–629 (2008)
58. Y. Liu, B. Hu, C. Fu, X. Chen, DCDB: drug combination database. Bioinformatics **26**(4), 587–588 (2010)
59. J.C. Ashton, Drug combination studies and their synergy quantification using the Chou–Talalay method. Cancer Res. **75**(11), 2400–2400 (2015)
60. S. Loewe, Die mischarznei. Klinische Wochenschrift **6**(23), 1077–1085 (1927)
61. C.I. Bliss, The toxicity of poisons applied jointly. Ann. Appl. Biol. **26**, 585–615 (1939)
62. N.J. Sucher, Searching for synergy in silico, in vitro and in vivo. Synergy **1**(1), 30–43 (2014)
63. T.C. Chou, Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. Pharmacol. Rev. **58**(3), 621–681 (2006)
64. D. Day, L.L. Siu, Approaches to modernize the combination drug development paradigm. Genome Med **8**(1), 115 (2016)
65. K. Pang, Y.W. Wan, W.T. Choi, L.A. Donehower, J. Sun, D. Pant, Z. Liu, Combinatorial therapy discovery using mixed integer linear programming. Bioinformatics **30**(10), 1456–1463 (2014)
66. L. He, E. Kulesskiy, J. Saarela, L. Turunen, K. Wennerberg, T. Aittokallio, J. Tang, Methods for high-throughput drug combination screening and synergy scoring, in *Cancer Systems Biology* (Humana Press, New York, 2018), pp. 351–398
67. D. Ferreira, F. Adega, R. Chaves, The importance of cancer cell lines as in vitro models in cancer methylome analysis and anticancer drugs testing. Oncogenomics and cancer proteomics-novel approaches in biomarkers discovery and therapeutic targets in cancer (2013), pp. 139–166
68. C.D. Doern, When does 2 plus 2 equal 5 A review of antimicrobial synergy testing. J. Clin. Microbiol. **52**(12), 4124–4128 (2014)
69. J. Wildenhain, M. Spitzer, S. Dolma, N. Jarvik, R. White, M. Roy, ... M. Tyers, Prediction of synergism from chemical-genetic interactions by machine learning. Cell Syst. **1**(6), 383–395 (2015)
70. J.D. Janizek, S. Celik, S.I. Lee, Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. bioRxiv, 331769 (2018)
71. P.J. Brézillon, P. Zaraté, F. Saci, Artificial intelligence approach in analysis of DNA sequences. Biochimie **75**(5), 337–345 (1993)
72. H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, ... R.M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging **35**(5), 1285–1298 (2016)
73. S. Kombrink, T. Mikolov, M. Karafiát, L. Burget, Recurrent neural network based language modeling in meeting recognition, in *Twelfth Annual Conference of the International Speech Communication Association* (2011)

74. D. Wang, A. Khosla, R. Gargeya, H. Irshad, A.H. Beck, Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718 (2016)
75. A.M. Abdel-Zaher, A.M. Eldeib, Breast cancer classification using deep belief networks. Expert Syst. Appl. **46**, 139–144 (2016)
76. H. Liao, A deep learning approach to universal skin disease classification. University of Rochester Department of Computer Science, CSC (2016)
77. R. Fakoor, F. Ladhak, A. Nazi, M. Huber, Using deep learning to enhance cancer diagnosis and classification, in *Proceedings of the international conference on machine learning*, vol. 28 (ACM, New York, 2013)
78. E.M. Mashhour, M.F. El Enas, K.T. Wassif, A.I. Salah, Feature selection approach based on firefly algorithm and chi-square. Int. J. Elect. Comput. Eng. **8**(4), 2338 (2018)
79. S. Park, S. Min, H. Choi, S. Yoon, deepMiRGene: Deep neural network based precursor microrna prediction. arXiv preprint arXiv:1605.00017 (2016)
80. S. Cheng, M. Guo, C. Wang, X. Liu, Y. Liu, X. Wu, MiRTDL: a deep learning approach for miRNA target prediction. IEEE/ACM Trans. Comput. Biol. Bioinform. **13**(6), 1161–1169 (2015)
81. O. Barca-Mayo, Q.R. Lu, Fine-tuning oligodendrocyte development by microRNAs. Front. Neurosci. **6**, 13 (2012)
82. A. Sharma, R. Rani, A Systematic Review of Applications of Machine Learning in Cancer Prediction and Diagnosis. Arch. Comput. Methods Eng., 1–22 (2021). https://doi.org/10.1007/s11831-021-09556-z

Deep Learning in Healthcare

V. Pavithra and V. Jayalakshmi

1 Introduction

In late years, deep learning (DL) has been an inspiring advancement of machine learning (ML). The computational roots of deeper education are easily grounded in traditional neural network examples. Only, unlike the extra traditional neural network use, a deep learning analysis is of course a technical advantage in reciprocal terms, employing more than two hidden neurons and layers [1]. The role of different neurons ensures full coverage of nonprocessed fats, only the successful nonlinear sorting of layer by layer creates a lesser dimension of the participation area. Each lower dimension projection is adapted to a higher level of imagination. If the organization is optimally weighted, the unrefined knowledge or photos are correctly conceptualized.

This high model level provides an automatic quality system, with crafted or personalized features that are otherwise required. The foundation of a basic feature without human interference, such as health information systems, provides a heavy slew of net income. It can create features that are more complicated and impossible to simplify in subtle ways in the consequence of a technically imaginative case. Fibroids and polyps could be selected and indiscretions such as tumors in the tissue pathology could be identified. These characteristics can also be picked out in regular bioinformatics nuclear progressions that could stick to DNA or RNA threads to proteins. Many designs are honored by a few methodological approaches to deep learning.

In a diversity of sectors, such as growth, transport, and algorithm governance, mechanical learning profoundly changed machine methods. The ability to better

V. Pavithra (✉) · V. Jayalakshmi
Department of Computer Applications, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_11

represent DL systems has contributed to the development of machine-sponsored analytical systems. Subsequent advancement, technology advances such as cloudy or leading-border networks, electronic networking as well as large-scale data mining benefit from the capabilities of healthcare ML systems. In summation of this scientific knowledge, ML/DL offers highly reliable statistical results and encourages a well-prepared and focused approach. This experience can play an important part in the development of the healthcare sector and will also offer specific benefits to rural and low-profit communities, such as remote healthcare.

While DL algorithms have been inhaled, many recent surveys have contributed to an increased fear of mere machine protection and force. Likewise, numerous types of data and model poisoning attacks against DL approaches have been planned, and numerous resistances to such plans have been proposed. Nevertheless, the effectiveness of the security mechanisms is too uncertain, and various studies have found that most defense strategies are dying to prevent an assault. The accuracy is that DL systems do not experience a healthy yet significantly realistic application of dangerous medical roles such as life-predictive operation. The strict implementation of ML/DL methods in the sector is also of utmost importance in ascertaining the dependability and character of DL systems and health information.

2 Deep Convolution Neural Network (CNNs) in Healthcare

In the setting of health Information Studies, Convolution Neural Nets have maximum impact. The structural design can be explicated by the challenging drain, which has degraded upgrades or pooling levels, as an alternate front level series. This physically driven structural plan is standardized to the method by which image cortexes incorporate image information in the variety of accessible areas [2]. The growing layer within the system initiates a high-level theoretical functionality. Certain reasons for deep learning include those caught up in the structure of Boltzmann Restricted Machinery, such as Deep Belief Networks, expand neural networks to a variety of points in that Deep Neural Networks are separated in Fig. 1.

The modern graphics processing unit innovations have had a significant impact on quick adoption and the deployment of deep learning. Some theoretical beliefs were predicted in the pre-GPU period following in-depth learning but in development in recent years to improve differentiation. The most basic arithmetical method can be transferred with complicated matrices such as matrix products, and the issue can be highly paralleled by profound learning formats such as CNN. An abundance of recent projects has therefore utilized deep learning technologies for medical information resources and have achieved comparable presentations or more than alternate methods in several instances.

However, there are massive problems to do with the use of deep learning throughout health information services. In addition to deep learning, large computer resources are needed without this research. The best essay on the network's free aspect will become harder to explain. Via integration problems and adaptation, a

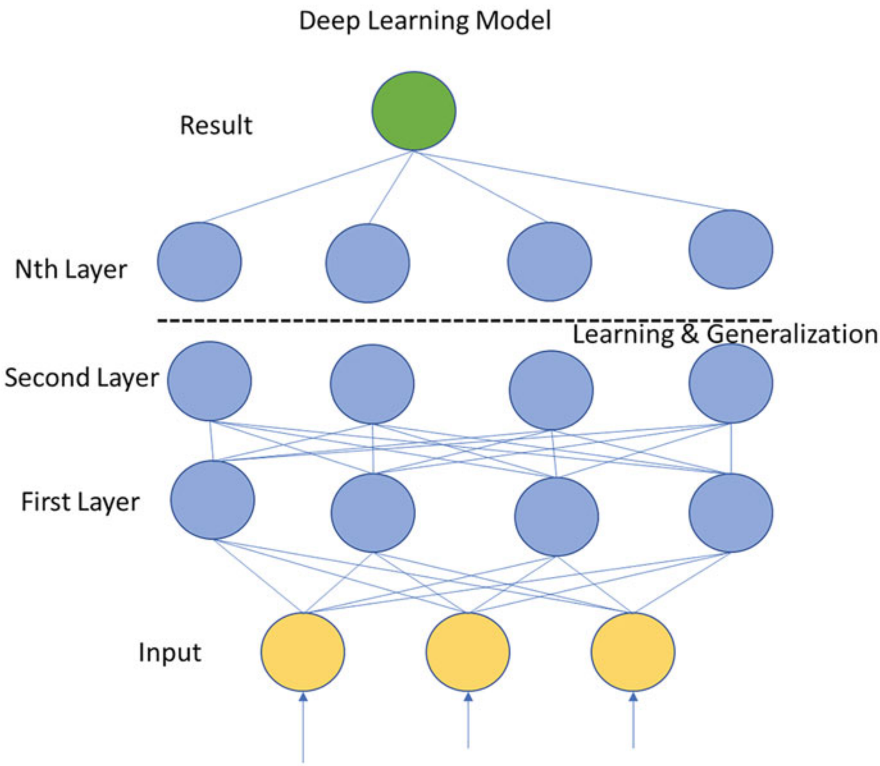


Fig. 1 Deep neural networks

deep program can also be tackled to implement additional learning strategies to resolve such problems. In the next section of this chapter, deep know-how is used in recent health information technology research to identify comparative strengths and possible matches.

3 Deep Learning for Genomics

At the feeling of every living being is their gene. The particles of DNA holding all the instructions to make the living being are their functioning parts. If a cell is a processor, then its gene cycle is the software it executes. And if DNA be software, reports are meant to be practiced by a processor, personal computers to investigate that reports [3]. But DNA is not a theoretical storage medium. It is a physical particle that acts in difficult ways. It also relates to thousands of other molecules, all of which participate in significant roles in maintaining, copying, directing, and carrying out the functions included in the DNA. The gene is an immense and difficult machine made up of thousands of parts. This takes us to the identical areas of genetics and genomics. Genetics treats DNA as theoretical information. It looks at samples of inheritance or searches for connections across people, to learn the relations among DNA cycles and physical features.

First analysis the basic image of genomics that is regularly educated in initial classes. Then, it will explain several ways in which the actual world is more difficult. DNA is a polymer, an extensive series of repeating elements strung simultaneously. In the folder of DNA, four elements can become visible through adenine, cytosine, guanine, and thymine. Nearly all the facts about living beings are ultimately determined in the precise model of these four repeating units that make up its genome.

The structure of a DNA particle consists of two sequences, each made of many adenine, cytosine, guanine, and thymine supports. The two chains are balancing; if DNA is the software, proteins are the significant hardware. Proteins are small devices that do approximately all the work in a cell. Proteins are also polymers, made up of repeating elements that are described as amino acids. Here are 20 major amino acids, and their physical properties differ extensively. Several prosperities are huge while others are minute. When the exact set of amino acids is together in the right order, it will spontaneously fold up into a 3D shape, all the parts placed right to let it function as a device.

The main purpose of DNA is to record the cycles of amino acids for an individual's proteins. It achieves this in a simple, easy method. Stretches of DNA directly correspond to proteins. All series of three DNA bases corresponds to one amino acid. From DNA to protein absorbs a different particle, RNA that serves as an intermediate symbol to hold information from one part of the cell to another. RNA is another polymer and is chemically extremely related to DNA. It also has four bases that can be sequenced mutually in random orders. To generate a protein, the information must be copied twice.

First, the DNA series is recorded into a corresponding RNA series, and then the RNA particle is converted into a protein particle. The RNA particle that holds the information is called a messenger RNA. An individual cell has thousands of dissimilar proteins to generate. There must be some sort of regulatory device to manage which proteins get made and when. In the conservative picture, this is completed by proteins called transcription factors. Each transcription factor identifies and is attached to a DNA series. Depending on the transcription factors and the place where it is attached, it can raise or reduce the speed at which nearby genes are transcribed. This provides an effortless, simple to realize the picture of it works. The occupation of DNA is to instruct proteins, extending of DNA systems for proteins using an easy, distinct way. DNA is transformed into RNA, which serves as an information shipper. The RNA is then transformed into proteins, which do the actual job. The method is extremely neat; and for several years, this picture was supposed to be acceptable. So, take a second to enjoy it before it spoils the vision by enlightening that reality is extremely messier and complicated.

4 Deep Learning in Medicine

The capability to extract significant facts from visual datasets can be useful for investigating microscopy images [4]. This capacity for managing visual data is similarly useful for clinical applications. All recent medicine needs doctors to seriously study clinical scans. Deep learning devices could probably make this investigation easier and faster and shown in Fig. 2.

Electronic Health Record Data

Traditionally, hospitals preserve paper charts for their patients. These charts would documentation the tests, medications, and other treatments of the patient, allowing doctors to follow the patient's physical condition with an immediate look at the chart. Unfortunately, paper health documents had huge trouble related to them. Transferring data among hospitals need a major amount of effort, and it was not simple to file or hunt proper health record data. There is a main push over the few years in various countries to shift from paper documentation to electronic health records. The adoption of the electronic health records systems has stimulated an explosion in research on-device learning systems that work with electronic health records data. These systems plan to use huge datasets of patient records to train systems that are capable of estimating data of patient results. In numerous ways, these are the logical successors of the professional methods and Bayesian networks

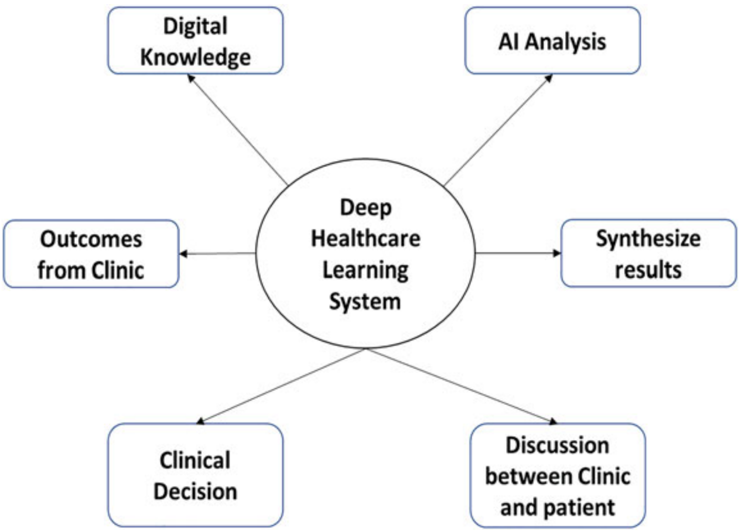


Fig. 2 Deep healthcare learning system

we just learned about. These previous systems of electronic health record systems seek to assist the method of analysis. Earlier systems wanted to assist doctors in making real-time analysis, these newer methods satisfied mostly with running on the backend.

A quantity of projects has tried to learn robust models from electronic health records data. While there are various remarkable successes, learning on electronic health records data remains difficult for practitioners. Because of privacy concerns, there are no several public electronic health records datasets accessible. Thus, a small set of selected researchers can plan these systems thus far. Besides, electronic health records data are extremely disorganized. Since human doctors and nurses manually enter data, most electronic health records data suffer from misplaced areas and all sorts of dissimilar meetings. Creating robust systems that deal with the missing data is difficult.

.

ICD-10 Codes

ICD-10 is a set of codes for patient illness and indications [5]. These codes establish broad adoption in current years because they permit insurers and governmental agencies to set regular practices, treatments, and handling charges for illness. The ICD-10 codes quantize the high-dimensional constant space of individual illness. By standardizing, they permit doctors to evaluate and group patients. Such codes will expect to prove relevant to developers of the electronic health records systems and models. This situation of issues is creating to change. Improved devices for preprocessing and for learning are in progress to facilitate efficient learning to occur on the electronic health records systems. The Deep Patient systems train and allot autoencoder on patient clinical records to generate a patient account which then uses to forecast patient results. A patient's record is changed from a set of unordered textual data into a direction. This approach of transforming different information types into vectors has widely successful throughout deep learning and seems poised to present meaningful developments in the electronic health records systems as well. Several systems created arranged the microelectronic well-being records systems have developed in the text, many are starting to include the newest tools of deep learning. While methods with these newest bells and whistles are still maturing, they are incredibly stimulating and supply pointers to where the area is likely to begin in a few years.

Probabilistic Diagnoses with Bayesian Networks

The main difficulty with expert system tools was they can only deliver deterministic forecasting. These deterministic predictions will not leave much room for insecurity.

It seemed that if expert systems can be customized to report for insecurity, this would allow them to attain victory.

This fundamental insight activates a host of jobs on Bayesian networks for medical analyses. These methods are experienced from similar limitations as the expert systems. It is still essential to seek structural information from doctors, and designers of Bayesian medical networks faced the extra challenge of seeking meaningful possibilities from doctors. This procedure added a slide to the development of adoption.

Various types of the Bayesian network need different algorithms but deep learning algorithms which use gradient descent methods work on approximately all network. Robustness of learning is regularly enabling widespread adoption [6].

5 Machine Learning for Microscopy

It illustrates a deep learning method for microscopy. It requests to recognize the organic structure of a microscopic image. Microscopy is the primary tool for the life sciences, and progress in microscopy has significantly advanced human science. Seeing is judging even for skeptical scientists, and able to visually examine biological units such as cells makes a spontaneous understanding of the fundamental mechanisms of life. Methods hope that they will allow robotic microscopy channels to become considerably more flexible. Deep learning techniques confirm guarantee at being capable to achieve several tasks a human image doctor can. Also, early investigation suggests that deep learning methods could significantly develop the ability of cheap microscopy hardware, potentially allowing economical microscopes to achieve analyses currently promising only using extremely difficult and costly apparatuses.

It is even achievable to train deep models that simulate experimental analysis. Such techniques are skilled in predicting the results of experiments without running the trial in question. This is an incredibly powerful ability, and which has stimulated the potential for deep networks in image-based natural science. Microscopy is the science of using material systems to view tiny objects. Usually, microscopes were simple optical tools, using finely ground lenses to develop the resolution of models. The field of microscopy is in progress to lean heavily on knowledge such as electron beams or physical probes to create high-resolution models.

Microscopy is attached intimately to the life sciences for several years. The Anton van Leeuwenhoek used early visual microscopes. The invention of high-resolution visual microscopes generated a revolution in microbiology. The increase of microscopy methods and the capacity to analyze cells, bacteria, and other microorganisms at scale allow the entire area of microbiology and the pathogenic representation of disease. It is tough to overstate the consequence of microscopy on the recent life sciences.

Optical microscopes are either easy or complex. Easy microscopes use a single lens for magnification. Complex microscopes use several lenses to attain higher resolution, but it cost additional difficulty in construction. The first practical compound microscopes were not accomplished until the nineteenth century. The main change in optical microscopy method design did not happen until the 1980s, with the arrival of digital microscopes, which permit the images captured by a microscope to be written to computer storage. As declared in the earlier segments, computerized microscopy uses digital microscopes to arrest huge volumes of images. These are used to perform large-scale organic testing that captures the causes of experimental perturbations. Functional super-resolution microscopy makes use of the physical properties of light-emitting materials embedded in the model being imaged. The case of fluorescent tags in biological microscopy can emphasize biological particles [7]. These methods permit typical optical microscopes to identify light emitters. Functional super-resolution methods can be generally split into deterministic and stochastic methods.

Deterministic Super-Resolution Microscopy

Several light-emitting matters have a nonlinear response to excitation. The plan is that arbitrary focus on a light emitter can be accomplished by turning off the extra emitters nearby. The physics behind this is a little difficult, but well-developed methods.

Stochastic Super-Resolution Microscopy

Light-emitting particles in biological systems are subject to random motion. This signifies that if the motion of a light-emitting particle is followed over time, its measurements can be averaged to profit a low mistake estimate of its exact position. Various methods filter this fundamental idea. These super-resolution methods had a remarkable effect in recent biology and chemistry because they permit relatively inexpensive optical equipment to investigate the performance of nanoscale systems.

6 Natural Language Processing in Healthcare

It concentrates on the analysis of text and phrases to carry on the interpretation of language. In this field, profound learning algorithms that are successful in producing sequential inputs, such as vocabulary, speech, and time sequence, play an important role. One lakh generates 50 thousand pieces of knowledge classically from a hospitalization alone. The possible benefits of this knowledge are significant. The electronic health information structures of this size reflect two lakh years of doctors' expertise and 100 million years of patient outcomes reports that include extraordinary illnesses and diseases. In this step, we have successfully translated the computer-generated text, and subtitled the picture [8] of healthcare systems such as the Electronic Health Reports, sequentially machine learning, and language innovations. A major hospital organization's electronic health reports will collect over ten million patients over a decade's medical service.

In the field of electronic health records, deep learning technologies are rapidly emerging. Figure 2 offers information on research steps for the establishment of deep learning frameworks for electronic recording. The initial purpose of aggregating unprocessed data is to ensure that a general context is established between the organizations. The curriculum is consolidated and tested briefly and by patients to make it suited to a specific training plan, which includes laboratory results, vitality, and community knowledge when predicting the date for supervised learning from incomplete planned data sets. The auditor tends to utilize weak learning strategies to take account of the unforeseen and predicted awareness of the electronic health records. There are auto-encoders in which networks can first learn significance by reducing and then recreate information that is not defined to evaluate certain situations. The modern usage of profound education copies the series of structured incidents to anticipate potential clinical experiences during the test of the patient with convolution and normal neural networks. This emphasis is on the MIM dataset for acute care, which involves patients in intensive care from a center. Although ICU patients are more enthusiastic than patients without ICU than non-ICU, non-ICU patients are greatly outnumbered. How effectively approaches received from this knowledge would simplify for larger communities is unknown.

7 Prediction Using EHR

To define patient schedules, the technical development of mechanical voice detection and data extraction systems would extend medical voices. Electronic wellness reviews recording burnout and time with providers are ingested practically multiple times in a single day. Additional scripting facilities are developed and assisted by computerized transcription. Consider RNN-assisted language translation, utilizing a process that moves directly from speech to text in another language.

This enables a patient source to be explicitly interpreted into a text document. The major concern is the arrangement and description of the dialogue in detail of the attributes and position of any therapy entity. While these methods are specialized in playing with the early design of human technologies, they must be generally scientifically applied. Future research may focus on enhancing algorithms to extend the plentiful and formless use of data from electronic health records. Medical records are often skipped or written down while expanding computational programmers. Here, large-scale RNNs begin to demonstrate promising forecast results through a semi-controlled combination of planned and formless details. This system helps them to analyze unique sources of data from broader populations and outstrip different methods utilizing practices such as culture, reading task, lifetime, and analysis [9].

8 Deep Learning Support in Healthcare

Health information collection as a formless resource is one of the functions of deep learning in the collection of health knowledge. The required documentation, including electronic health records, typically integrates a wide quantity of materials to include a thorough explanation of the patient's diagnosis, pathology, care, evaluation, and performance. In the cytological notes of the cancer diagnosis, clinical imaging will provide convincing information regarding the process and its dissemination. This information helps a detailed view of the illness or disease of a patient to be developed and an excellent conclusion to be drawn. Powerful inference by in-depth analysis and artificial intelligence can improve the coherence of medical assessment support systems. There are so many academic topics to be addressed. It is challenging to access medical and diagnosis information because healthy patients form a large portion of a standard data collection for healthcare.

In systems that have equal sets of data into which Artificial Intelligence is implemented to achieve equity, deep learning algorithms are typically used. The following description must be understood concerning the authenticity of biological fictional findings. The usefulness of the NNs must, however, be returned in this view. Another difficulty is that in-depth learning largely depends on a broad variety of results. This role allows for traditional data accessibility and privacy problems, which are essential to machine learning. In future research [10] the improvements made in the manufacture of faulty and effective functional measurement and calculating devices play an important role.

The machinability problems predicted in the coming years are reported and made informal hardware solutions for neural networks and deep learning commercially usable. Data acquisition, bulk storage, and distribution facilities tend to enhance the profound expertise of big IT firms that understand vast quantities of local and core industry's rights. It is recommended to thoroughly learn about multiple data mining and the issue of model recognition connected with health information technology, in the expectation that free bundles are common for this study.

9 Applications of ML in Healthcare

Healthcare test services trigger a tremendous volume of various details and knowledge per day, which complicates the investigation and application of traditional techniques. The techniques of ML/DL enable this statistic to be analyzed easily for valuable insights. There are mixed data streams, such as genomics, biomedical data, social network knowledge, and environmental data, that may extend healthcare data. Prediction, analytics, regeneration, and therapeutic workflow [11–14] are the four key roles of healthcare that may benefit from the ML/DL processes.

1. *Applications of ML in Prognosis:*

That is the manner that the disorder in clinical practice has been predicted to evolve. This requires a knowledge of illness symptoms and signals, as well as if they are deteriorating, healing, or stabilizing over time and identifying associated residual health conditions, conditions, and daily jobs.

Multimodal patient knowledge, including findings from phenotypic, genetic, test pathology, and clinical photos, is gathered in the diagnostic sense, which allows the ML structures to help to forecast, assessing and treat the diseases. Thanks to the latest translation work aimed at enabling customized medicine, the possible roles of ML for diagnostic diseases are the forecast of disease signs, risks, recovery, and recurrence. However, it is exciting in the field of personalized medicine that substantial advances in neighboring fields such as bioinformatics, solid validation plans, and seemingly sound ML systems implementations would have a massive translational effect.

2. *Applications of ML in Diagnosis:*

   (a) *Electronic Health Records:*

Hospitals and other healthcare facilities produce vast quantities of electronic health reports regularly and provide expected and informative documentation and a full history of the patient's treatment. For the extraction of medical characteristics, ML techniques have been used to promote the research process.

Taxonomy is the primary objective of medical picture surveys to enable physicians and radiologists' analyses and forecast diseases efficiently. The main activities for the clinical image exam include identification, description, segmentation, rehabilitation, restoration, and documentation of images. The next-generation healthcare services are projected to consist of completely digital intelligent clinical image recognition systems [12].

- *Improvement:* Deteriorated diagnostic photos are a crucial preprocessing phase that directly influences the analytical technique. The clinical picture production approach has a range of sound sources and challenges that compromise images' excellence and significance. Several gestures will decide the artifact in the picture collected. The output picture is often used to add many forms of mechanical noise. The numerous DL systems are used for the demining of health photos, such as convolution demining, autoencoders, and GANs. For cleaning up of operation, objects contained in MRI pictures were used effectively. Excellent resolution is another essential and efficient technique for the creation of medical photographs.
- *Analysis:* Traces of illness or defects in clinical photographs are defined in the analysis method. The abnormalities in the normal medical procedure are reported by expert radiologists and physicians that typically take time and effort. The significance of this study is contained in DL research and a few literature findings on disease exposure. CNN approaches the critical region of histopathologic

photographs of nuclei colon cancer for diagnosis and description. A composite approach of photographs of breast cancer used to diagnose stages of mitosis or division of cells.

- *Classification:* DL model has been reported to have high efficiency in the medical imagery when tested using state-of-the-art nonlearning structures in the challenging and complicated neural networks; the literature has typically measured modality description, dissimilar organs identification, and anomalies in an anatomical illustration by means in CNNs. A multi-instance identification strategy with CNN for different corporeal bodies is done and a CNN-based categorization methodology for interstitial lung illnesses is presented. CNN is specialized in the categorization of pulmonary nodules in an additional analysis. For medical picture arrangements, transfer learning methodology was also used.
- *Segmented:* The tissue and organ segmentation of diagnostic photos enable a detailed examination of clinical conditions of anomalies, the extent and type of cancer in the brain pictures being accidentally calculated. Furthermore, the elimination of these scientifically relevant features is a significant and significant phase in computer-aided diagnostic systems addressed later in this chapter. The segmentation approach includes splitting the picture utilizing the predefined criteria, such as important color, texture, and contrast, which are associated with multiple overlapping sections. The literature and regular structural structures used for medicinal picture segmentation are generally measured to solve the complexity of segmentation using various DL models. For multimodal images such as the brain, skin cancer, and CT pictures as suit as volumetric image segmentation, different DL architectures are calculated.
- *Reconstruction:* The way unprocessed data from the imaging sensor are produced is defined as the modernization of a medical picture. The key issue in the reconstruction of medical images is speeding the necessarily intentional data collection process, which promotes the wrong opposition in agreeing on the input of the device provided its performance. Many essential techniques in medical imaging require a lot of time to create a picture from the unprocessed data examination. Thus, in the preservation of medical photos, the goal is to minimize picture duration. The study on medical picture renewal of deep model models is increasing dramatically, with the restoration of MRIs and CT images utilizing various model DLs such as CNN and autoencoders. For example, a GAN-dependent RIM system may be obtained and the motion objects remain uncontaminated.
- *Picture registration:* Image registration is the creation of high-opinion photos for mapping and the first step in image synthesis. Image registration has several uses outlined in the medical picture breakdown. However, the usage is quite ineffective in actual therapeutic implementations. The picture checkup is also used during spinal procedures or neurosurgery to find the spinal cord or a tumor, appropriately, to ease the implant or the removal of the tumor. Multiple matches and reference points have been configured to compare the felt picture with the reference image. An organization, which uses the big deformation metric mapping representation as a patch-wise forecasting law for deformable image

logging is called Quicksilver. An unverified picture approach is also proposed for the picture operation. A 2D/3D picture listing CNN-based solution is possible that tackles two basic limitations of a current image capturing techniques, which are amplitude-dependent, that is, the limited variety of captures and the sluggish estimate.

- *Retrieval:* From the large-scale picture and video processing to big data, the newest period has seen the revolution in digital participation. This creation is also suitable for medical imagery. Thousands of medical photographs are generated every day in various ways with each hospital and clinic having radiological facilities and this extends multimodal collections with wide-scale medical images. This ensures that managing and querying such large datasets is challenging. For multimodal medical info, this is a new task. Traditional approaches are not appropriate to assist in the development and control of multimodal medical data and different ML/DL strategies are expected. As expected, doctors also balance the latest cases with the prior ones, primarily to prepare the analyses and therapies of the patient.

3. *Applications of ML in Treatment:*

   (a) *Image interpretation:* Contaminants are widely seen in clinical practice and these images are processed and interpreted by professional doctors and radiologists. The textual knowledge of all the organs studied in this analysis is analyzed and numbered to clarify photo findings. Nevertheless, in some instances, it is very difficult to draught those papers since fewer specialist radiologists and local physicians do not recognize the characteristics of medical facilities. The high-quality reporting method may often be monotonous and time-consuming with the additional signal, with qualified surgeons and oncologists, and many more consumers attending almost every day might step up. Therefore, through using computer vision and ML techniques some scientists have tried to solve this issue.

   (b) *Successful clinical surveillance:* This is a crucial observation of critical patients and the main recovery management feature. Continuous wearable health-tracking, IoT-sensors, and smartphones draw interest from one person to the next. When health surveillance is usually constant, health data are composed of a wearable computer and a mobile and then sent into the cloud for review utilizing an ML/DL device. The findings are then referred to as the required action unit. An equivalent device architecture concept is provided, for example. To track the heart rate using PPG signals, the scheme has been installed into the smartphone and cloud. Similarly, an overview of wireless wearable interface wearable ML strategies for the identification of human behaviors is given for unavailable tracking. For further study, the exchange of health data with clouds faces various isolation problems and security that is addressed in the next section.

10 Applications of ML in Clinical Workflows

(a) *Condition detection and diagnosis:* Early forecasts and diagnostic knowledge illness research are one of ML's relaxing applications. Different experiments have demonstrated how predictive medicine can be utilized to manage illnesses on time. For example, the case of cardiovascular risk estimates utilizing various clinical data ML algorithms has been studied and the studies have been found to have increased prediction performance by ML techniques. The ability to use ML-based cancer diagnostic and prediction approaches are emphasized.

(b) *ML in computer-assisted diagnosis or identification:* Computer-aided identification or computer-aided applications are usually specifically created for medical concepts to be routinely understood to support radiologists' practice throughout their clinical practice. The classification operates across various mechanisms, like ML/DL, normal computer vision, and image processing techniques, and relies highly on these techniques. CADx framework is built through the incorporation of multiple approaches such as ML. However, the computerized usage of ML/DL models may be assumed to be CADe or CADx structures in both medical figures and signal tests.

(c) *Clinical strengthening learning:* In strengthening a learning, the primary goal is to define the strategic role in drawing exact judgments to optimize accrued compensation in an unpredictable setting. RL for analyses and evaluation of patients with human differences may be included in clinical medicine. Sepsis is a serious organ failure contamination and a leading cause of death owing to lavish and under-optimum care. Easy and tabular Q-learning will classify appropriate policies for the management of sepsis and their output is close to a comprehensive state-space technique.

(d) *ML for a clinical time series data:* The medical time sequence information is one of the activities in the clinical workflow. The aim of clinical time-series modeling is to measure an intensive care unit with CNN and LSTM clinical relation, to estimate mortality in patients with troubling damage to the brain and to infer average blood pressure in patients with reasonable blood pressure, and to render significant indicators of cerebra vascular auto-regulation. A current learning understanding system is used by integrating clinical descriptions with multivariate and instant-series data for the organization of ICU's forecasting activities. Specific research explores the problem of unforeseen breathing dissociation utilizing ML techniques.

(e) *Therapeutic natural language processing:* The system used to get in contact with medical problems was widely used by doctors. The usage of therapeutic contents is vital and even the most relevant material is contained. In potential clinic software for theft of relevant knowledge from unstructured clinical findings in the area of enhancements to clinical results and study, the development of clinical NLP techniques is envisaged. Specific problems involve the usage of acronyms, vocabulary disparity, incomplete systems, and qualitative uncertainty of the therapeutic NLP. Discussions and clinical NLP options with languages other than English are possible together with the review of the clinical NLP techniques. For purposes of various standing of the NLP methods for clinical document analysis, the instigator suggested a toolkit called CLAMP.

(f) *Professional speech and audio processing:* Doctors must schedule professional descriptions and release summaries as well as details on radiology in a hospital environment. According to physicians, 50% of their case is reported in clinical situations and is highly oblivious to clinical practice, routine activities, and a short free time. In comparison to cooperating individually with patients, they typically take sufficient resources to organize health reports. Two identifiable problems for therapeutic speech recognition are disparity and utterance segmentation.

11 Secure, Private, and Robust ML for Healthcare Solutions

In this segment, it summarizes numerous projected approaches to ensure that healthcare applications enjoy healthy, confidential, and powerful ML [13].

Data Protection-ML: ML

The protection of the user's isolation in healthcare is supreme since it is a user-centric inquiry and the implementation of personal details. Preserving secrecy ensures that no explicit details regarding the topics from which data are composed should be made available in the planning and analysis for the ML model. ML/DL requires qualified data housed in a middle repository that may provide private user data that raise different stresses and utilize anonymization processes to cope with certain issues. However, significant details can be indirect regarding personal data even though the data are anonymized. It has been mentioned in the literature. The privacy concerns in the usage of ML have been solved through various literary attempts. Three separate protocols for the two-server model are presented where the data owners share private data between two nonconflicting data servers and those servers use protected two-party tests to train the ML models on the shared data.

1. *Cryptographic approaches:* A cryptographic push to the conditions in which the ML image requires various parties' encrypted estimates. The commonly employed methodology involves homomorphic coding, secret divisions, twisted circuits, and next-generation processors.

   (a) *Homorphism encryption:* It allows protected data to be obtained utilizing techniques including adding and multiplying that can be used as the basis for compound tasks to be defined. The database is classically protected using the innovative data holders' ciphertext and cryptographic keys.

   (b) *Garbled circuits:* The garbled circuits are old because two collectors request reports dependent on their proprietary data to be measured. He sends the sense along with the feedback in the shape of the chopped circuit. After the twisted definition of Alice's input was accomplished with insensitiveness, the twisted input with the twinkled circuit was then divided, if appropriate, into the impact of the feature needed. The usage of homomorphic encryption and rubbing circuits is available, to maintain ML replication and new samples submitted for the conclusion to increase three organizational performers, explicitly resolution trees, and hyperplane evaluation.

(c) *Hidden share:* Hidden exchange is regarded as a confidential exchange when it includes allocating information to many people. The underground will only be reconstructed if all appointed species are shared; otherwise, it is rare. In certain places, the secret has been revived using t tokens, which are not consistent with all tokens. A framework for private exchange is provided to compute parallel key component analysis (PCA). In a simultaneous analysis, the tag is created using the Hidden Sharing Approach to aggregate model updates provided by different input parties. There is a system for privacy-preserving sensation detection.

(d) *Protected processors:* Safe processors have formerly been built with rogue applications to guarantee the reactive code from illegitimate access is secret and honest at a higher level.

2. *Differential secrecy:* Differential confidentiality applies to the approach of absolute interference of hidden details in the intelligence case. Differential secrecy was a well-built feature of the company's lone lability for researching algorithms in aggregated datasets, and it is obvious that neighboring figures are a valid theory. Differential privacy is particularly useful for applications such as healthcare because of its various features, such as community privacy, computerization, and secondary data tightening. Grouping of data protection included related trials requires the elegant tragedy of insulation. The formula makes the algorithm schedule modularity if the systems are labeled differently. Robustness to endorse details would not artificially delete the system utilizing any side information understood in opposition [14]. The seclusion of the system is an alternative. Researchers may also locate coded and infiltrate data sets for building ML to permit healthcare uses and avoid separation breaches.

12 Conclusion

Deep learning has increased the fundamental situation of digital culture and pattern recognition. In this chapter, how the development of data-based applications in health informatics has been enabled by digital technologies that eliminate human involvement. This is suitable for many health awareness concerns and consequently encouraged much-sophisticated equipment such as diagnostic testing, medical equipment, and bioinformatics. Looking at the area, it promoted an enticing machine learning pattern, while we should not accept profound learning as a silver shell for each solo test set by health-related IT. It has been pointed out that it is still uncertain if the large variety of training expertise and computational capital required to obtain a deep understanding of optimum performance is warranted, considering extremely quick learning algorithms that can produce consistent outcomes with less energy, fewer parameterization, tuning, and advanced interpretability. Thus, this chapter concludes that the optimistic revolution of NNs and networking has been achieved by integrating the latest advantages in simultaneous treatment that co-processors enable. But building new machine learning algorithms with more knowledgeable computational resources and interpretability should be confined to a sustainable computer science research area, almost exclusively deep learning.

References

1. Y.W. Chen, L.C. Jain, *Deep Learning in Healthcare* (Springer, Cham, 2020)
2. K. Muhammad, S. Khan, J. Del Ser, V.H.C. de Albuquerque, Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey, in *IEEE Transactions on Neural Networks and Learning Systems*, (IEEE, 2020), pp. 507–522
3. M.G. Seneviratne, N.H. Shah, L. Chu, Bridging the implementation gap of machine learning in healthcare. BMJ Innov **6**(2), 45–47 (2020)
4. S. Pitoglou, Machine learning in healthcare: Introduction and real-world application considerations, in *Quality Assurance in the Era of Individualized Medicine*, (IGI Global, Hershey, 2020), pp. 92–109
5. F. Ali, S. El-Sappagh, S.R. Islam, D. Kwak, A. Ali, M. Imran, K.S. Kwak, A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Inf. Fusion **63**, 208–222 (2020)
6. M. Alloghani, T. Baker, D. Al-Jumeily, A. Hussain, J. Mustafina, A.J. Aljaaf, Prospects of machine and deep learning in analysis of vital signs for the improvement of healthcare services, in *Nature-Inspired Computation in Data Mining and Machine Learning*, (Springer, Cham, 2020), pp. 113–136
7. X. Zhou, W. Liang, I. Kevin, K. Wang, H. Wang, L.T. Yang, Q. Jin, Deep learning enhanced human activity recognition for internet of healthcare things. IEEE Internet Things J. **7**, 6429–6438 (2020)
8. M. Sendak, M.C. Elish, M. Gao, J. Futoma, W. Ratliff, M. Nichols, et al., "The human body is a black box" supporting clinical decision-making with deep learning, in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, (The Association for Computing Machinery, New York, 2020, January), pp. 99–109
9. V. Jain, J. M. Chatterjee (eds.), *Machine Learning with Health Care Perspective: Machine Learning and Healthcare*, vol 13 (Springer Nature, Cham, 2020)
10. M.Z. Uddin, M.M. Hassan, A. Alsanad, C. Savaglio, A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. Inf. Fusion **55**, 105–115 (2020)
11. J.R.A. Solares, F.E.D. Raimondi, Y. Zhu, F. Rahimian, D. Canoy, J. Tran, et al., Deep learning for electronic health records: A comparative review of multiple deep neural architectures. J. Biomed. Inform. **101**, 103337 (2020)
12. H. Naz, S. Ahuja, Deep learning approach for diabetes prediction using PIMA Indian dataset. J. Diabetes Metab. Disord. **19**(1), 391–403 (2020)
13. E.S. Kumar, P.S. Jayadev, Deep learning for clinical decision support systems: A review from the panorama of smart healthcare, in *Deep Learning Techniques for Biomedical and Health Informatics*, (Springer, Cham, 2020), pp. 79–99
14. O. Costilla-Reyes, R. Vera-Rodriguez, A.S. Alharthi, S.U. Yunas, K.B. Ozanyan, Deep learning in gait analysis for security and healthcare, in *Deep Learning: Algorithms and Applications*, (Springer, Cham, 2020), pp. 299–334

Self-Organized Deep Learning: A Novel Step to Fight Against Severe Acute Respiratory Syndrome

Shubhangi Rathkanthiwar

1 Introduction

The COVID-19 pandemic poses major challenges particularly in densely populated countries like India, which tackles with existing problems such as mediocre health infrastructure, poverty, and illiteracy. These issues are significantly hampering the processes of detection, testing, isolation, and quarantine, which are quite essential for decelerating the proliferation of COVID-19. Patients diagnosed "positive" with this virus pose a foremost threat of infection spread to every person coming directly or indirectly in their proximity such as doctors, nurses, and family members, despite an appropriate use of personal protective equipment (PPE). Through the proposed project, we wish to arrange an equipment which will be not only helpful for elucidation for Corona patients, but will also be an end-to-end medico engineering solution for saving every Corona warriors, involved in providing courageous and chivalrous field services.

Overall Theme

The proposed work in this project, "COROBOT," revolves around developing a cost-effective Robot and associated kit(s) to provide a "contactless" alternative to regular services to patients diagnosed with COVID-19. The main motivation behind this project is the point-of-use provision of necessities to patients while ensuring the suppression of risks associated with person-to-person physical interaction. Our

S. Rathkanthiwar (✉)
YCCE, Nagpur, India
e-mail: svr_1967@yahoo.com

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_12

broad goal is to nurture the sophisticated environment in hospitals, quarantine centers, and homes. The proposed product would not only directly benefit the COVID-19 patients but also provide the passivation to the associated people from the virus.

The first phase of this project would focus on developing the following primary capabilities.

- Provision of automated hand sanitization for the patient and sanitization in and around the patient's bed
- Measurement of body temperature, , heartbeat rate, blood pressure, blood sugar, etc. using pulse oximeter, and interfacing this information with a monitoring system, especially designed to update the history of an individual patient
- Facilitation of video conferencing between doctor(s), patient, and family members
- Sending alert to doctor(s) in case of emergency situation through a panic switch
- Provision of meals and medicines to the patient
- Collection of medical garbage around the bed
- Regular collection of blood samples from the patient for investigation procedures

2 Background

Researchers across the world are involved in providing solutions to the problems associated with the COVID-19 pandemic. Primary solutions include breaking the chains by upholding social distance, as the disease/virus has inflicted devastation around the world. Secondly, we have to save the lives of Corona warriors, who are providing courageous support despite the challenges they have to face in tough situations. The motivation for presenting our engineering solutions supporting the healthcare system, incorporating "Telemedicine" concept through this book chapter is multifold. Our engineering research in this area would step up the medico spotlights, and help actual healthcare workers to make them provide better services to the Corona patients.

Medical science, microbiological sciences, engineering, and technology, putting hand in hand, are playing a vital protagonist in providing elucidations on an authoritative basis, through the development of inoculations, equipment, diagnostic, and ICT tools. Strategic planning should be defined to include the role of academicians, since their balanced role in research administration, social, and academics could be the best combination to deploy resources to combat this pandemic. Motivations to carry out research and bring the product to address the current pandemic would not end at generating COROBOT. Our research team is ready to face the challenges ahead. Our efforts in carrying out multidisciplinary research, our research collaborations with medical field professionals, and entrepreneurial networks set through collaborative funding, all these factors will help clout our expertise. It will craft synergy, which can lead to the generation of state-of-the-art, transformative solution for a problem like COVID-19.

For this book chapter, we have considered a scenario of densely populated country like India, where the chances of widespread of the disease are quite obvious, even after attempts are made to take care from all aspects. When we consider diversified geographical conditions of India and also the urban–rural divides, we realize the challenges and difficulties in providing healthcare services. Telemedicine is the best solution through which we can diagnose, monitor, and cure the patients, keeping social distancing, thereby providing safety of healthcare workers.

In our project, we are implementing Telemedicine by facilitating the collection of patient's data and providing them to doctors placed remotely. This conceptualization is definitely making a very positive contribution to healthcare during this challenging phase.

The existing products in India include robots devised to deliver food and medicines, patient screening and providing sanitization. Autonomous robots namely 'Zafi Medic robot' and 'KARMI-Bot' are able to disinfect their premises using ultraviolet radiation. Robot 'Mitra' can estimate patient's parametric data, prior to its actual interaction with health workers. Our COROBOT was designed looking after these features.

The novelty of our proposed project lies in the provision of additional features such as continuous monitoring, data collection, and updating logs in an easily comprehendible graphical format which the doctors can immediately refer to and make appropriate decisions. Prototypes of robots with some of the proposed features are complete, which have been designed for Tele-monitoring, Teleconsultation, and Tele-diagnosis. The models were developed using National Instrument's MyRIO setup, which required biomedical sensors and IoT platforms. Experimentation has been done in one of the quarantine centers. One of the developed prototypes is based on a low-cost solution for health monitoring of bedridden patients at home, the primary feature being an ability to send alarm signals to doctors and relatives.

Contextual Feature of the Book Chapter

Few years back, our research team signed an ethical agreement with medical institute, and the institutional ethics committee agreed to share some medical records and data. The Scientific Scrutiny Committee and Departmental Research Committee helped in finalizing research topics and preparing an outline of the research. Research themes were highly multidisciplinary projects incorporating deep knowledge of medical science and engineering. The time required for the disposal of research was considered before framing objectives and methodology, and a frame work was prepared after a long-thought and critical review of literature. Some topics were selected by officials and experts from medical science and they urged engineering solutions after facilitating the high utility of their available laboratories and medical expertise in the field of study. The era of work with the concept of "Telemedicine" began and our research team started developing models using NI LabVIEW, MyRIO, Bio-Medical Sensors in ambulatory services, and IoT platforms were developed [1, 2]. As far as intellectual property rights are concerned, some of the application processes for filing pertinent patents are in progress.

Need of Telemedicine for COVID-19

Across the world, every country has been adversely challenged by the COVID-19 pandemic. As the COVID-19 virus has inflicted devastation around the world, primarily, it has become indispensable to break the chains by upholding social distance. Secondly, we have to save the lives of corona warriors, who are providing courageous support despite the challenges they have to face in tough situations. The motivation for presenting our engineering solutions supporting the healthcare system, incorporating "Telemedicine" concept through this book chapter is multifold. Our engineering research in this area would step up the medico spotlights, help healthcare provider organizations to better respond to the needs of patients who have contracted the Corona virus.

Medical and microbiological Sciences keeping hand in hand with Engineering and Technology must play a vital protagonist role in finding solutions for the global challenges, on imperative basis, through the development of new vaccines, devices, diagnostic tools, and information & communication tools. Some strategies must be defined to include the role of academicians to help communities and nations manage and deploy resources to combat this pandemic. Global challenges call for global collaborations and partnerships, bringing together the best and brightest scientists, engineers, academicians, and entrepreneurs to work together to find solutions. Motivations to carry out research and bring the product to address the current pandemic would not end at generating COROBOT. Our research team is ready to face the challenges ahead. Our efforts in carrying out multidisciplinary research, collaborations with medical field, and entrepreneurial networks will help to leverage and share our expertise across communities. It will also create synergy, which can lead to the generation of innovative and transformative solutions.

For this book chapter, we have considered the scenario of densely populated country like India. When we consider diversified geographical conditions of India and also the urban–rural divides, we realize the challenges and difficulties in providing healthcare services. Telemedicine is the best prescription in bridging the gap; however, bringing technology for healthcare delivery to reach remote and far-flung areas was the challenge. Telemedicine is making a very positive contribution to healthcare during the pandemic and is being used in a variety of ways. But telehealth technologies do have certain limitations when it comes to treating patients during a pandemic. Further, there is a chance Telemedicine could add to hospitals being overwhelmed, unless it is used well. But hospitals are learning to adapt to telehealth during a pandemic.

When the factors including benefits, limitations, burdens, and adaptation of Telemedicine for COVID-19 were considered, and the possibility of using Telemedicine in the context of COVID-19 was discussed with doctors and other medical professionals, few points were noted as a part of suggestion, opinion, prior to actual collaborations. They opined that one of the greatest advantages of Telemedicine technology for COVID-19 patients is to provide precautionary treatment. Many COVID patients can contact doctors from their respective homes, with pre-scheduled Teleconsultations, to avoid face-to-face clinical visits. This can also minimize their risk of health workers from actually contracting the virus.

Previous Work

The aim of this text is to explore possibilities of extending our previous research in the healthcare area for carrying out the futuristic work of our project "COROBOT," the background theme of which revolves around developing a cost-effective Robot and associated test kits with auto-open and close facilities. Our basic objective is to provide a "contactless" alternative to regular services to patients diagnosed with COVID-19. In the implementation part, algorithmic part for deep learning will be reconnoitered, which can help in Tele-diagnosis. The passivation thus provided to the Corona warriors from medical field from the virus would be highly demanding, because while providing treatment to the patients diagnosed with COVID-19, the life of medical staff including doctors, nurses, and ward-boys in contact with Corona patients is not safe. Our deep research revolves around problems related to multipath propagation and trans-reception of medical data. This analysis is required to enhance reliability of the signal detection at the monitoring end. Otherwise signal transmission through wireless media mainly cause deteriorations of the medical data, which may result in wrong diagnosis. If certain features of computational intelligence were added through deep learning approach, we expect performance improvement in overall working.

Our research team has developed an IoT (Internet of Things) module for Telemonitoring [1]. Development of functional methods for connecting Telemedicine Facilitation Devices with Internet of Things was the first step of our research. This project was highly appreciated by doctors from the hospitals with which we had signed the ethical agreements. They opined that this could be used as telecare services for senior bedridden individuals at homes, crisis checking, perception of constant ailments like individuals with heart problems or diabetic patients, and telecare of pregnant ladies, especially from rural areas. The IoT model proved to be a cost-effective product for Teleconsultation and Tele-diagnosis and was performing pre-diagnosis tests effectively. Consistently, real-time patient monitoring was possible for ECG, heart rate, body temperature, weight, pulse rate, blood oxygen immersion, and transmitting the data from remote location to doctor was also possible additionally including CTG (Cardio-topography) and pulse oximeter. This model was used as a basic model, as shown in Fig. 1, in the preliminary work of the proposed COROBOT, with basic of data collection on a continuous basis.

One of our research teams has developed a model using the internet of things (IoT) approach for providing smart operations in healthcare services. The main

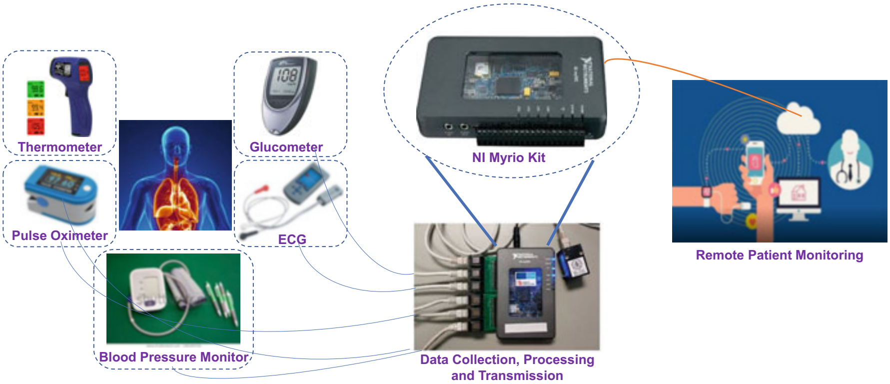


Fig. 1 Basic system for tele-monitoring

focus was on two parameters, namely vital signs early detection (VSED) and physical signs early detection (PSED) [2]. Our research team was successful in drawing important conclusions even for multiple chronic disease parameters measuring system using two prominent technologies namely, Internet of Things (IoT) and Thing-Speak. Doctor's team was satisfied with the data collection and monitoring technique, which helped them to achieve pre-diagnostic outcomes of chronic patients. Their diagnosis was enriched by identifying the physiological deterioration through early scores. Additionally, the system was involved in quantifying multiple chronic parameters like blood pressure (BP), in which blood pressure was measured in the arteries when the heart rests between beats called diastolic and pressure in the patient arteries during the tightening of patient heart muscle called systolic, and body temperature and pulse rate were also properly calibrated. Data were propelled to the cloud using Thing-Speak.

3 Working of Proposed Model

The cabinet of the COROBOT is proposed to be arranged near Doctor's table. It will perform various tasks, as defined by the expert physician. It will take round in the patient's ward, as if doctors and nurses are taking periodical rounds on daily basis. Figure 2 shows the basic structure of 'COROBOT', which is a multifunctional, line follower robot. It can detect and follow the lines drawn on the floor, and in the ward, lines/paths are drawn from Doctor's table to the patients admitted. Invisible magnetic path is preferred for the project. We have used IR reflective line/object sensor, a QRD1114 sensor and a reflective optical sensor for left/right movements, CNY70.

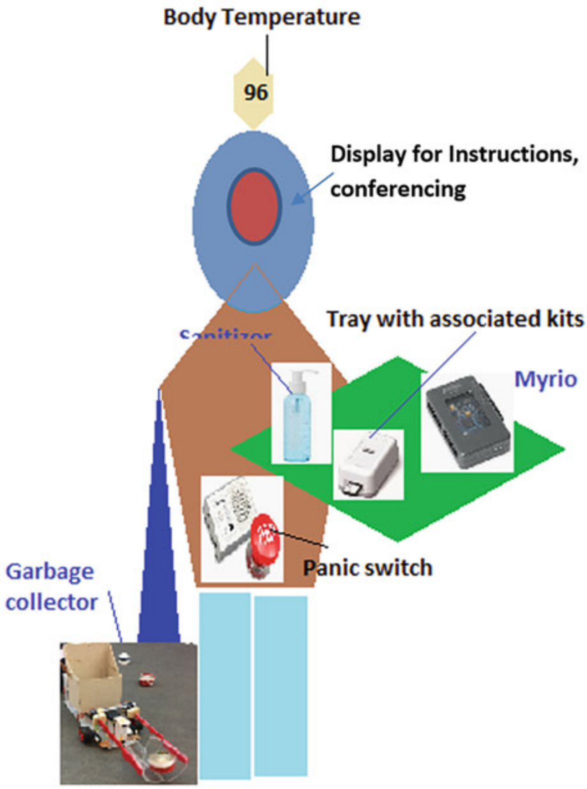


Fig. 2 Basic structure of "COROBOT"

COROBOT would focus on developing the primary proficiencies, including

1. *Collection of medical garbage around the bed and provision of automated hand sanitization for the patient and sanitization in and around the patient's bed:* When COROBOT goes near one patient's bed, it will instruct patient, to cover the eatables, food items, and his/her face. It will collect garbage and medical debris around the bed and rack. Once this task is over, it will again take a round of complete bed and will provide sanitization.
2. *Measurement of body temperature, heartbeat rate, blood pressure, blood sugar, etc. using pulse oximeter, and interfacing this information with a monitoring system especially designed to update the history of individual patient:* These data will be recorded in a particular patient's folder, which gets automatically updated. Information will be recorded continuously on cloud, so that doctors can anytime check the updates and history as well.
3. *Facilitation of video conferencing between doctors and patients:* This is a highly required and important feature of COROBOT because through this feature,

medical staff can safely communicate with the patient. Family members can also avail this facility and can have chat with the admitted patient.

4. *Sending an alert to doctor(s) in case of emergency situation through a panic switch:* A panic switch is a strategically placed electronic device designed to assist a patient in alerting doctors and nurses in emergency situations, where a threat in terms of emergency to the patient arises.
5. *Provision of meals and medicines to the patient:* Since COROBOT has been concocted with a wide tray, it can also serve medicines and meals to the patients. This can reduce the rate of contact of medical staff including ward boys and nurses.
6. *Regular collection of blood samples from the patient for investigation procedures:* This can also be a part of regular practice.
7. *Provision of test kits:* When COROBOT reaches the patient's bed, it will instruct the patient to cooperate with the process of collecting test samples. The kit will be opened automatically, the sample collected, sticker with patient's name will be wrapped around the sample, and safely transferred to a compartment. This is the most important feature of our project.

4 Deep Analysis

Motivation for this project and applying analysis based on computational intelligence as well was the unfortunate worldwide rise in the number of patients and death rate, which is the most disastrous factor. COROBOT is an equipment, which can be substantiated as an elucidation for Corona patients and an end-to-end medico engineering solution for saving every Corona warrior, including doctors, nurses, and all health workers, involved in providing courageous and magnanimous field services. Though the factor "identifying patient who is at risk" was a priority, we were interested in applying the concept "Tele-monitoring" and "Tele-diagnosis" at top priority, thereby enhancing the level of social distancing.

Most important parameter in carrying out Telemedicine-related services is the reliable detection of medical signals. This is one of the most challenging issues in present and future Telemedicine-related wireless communication systems. This research is important, since in any case, every data from the patient must reach doctors and analysts without any disturbance and deterioration. This research is important, since in any case, every data from patient must reach doctors and analysts without any disturbance and deterioration. Factors resulting deterioration of signal, or degradation of trans-reception, such as attenuation, noise, multipath effects, interference, time variation, nonlinearities, Doppler spread must be strictly avoided. Through this book chapter, an attempt has been made to focus on a deep learning approach, which is a self-organizing neural network-based intelligent system. Major elements that contribute to multipath fading effects are also considered and to account for major degradation categories, various fading channel models are considered [4, 5].

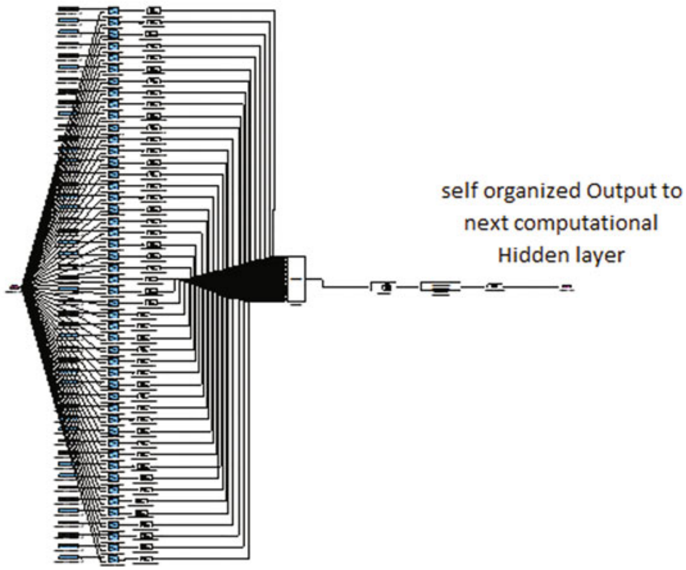


Fig. 3 Self-organizing unit

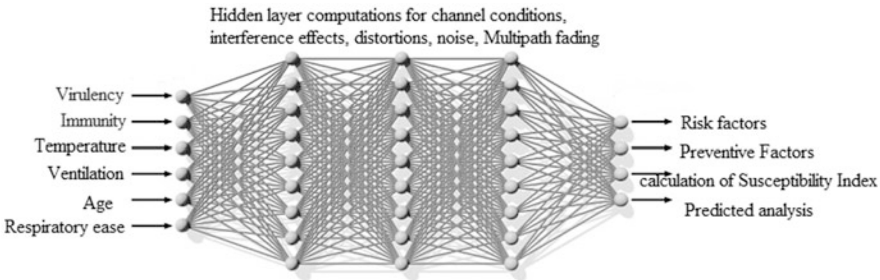


Fig. 4 Deep learning with self-organizing units

Due to the constrictions of intellectual property rights, results for the entire scheme cannot be disclosed: however, an attempt at bringing performance improvement through a self-organized approach has been done.

Self-organized deep network will classify data with pre-trained networks with parametric data including virulence, immunity, temperature, ventilation, age, and respiratory ease, process the data in computational hidden layers, provide output in terms of risk factors, preventive factors, calculate susceptibility index, and do the predictive analysis. Figure 3 shows a single self-organizing unit, and in deep analysis, many such are required. Deep neural net is shown in Fig. 4.

It can be observed that deep learning processes unstructured data, including virulency, immunity, temperature, ventilation, and ease in respiration (the data collected from COVID patient) through several layers of self-organizing feature

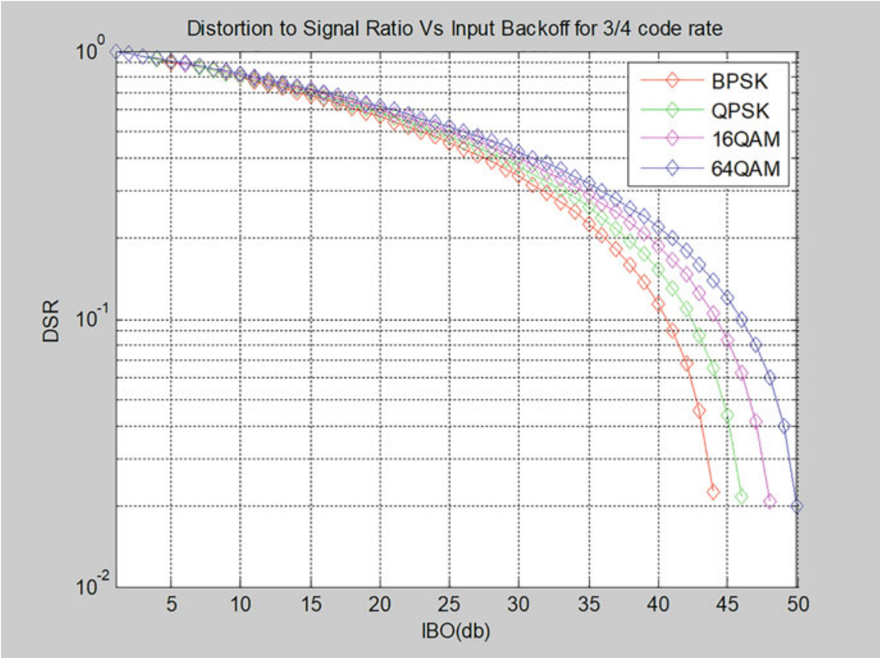


Fig. 5 DSR plot

maps, which is passed to the next layer [3]. The data become highly unstructured, when it is to be made available to doctors, through Telemedicinal OFDM-based equipment due to multipath fading, delay spread, impulse noise, frequency-selective fading channels, Doppler spread, phase noise, frequency offset, and time and frequency synchronization errors.

DSR plot shown in Fig. 5 clearly indicates the performance of our proposed deep learning model used.

In the COROBOT project, some industry partners, collaborators from India and the United States are also involved. Intellectual property will be shared among the partners, who would be actually involved in design and development of the proposed project. The relationship between the US partner and Indian partner is equal in all respects, including the filing of patent application, to complete all stages of IPR (publishing and examination) applying for PCT applications, so that it would help to get business and capture market internationally. In the process of commercialization, partners from both countries will take initiative to commercialize the proposed prototype internationally and get all the certification which is required for product selling. Both the partners will decide the manufacturing, marketing, and selling policy under one roof so that product will create standard in all respects nationally as well as internationally.

After developing prototype of the COROBOT, experimentation work has been carried out in the identified quarantine center. In the second phase of the project execution, business plan and commercialization strategy would come in picture.

5 Conclusion

1. The unique features of the proposed project, "COROBOT," designed specifically for COVID-19, are as follows:
   - Cost-effectiveness, compactness, and point of use.
   - Movement control through mobile phones.
   - User-friendliness.
   - Ability of robots to read and learn the physical path it traverses by mapping. the information in its memory. This will enable an auto-mode for the subsequent maneuvers.
   - Video Conferencing at Wi-Fi range.
   - Facility of telemedicine.
   - Screen for displaying essential instructions and messages.
   - Automated sanitization around the bed.
   - Display of patient's medical report in a graphical format.
   - Provision of "Panic Switch."
   - Automated medical garbage collection.
2. We have analyzed the trans-received multicarrier Telemedicine signal, which traveled through multipath environment. Rician and Rayleigh channel models were considered for inspecting possible LOS and NLOS paths. Nonlinear distortion resulting due to high peak to average power ratio was considered a major parameter for degrading. Self-organizing neural network was used to perform classification and function approximation.

References

1. Y. Ahir, S.V. Rathkanthiwar, Y. Kale, A review on teleconsultation and telediagnosis using IoT module, in *2019 International Conference on Communication and Signal Processing (ICCSP)*, (IEEE, Chennai, 2019), pp. 0153–0156. https://doi.org/10.1109/ICCSP.2019.8698073
2. Y.S. Kale, D.S.V. Rathkanthiwar, Smart healthcare system using low cost IoT module. Int. J. Futur. Gener. Commun. Netw. **13**(3), 9 (2020)
3. A.S. Rapatwar, S.V. Rathkanthiwar, Design of channel selection filter for wireless receiver, in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, (IEEE, 2015), pp. 809–812. https://doi.org/10.1109/ECS.2015.7125024
4. S. Rathkanthiwar, C.G. Dethe, K.D. Kulat, A neural approach for overall improvement of OFDM system, in *International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2011)*, (IET, Chennai, 2011), pp. 764–769. https://doi.org/10.1049/cp.2011.0466

5. S. Rathkanthiwar, C. Dethe, K. Kulat, Performance evaluation of a novel WLAN receiver performance in highly dispersive nonlinear environment, in *2012 2nd International Conference on Power, Control and Embedded Systems*, (Allahabad, India, 2013, April 29). https://doi.org/10.1109/ICPCES.2012.6508087

Dr. Shubhangi Rathkanthiwar has 30 years of teaching experience, and her area of specialization includes signal processing in wireless communications, wireless adhoc networks, artificial neural networks, embedded systems, IoT modules in Telemedicine, etc. She has authored more than 97 papers published in refereed international journals and conferences, and has five books, three book chapters, three patent applications, and two copyrights to her credit. She served as a reviewer for IEEE Trans. on Vehicular Technology, Industrial Electronics, IEEE Communications Survey and Tutorials, IEEE Access, and Elsevier Journal on Applied Soft Computing. She served as a TPC member and reviewer of several IEEE conferences. Dr. Rathkanthiwar is the recipient of "Best scientist Award-2017," "Shiksha Ratna Award 2018," "Best research paper award" (2011), "Best Teacher award" (2010), and "Shikshak Ratna award" (2011) conferred at international and national levels. Her research was nominated for award in the University of California, Berkley, USA. She has visited nine countries, including the United States, Singapore, England, Scotland, etc., for presenting research and for academic collaborations.

Clustering Algorithms in Healthcare

Neerja Negi and Geetika Chawla

1 Introduction

Nowadays, data mining is an essential technique to determine the most useful information from the data set [1]. It provides the use of automated data investigation techniques to exhibit the relationship of patterns within data sets. It offers various facilities in the healthcare sector. It helps in determining the early diagnosis of illness, providing therapeutic solutions to the person at a cheaper price and the identification of most suitable treatment methods. It also supports researchers in the development of drug recommendation methods and to draft healthcare policies [2]. Data mining methods are generally classified into supervised and unsupervised learning [3]. Supervised learning tries to identify unlabelled data according to the labelled training input data. It is used in predictions to determine the value of the output variables. In this technique, based on the known input and output variable, a model can be formed from a training data set [4]. After that it can predict the class label for the unknown output variable. In supervised learning, the model requires an ample amount of labelled data for the learning process. Unsupervised learning reveals to learn the hidden patterns from the unlabelled information. In this technique, there exist none of the output data for prediction as compared to supervised learning. According to the relationship within the data points, this technique discovers the patterns in the data set.

N. Negi (✉)
Faculty of Computer Applications, MRIIRS, Faridabad, Haryana, India

G. Chawla
Department of Computer Science & Engineering, BSAITM, Faridabad, Haryana, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_13

2 Clustering

Clustering is a task of partitioning the objects into groups of data points such that the data points in a cluster have more resemblance as compared to data points in other cluster [5]. Clustering algorithms are used to classify every data point into a particular group with a given set of data points. Data points of same cluster must have similar properties whereas data point of diverse cluster must possess dissimilar properties. Clustering is unsupervised learning that is used to assist professionals in finding hidden patterns in a data set. It results in exhibiting similar and dissimilar properties for the different groups. Let us understand this with an example. Suppose, as the head of a departmental store and to understand preferences of customers to enhance the business. It is impossible to look at details of each customer and plan a unique business strategy for each one of them. So, the easy way to do is to cluster all of our customers into say five groups based on their past history of shopping and use a separate strategy for customers in each of these five groups using clustering. Clustering algorithms can be used in different sectors, that is, for the classification of diseases in medical era and customer interest in market study. There is no universal method for clustering, so various methods are used for diverse clustering purposes.

K-Means Clustering

Now, that we understand what is clustering. Let us take a look at the types of clustering.

K-Means Clustering Algorithm

This is one of simple clustering algorithm since it is straightforward to implement. It is a form of unsupervised learning used for data without defined groups. This algorithm works repeatedly to allocate each data point to one of K groups based on the characteristics that are provided. K-means clustering algorithm has been found to be very helpful in grouping new data. Few applications which use k-means clustering are sensor measurements, activity monitoring in a manufacturing practice, audio detection, and image segmentation [6, 13].

Steps:

1. Selecting the amount of classes to be used and arbitrarily initializing their own centre point.
2. Categorize every data point by evaluating the distance between that point and centre of every group. Then classify the points to be placed in the group having nearest centre.
3. Recompute the group centre by taking the mean of all the vectors in the group.

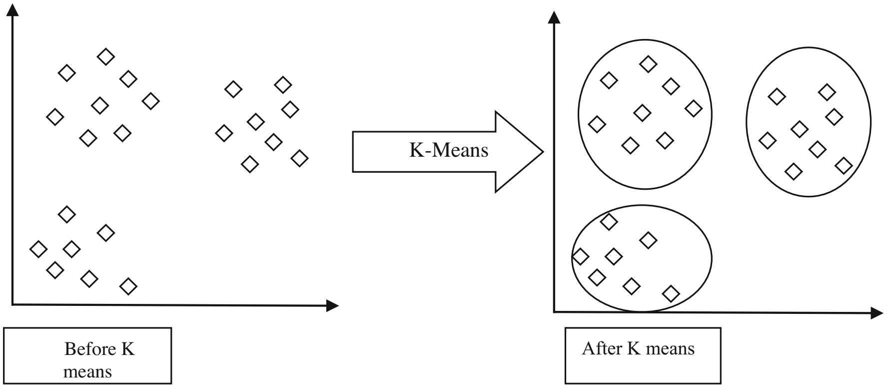


Fig. 1 K-means clustering algorithm

4. Do these steps repeatedly until the group centres do not alter to a great extent between iterations. Now, initialize the group centres randomly and then choose the one that seems to provide the best result, as shown in Fig. 1.

Pros:

- Simple to execute.
- It is a fast method due to fewer computations.
- For large number of variables, K-means may be faster than hierarchical clustering (if value of K is small).
- It may produce higher clusters than hierarchical clustering.

Cons:

- The challenging aspect is identifying and classifying groups.
- Because of arbitrarily selection of centre of cluster, the result may be inconsistent.

Mean-Shift Clustering Algorithm

Mean shift is a kind of sliding window algorithm [7]. It is useful to discover the crowded region of data points and to trace the centre points of each group. Within the sliding window, it updates the candidate for the centre points as the mean of the points which can be used to eliminate the duplicate values. So, the result is arrangement of final set of centre points with their related groups. The distinction between k-means algorithm and Mean-Shift is that in Mean Shift, there is no need to state the quantity of clusters in advance because it will be determined based on the data. Mean Shift clustering algorithm is mostly useful in Computer Vision problems, Image Processing, Video Tracking, and Image Segmentation.

Working of Mean-Shift clustering algorithm with the help of these steps:

- Step 1 – Begin with the data point that a cluster possessing itself.
- Step 2 – Now calculate the centroids.
- Step 3 – Modify the position of new centroids.
- Step 4 – Repeat the process and move to the high density area.
- Step 5 – Stop once the centroids reach at position from where it cannot shift further.

Pros:

- Unlike the k-means clustering algorithm, selecting the quantity of clusters is not necessary.
- The cluster centre should meet towards the point of maximum density.
- It is much better as compared to k-means due to the reason that there is need to give the value of 'k', that is, the number of clusters.
- This algorithm takes only one input, that is, the bandwidth of the window.

Cons:

- It may be computationally expensive due to lot of steps in this algorithm.
- The selection of the bandwidth is an important issue.
- If the bandwidth is very small, few data points may be missed, and may not reach at convergence.
- If the bandwidth is very large, some clusters might be missed entirely.

DBSCAN Clustering

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is extensively used in density-based algorithm [15, 16]. DBSCAN has been found to be useful in the discovery of non-linear shapes based on the density.

Let P = {p1, p2, p3... pn} is a set of data points. DBSCAN requires two parameters: € (ep) and the least amount of points required to make a cluster (minpt).

Step 1 – Begin with an arbitrary point not explored so far.

Step 2 – Remove a neighbour of this point using € (data point within the € distance are neighbours).

Step 3 – Clustering process starts if there are enough neighbours around that particular data point and mention them as explored, else consider that point as false point (that point may be a part of the cluster after some time).

Step 4 – If a point is present in a cluster then its € neighbour is also the member of the cluster. So, iterate from step 2 for all € neighbour points until all data points in the cluster are determined.

Step 5 – Then retrieve and process the new unvisited point that is used to identify a point of cluster or false data.

Step 6 –Iterate this process until all points are mentioned as visited (Fig. 2).

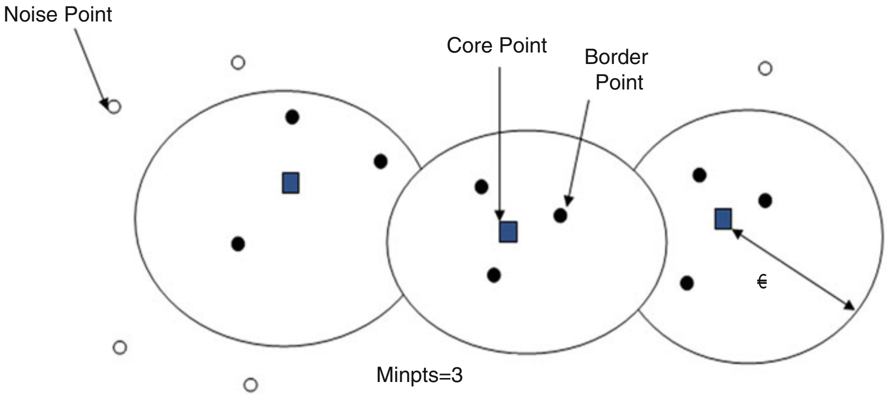


Fig. 2 DBSCAN clustering

Pros:

- It is better than other cluster algorithms because there is no need to specify number of clusters in advance.
- It recognizes outliers as false data.

Cons:

- This clustering algorithm may not be very helpful when there are varying density clusters. In case of change in the density level, there may be deviation in the location of the threshold distance ε and the least amount of points to identify the neighbours.
- It would be a challenge to determine the threshold distance ε for high dimensional data.

Expectation-Maximization (EM) Clustering Using Gaussian Mixture Models (GMM)

The Gaussian mixture model (GMM) clustering provides more flexibility as compared to other clustering techniques [6]. In GMM Clustering, initially, it is assumed that all the data points are distributed according to the Gaussian distribution. To determine the shape of each cluster Mean and standard deviation are the main parameters. The standard deviation can be in both P and Q directions, each cluster can draw ellipsoids, for multivariate models. They do not need to have a spherical shape. Expectation-Maximization (EM) is an optimization algorithm that is used [6, 14] to obtain the parameters regarding the Gaussian for every group.

Similar to the k-means clustering algorithm, firstly, it selects the number of clusters and arbitrarily initializes the Gaussian distribution parameters. After that the GMM algorithm follows the following steps.

1. In each cluster, estimate the probability for every point. If the point is closer to the Gaussian's centre, then there are better possibilities that it can fit into the group.
2. Now, according to these calculations, a new kind of parameter is located for the Gaussian distributions that increase the probabilities of data points inside clusters. The weight factor is an important parameter to determine the probability of the data point relating to the particular cluster.
3. Repeat the above-mentioned steps until convergence.

Pros:

- GMM uses the concept of standard deviation. So, it provides more flexibility in terms of cluster covariance.
- In k-means, a data point relates to one and only one cluster. On the other hand, in GMM, a data point relates to each cluster to a precise degree. The degree is based on the probability of the point being produced from each cluster's normal distribution, with cluster centre as the distribution's mean and cluster covariance as its covariance.

Hierarchical Agglomerative Clustering

Hierarchical clustering is the extensively used method to do analysis of social network data. The data are compared with one another based on their resemblance in this clustering method. The group of nodes are combined to form larger groups based on their similarity. So, the core task in hierarchical agglomerative clustering is to iteratively merge the two nearest clusters into a larger cluster [8].

Working of hierarchical clustering algorithm:

Suppose there are six data points' p**, q, r, s, t.**

Step1– Assume each alphabet as a single cluster and estimate the distance of one cluster from all the other clusters.

Step 2 – Now similar clusters are combined collectively to make a single cluster. Suppose cluster (q) and cluster (r) are similar to each other, so combine these in the second step similarly with cluster (s) and (t) and at last, the clusters [(p, q), (r), (s, t)] are obtained.

Step 3 – Recompute the closeness of points according to the algorithm and join the two nearest clusters [(r), (s,t)] together to form new clusters as [(p,q), (r,s,t)].

Step 4 – Lastly, the remaining clusters are merged together to form a single cluster [(pqrst)] as depicted in Fig. 3.

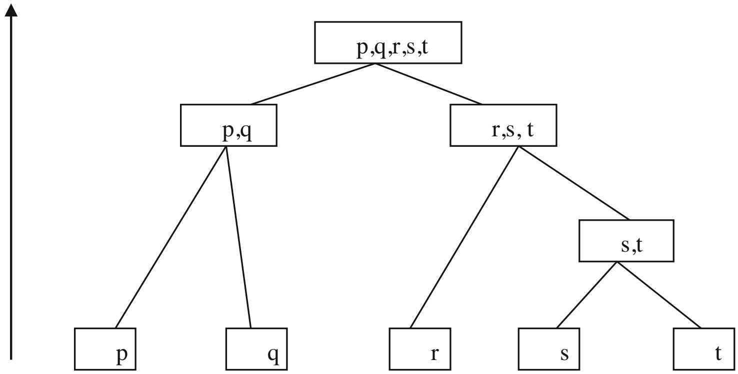


Fig. 3 Hierarchical clustering

Table 1 Description of sample data set

Pregnancies	Glucose	BP	ST	Insulin	BMI	DPF	Age
6	148	72	35	0	33.6	0.627	50
1	85	66	29	0	26.6	0.351	31
8	183	64	0	0	23.3	0.672	32
1	89	66	23	94	28.1	0.167	21
0	137	40	35	168	43.1	2.288	33

Pros:

- It is not necessary to specify number of clusters in advance.
- It is insensitive to the selection of distance metric.

Cons:

- The algorithm can never undo any previous steps. So for example, the algorithm clusters 2 points, and later on if it has been noticed that the correlation was not a good one, the program cannot undo that step.
- Due to time complexity, the computation time may be high as compared to other methods like k-means.

3 Experimental Analysis

The experiment has been conducted on windows 7 OS,4 GB RAM. The implementation has been done in Python. The data set on diabetes from the UCI Machine Learning Repository [9] has taken into consideration. It consists of eight attributes and 768 observations. It contains the records of 500 non-diabetic persons and 268 diabetic persons (Table 1).

Code 1: Load the Data Set

```
import pandas as pd
data= pd.read_csv('diabetes.csv')
data = data.drop("Outcome", axis = 1)
array = data.values[:,0:8]
print(data.head(10))
```

The diabetes data set has some missing values. Therefore, it is necessary to preprocess the data before using it. The data processing techniques enhance the overall quality of the models mined and also decreased the time needed for real mining. In the Diabetes Dataset, some of the entries are having zero values such as glucose level, body mass index, diastolic blood pressure, skin thickness, and physically impossible insulin level. Therefore, there is a need to preprocess the data.

The preprocessing has applied on the data set to normalize the values as shown in Table 2.

Code 2: Preprocess the Data

```
from sklearn.preprocessing import scale
X = pd.DataFrame(scale(data))
print(X.head(n=5))
```

After that apply the k-means clustering technique to identify the clusters as shown in Fig. 4.

Code 3: K-Means Clustering

```
clustering = KMeans(n_clusters=4)
clustering.fit(X)
```

To observe the performance of hierarchical clustering, agglomerative hierarchical clustering has also applied on the diabetes data set (Figs. 5 and 6).

Code 4: Apply Hierarchical Clustering

```
clust = AgglomerativeClustering(n_clusters = 4,
affinity = 'euclidean', linkage = 'ward')
clust.fit_predict(pat_data)
```

Table 2 Preprocessed data

Pregnancies	Glucose	BP	ST	Insulin	BMI	DPF	Age
0.639947	0.848324	0.149641	0.907270	−0.692891	0.204013	0.468492	1.425995
−0.844885	−1.123396	−0.160546	0.530902	−0.692891	−0.684422	−0.365061	−0.190672
1.233880	1.943724	−0.263941	−1.288212	−0.692891	−1.103255	0.604397	−0.105584
−0.844885	−0.998208	−0.160546	0.154533	0.123302	−0.494043	−0.920763	−1.041549
−1.141852	0.504055	−1.504687	0.907270	0.765836	1.409746	5.484909	−0.020496

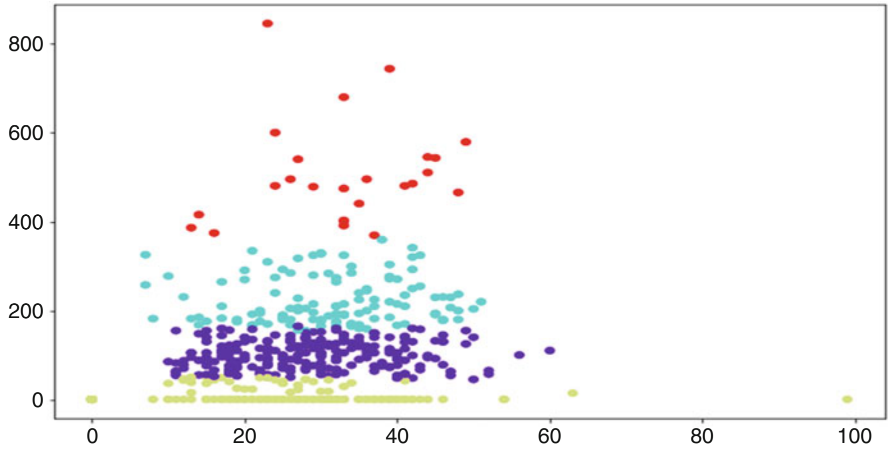


Fig. 4 Clusters according to k-means clustering

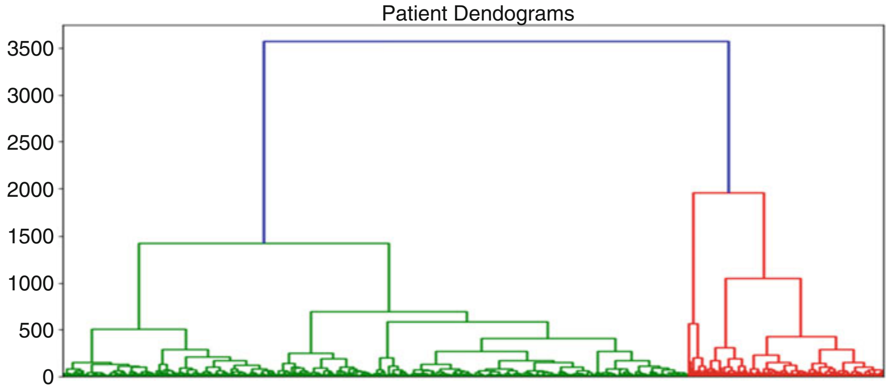


Fig. 5 Dendograms

Performance Analysis

The following performance metrics have been taken into consideration for measuring the performance of the clustering algorithm.

Davies Bouldin (DB) Index It measures the average correlation among a cluster and its most related equivalent one. It is required that the clusters should be distinct from the other cluster. Thus, a clustering that reduces the index value is the perfect one [10].

Silhouette Analysis It is employed to resolve the distance of a specific object into one cluster with respect to another object in different clusters. Its score values lie among −1 to +1. Here +1 indicates the object lies in the correct cluster and −1 indicates that objects are not accurately clustered [7, 11].

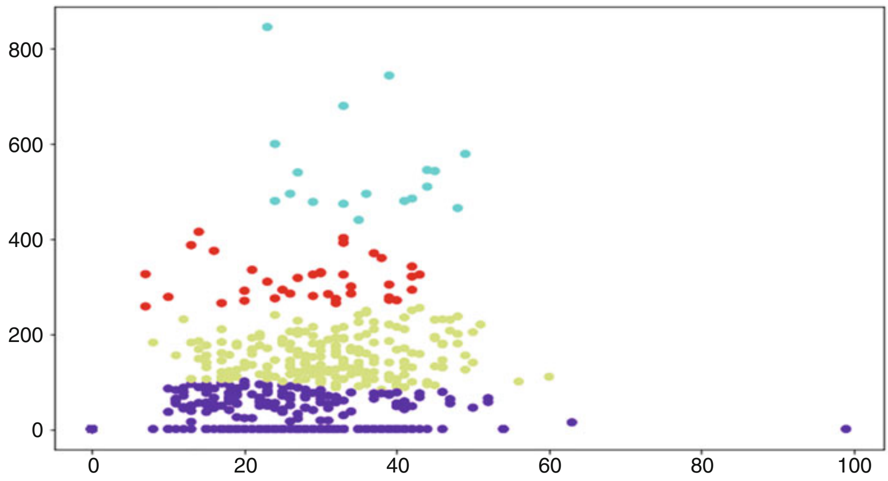


Fig. 6 Cluster according to agglomerative clustering

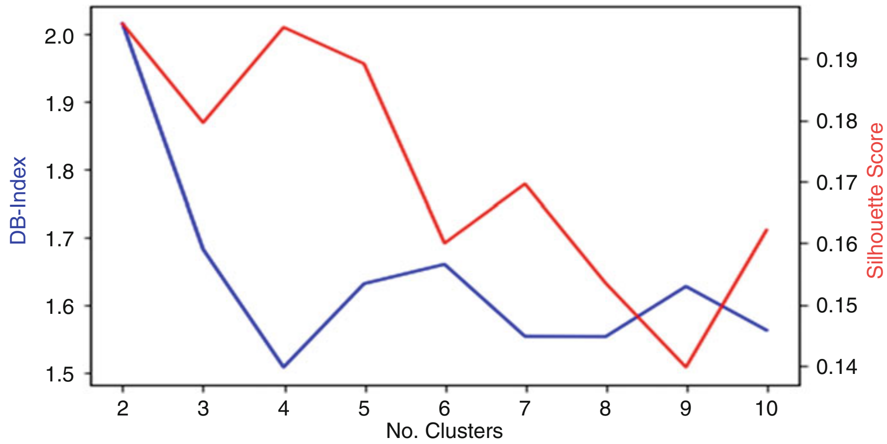


Fig. 7 DB index and Silhouette analysis

It has been depicted in Fig. 7 that according to both the indices suitable number of clusters is four.

It has been observed from Table 3 that agglomerative clustering has better execution time as compared to k-means clustering.

Data set 2: To group, the unlabelled data k-means clustering has been applied to the Wisconsin breast cancer data set [12]. In the data set, there are some records of patients, and it has to be determined whether the patients have cancer or not at the moment when the information was gathered. The cluster discovered here will be the foundation for additional research.

Table 3 Performance comparison of k-mean and agglomerative clustering

Clustering algorithm	No_of_cluster	Homogeneity	Completeness:	V-measure	Adjusted rand index:	Adjusted mutual information	Silhouette coefficient	Execution time
Agglomerative clustering	4	0.021	0.016	0.018	0.051	0.015	0.487	18.8
K-means clustering	4	0.035	0.021	0.026	0.027	0.024	0.426	58.0

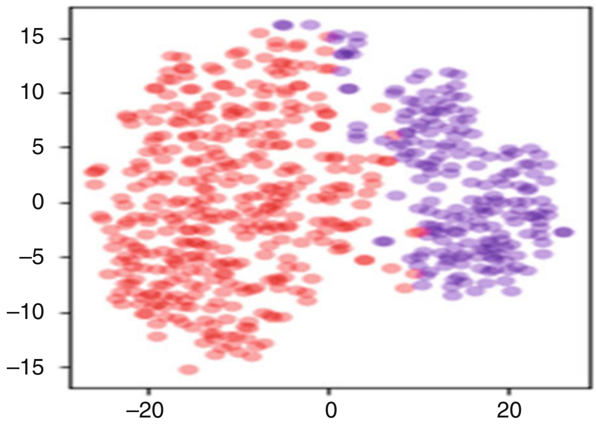


Fig. 8 K-mean clustering

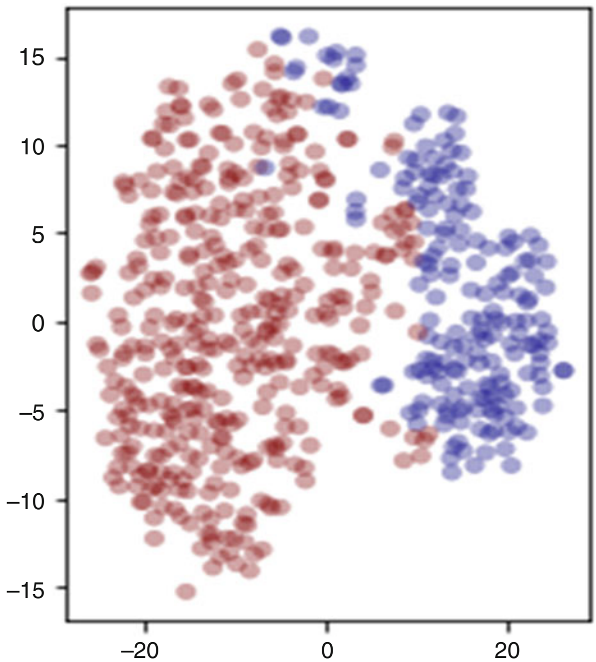


Fig. 9 Agglomerative clustering

Code 5: Load the Breast Cancer Data Set

```
Load data
 from sklearn import datasets
 dataset = datasets.load_breast_cancer()
 X=dataset.data[:,0:2]
```

After applying the k-means clustering and agglomerative clustering, the obtained clusters have been depicted in Figs. 8 and 9 (Tables 4).

Table 4 Performance comparison of k-mean and agglomerative clustering

Clustering algorithm	No_of_cluster	DB- index	Silhouette coefficient	Execution time
Agglomerative clustering	4	1.36	0.339	15.482
K-means clustering	4	1.31	0.345	39.673

It has been observed from the results that agglomerative hierarchical clustering is performing better than k-means clustering. It provides more flexibility as compared to k-means clustering and it has fewer hidden hypotheses regarding the distribution of the data. Sometimes, k-means give unpredictable results if the data are not well-separated into sphere-like clusters. In contrast, hierarchical clustering has fewer hypotheses regarding the distribution of data. It typically 'associates' nearby objects within a cluster, and then successively combines nearby objects to the most related collection. It can be computationally expensive but usually produces more intuitive results.

4 Conclusions

In this chapter, the performance of different clustering algorithm for healthcare data set has been evaluated. The performance of the algorithms has been analysed according to the number of clustered instances. According to the investigation, both the k-means and the agglomerative algorithm have efficient intra-cluster cohesion and inter-cluster distribution. The agglomerative algorithm has better performance accuracy and execution time as compared to the k-means algorithm. It is judged from the result that the agglomerative algorithm performs better as compared to k-means clustering.

References

1. A. Alsayat, H. El-Sayed, Efficient genetic K-Means clustering for health care knowledge discovery, in *2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA)*, (IEEE, Piscataway, 2016), pp. 45–52
2. D. Tomar, S. Agarwal, A survey on data mining approaches for healthcare. Int. J. Bio-Sci. Bio-Technol. **5**(5), 241–266 (2013)
3. G.R. Hobbs, Data mining and healthcare informatics. Am. J. Health Behav. **25**(3), 285–289 (2001)
4. M.C. Young, H.S. Kim, K.C. Tark, H.J. Park, S.H. Ho, Analysis of healthcare quality indicator using data mining and decision support system. Expert Syst. Appl. **24**(2), 167–172 (2001)
5. R. Nithya, P. Manikandan, D. Ramyachitra, Analysis of clustering technique for the diabetes dataset using the training set parameter. Int. J. Adv. Res. Comput. Commun. Eng. **4**(9), 166–169 (2015)
6. https://www.analyticsvidhya.com/blog/2019/10/gaussian-mixture-models-clustering/

7. G. Bruno, T. Cerquitelli, S. Chiusano, X. Xiao, *A Clustering-Based Approach to Analyse Examinations for Diabetic Patients*. 2014 IEEE International Conference on Healthcare Informatics (Verona, Italy, 2014), pp. 45–50. https://doi.org/10.1109/ICHI.2014.14
8. https://towardsdatascience.com/understanding-the-concept-of-hierarchical-clustering-technique-c6e8243758ec
9. Diabetes Data Set. http://archive.ics.uci.edu/ml/datasets/Diabetes
10. U. Maulik, S. Bandyopadhyay, Performance evaluation of some clustering algorithms and validity indices. IEEE Trans. Pattern Anal. Mach. Intell. **24**(12), 1650–1654 (2002)
11. G. Ogbuabor, F.N. Ugwoke, Clustering algorithm for a Healthcare dataset using Silhouette Score value. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **10**(2), 27–37 (2018)
12. Kaggle Breast Cancer Dataset. https://www.kaggle.com/uciml/breast-cancer-wisconsin-data
13. S. Preheim, A. Perrotta, A. Martin-Platero, A. Gupta, E. Alm, Distribution-based clustering: using ecology to refine the operational taxonomic unit. Appl. Environ. Microbiol. **79**, 6593–6603 (2013)
14. B. Jiang, J. Pei, Y. Tao, X. Lin, Clustering uncertain data based on probability distribution similarity. IEEE Trans. Knowl. Data Eng. **25**, 751–763 (2013)
15. H. Kriegel, P. Kroger, J. Sander, A. Zimek, Density based clustering. Wiley Interdiscip. Rev. **1**, 231–240 (2011)
16. G. Sheikholeslami, S. Chatterjee, A. Zhang, Wavecluster: A multi-resolution clustering approach for very large spatial databases. Proc. Int. Conf. Very Large Data Bases **24**, 428–439 (1998)

Multimodal Detection of COVID-19 Fake News and Public Behavior Analysis—Machine Learning Prospective

V. Kakulapati and S. Mahender Reddy

1 Introduction

Coronavirus-19 (COVID-19) is a varied set of viruses that can contribute to disease in living animals. Many coronaviruses have been proved to affect pulmonary infectious diseases ranging from common cold diseases to more severe conditions, such as Middle East Respiratory Syndrome (MERS). COVID-19 is the most recently identified virus. COVID-19 is an infection that is extremely contagious and has massive outbreaks worldwide. At the end of December 2019, it came out of Wuhan province of China. The 2019-nCoV infection primarily affects the ventilation system and makes it more likely to develop end-stage influenza. It can quickly be moved from human to human, and then from country to country in the form of traveling. COVID-19's accelerated dissemination risks people's lives. This pathogen affects all developing nations, such as China, Italy, Spain, and Germany. Related epidemics have existed worldwide in current history, but the virus reached 213 countries before 28 June 2020 [1].

Temperature, coughing, and lethargy are the most typical problems of COVID-19. Ailments, difficulty in breathing, fever, eye infections, sinus infection, diarrhea, loss of appetite or scent or skin rash, or finger or toe discoloration are some other signs that are less common and can affect most patients. Typically, specific symptoms may occur and begin eventually. Many individuals are becoming infected but have only very mild symptoms.

V. Kakulapati (✉)
Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

S. Mahender Reddy
Otto-Friedrich University of Bamberg, IsoSySc, Bamberg, Germany
e-mail: mahender-reddy.sheri@stud.uni-bamberg.de

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_14

Without seeking medical treatment, some patients (approximately 80%) survive from the disease—around 1 in 5 individuals who will get COVID-19 becomes critically ill and has breathing issues. Older people as those with underlying chronic conditions, such as blood pressure, complications with heart issues and lungs, diabetic mellitus, or malignancy, are at increased risk of developing severe disease. Anyone may catch COVID-19 and get severely ill. Individuals of all ages who are suffering with temperature and/or coughing combined with breathing difficulties/breathing shortness, chest pain/pressure, or lack of speech or movement should immediately seek medical attention.

With the advent of new technologies, users can connect with families, employers, and contacts on online media such as FB (Facebook), Twitter, Instagram, and LinkedIn. As a result, social practices are moving to virtual machines from real things [2]. It is difficult to evaluate people's actions on social networking sites since many strategies are used. The users' activities can be analyzed by collecting information from various databases and then by analyzing such data.

During the disease outbreak, digital platforms can be useful for alerts with vital details on existing preventive measures and risk management. It will also help bind doctors worldwide with family or friends with common reactions and those who are socially distant. The e-supplement gives more details on the power of 293 social networks during the contagion. Fake news [3] has risen since the COVID-19 epidemic, prompting real journalists and healthcare staff to provide unbiased information about the outbreak round the clock. BOOM has released 178 accurate tests on hoaxes/spreading of misinformation on the contagion since the beginning of the crisis.

As much of the misleading and disappointing claim consumers scatter over brief videos (35%), a significant percentage of voice messages (29.4%) and images (29.4%) are misinterpreted and physicalized by prominent actors by fake remedies therapeutics or fragments. A few audio-clips (2.2%) have even gone viral with fake meanings of real tests on media organizations' news coverage (4%). Many of these reports have been described as having misleading allegations about a specific culture.

In social media [4], users have learned about India TV, a media outlet that is respectful of the facts and leads the way in the battle against the outbreak. False information was spreading about social media posts that a few of the Indian TV network staff were influenza-contaminated by a man named Rajesh. The deception was not only about the network but also its workers. The person said, keep away and be careful if you live anywhere near any Indian Television employees. Then, maybe the user would still be talking about India Broadcasting. However, this guy intentionally misleads regarding the current Indian TV employees. Considering false propaganda about their employees, India TV registered a case against the person, after which the Noida police arrested him.

Stories have circulated across digital platforms in many ways, including email, photographs, and organizations. Online evidence checking helps to show that convergent information is collected and consolidated. The RNN multimode extraction and synthesis structure for gossip detection [5] are examples of such techniques. For this reason, neural adverse networks have been established by training about the invariable perception of an event [6] to recognize multimodal false information, which reduces tight reliance on particular circumstances in the training data to increase these events' generality. In the early phases of news dissemination [7], three connections are rooted in publishing new relationships, and the experiences between consumers and news at the same time to detect false information.

Multimodal false news identification typically focuses on the extraction of news source elements, namely news provider and text information. The false news is classified using frequency and multiple machine learning algorithms—a multimodal news identification system. The architecture suggested is intended to identify whether a single news story is true or false.

2 Related Works

A variety of hoax identification strategies have been recently acquired. A strongly connected scientific research on identifying counterfeit posts on social media is addressed in this segment. In compliance with the nature and format of usable false news data and technological approximations, fake news identification has been investigated using different methods [8]. The roles of false news data sources include checking whether the title refers to the body of the message, detecting incorrect phrases in the body text, and recognizing fake news dissemination through social media platforms. In general, such approaches use techniques, such as deep learning (DL) [9], machine learning (ML) [10], and rule-based methods [11] for detection purposes.

The topic around COVID-19 has started to grow, and peer relationships are going online with an increasing array of newspapers and corporates employed in social media [12]. The technical and social infrastructure, which enables us to remain linked even through crises, has become core hubs like Twitter. The authors describe the online discussions with the research group on Twitter on COVID-19. People worldwide use Twitter to share their views and join in dialogues at an official forum, which has proved to be an excellent platform for multitude of discussions with Twitter's open application programming interface (API). The intellectual community has been using Twitter for years to understand trends that can be found in online social networks, varying from press releases to the spread and influences [13] of bots and misinformation [14]. Most notably, Twitter allows researchers to analyze social media's role in public health crises during the latest COVID-19 pandemic [15]. Users hope that this knowledge will inspire more studies on the social aspect of the outbreak.

The negative side of digital media is constantly rising and harms communities with people and consumers' rights [16]. Researchers argue that certain outlets encourage psychologically manipulative behaviors, such as self-advertising, moral dirtiness, replication, aggressivity, and sociopathy [17]. These adverse effects of social media significantly impact physiological, physical, social, and mental health [18]. Also, the harmful elements of social networks endanger our culture and businesses as a whole. The most significant adverse effects concern Internet explosions and antagonism shared [19], the perception that companies are stressed at work, and work–life conflicts [20], and the dissemination of users' smears or false news [21].

Counterfeit news on online media belongs to the video, audio, and/or texts which propagate false information [22]. This paper explores the use of social media and activity regarding "fake news," undertaken in the United States in early 2018, and general news that does not apply directly to circumstances of catastrophe. In this sense, though SM users' general conduct is vital for crises, it is conceivable that consumer's behavioral habits in daily life are more significant for emergencies, as people are willing to communicate rapidly. While Hughes and Palen in a Twitter analysis do not contribute specifically to false stories, it seems that other information-sharing practices take place throughout crises [23]. Gupta and collaborators have also shown that misinformation in medical emergencies can lead to confusion and disruption in social media [24]. Additionally, more study is required if "regular" activity is connected to evaluations of messages' reliability.

The wide spread of misleading information will adversely affect people and culture. Second, the credibility of the media outlet may be breached by false stories. For instance, Facebook's most common misinformation was much more comprehensive than the most commonly recognized authentic conventional news during the US Chairperson election in 2016 [25]. Second, propaganda deliberately persuades consumers to adopt partial or myths. False information is commonly a distorted information to communicate political ideas through propaganda machines. For instance, few surveys indicate Russia's fake profiles and SocBots to disseminate fabricated stories [26]. Third, the perception of fake news and how people react to actual news. For instance, some false information was only generated that caused mistrust and misguided people, preventing them from differentiating between reel and real [27]. To minimize the detrimental impact of yellow journalism—both for the community and for the news ecosystem—there is a need to build ways to classify hoaxes instantly.

There are some similarities between digital rumors and misinformation reporting, and both require knowledge transfer. An increase in negatively biased news sharing in social media is increasing very alarmingly, primarily when more people rely on digital media, for example, up to 62% [28]. In a recent survey, a substantial portion of people expressed confidence in the pseudo-news they saw during the 2016 US Presidential Election [29], and the dangers raised by the dissemination of misleading information are also illustrated. The full impact of information exchanged on digital media is that it is challenging to accurately identify its authenticity, hence emphasizing the need to improve systems efficiency to avoid propagating gossips. Nevertheless, there is a scarcity of observational research investigation exploring the multiple histories of false news dissemination by social media users, who have an essential requirement to gossip.

Unimodal technology has been able to deliver positive outcomes. In knowledge retrieval, brief and informal social media data are still a challenge. The investigators began to test the characteristics derived from different approaches (i.e., text and images) to solve this constraint and condensed it on a more robust statistical analysis. Works detailed in Ref. [30] are the most impressive experiments in the identification of multimodal content. In [31], the authors developed an end-to-end structure for the title and denial of misinformation. This is regarded as a multimodal false news analysis, event adversarial neural networks (EANN). There are two factors in their approach. The word component integrating vector was used as a component, and message interpretation was created with the conventional neural network [32].

The "probabilities that anything will develop, provided that perhaps it has already happened" from Bayes theorem derives Naive Bayes, which is utilized to quantify conditional distribution [33]. Thus, by using previous experience, users can determine the chance of a given outcome.

3 About COVID-19 Fake News

Deception is inaccurate news related to facts. It also attempts to harm an individual or entity's image or raise money through publicity revenues. Formerly prevalent in newspapers, with the development of social media, particularly the FB News Stream, the frequency of fake news has risen. For brands and products, the dissemination of misleading information poses significant challenges. Indeed, a counterfeit report which promotes, but cannot be right, a particular perspective or impression on a product, brand, or organization may be purposely conceived to manipulate consumers [34].

Also, as the specific strain of infections, the COVID-19 pathogen occurs, it has an extensive knowledge deficit. This has contributed to a global need for searched and recorded contents of the new virus, while also supplying people with an incentive to fill this vacuum with misleading info. With the immense volume of material entering social outlets, more by the minute, the distinction between the believable and the fake becomes increasingly smooth.

The propagation of deception across COVID-19 led to policymakers and technology leaders rushing and taking stringent action to stop it. In some countries, prosecutions have been given, and other extreme fines have been levied to limit misinformation dissemination [35]. The pandemic of coronavirus has spawned many reports, including conspiracy theories and fake information, and has saturated Facebook's algorithm with ample postings to train its circuit breakers program to recognize problems that could be problematic [36]. A knowledge center on its software is part of other initiatives to curb "myths" about COVID-19—approximately, 8.3 billion views on FB users are created by misleading healthcare platforms in 2019.

The center of the problem with misleading information [37] is known in such severe situations in segments with two most extremely negative and very far-reaching results; first, any misleading news broadly distributed will cause widespread confusion and panic, and negatively influence health status and user being. For instance, if rumors circulate that apply to such an area of a city that is no longer locked up, people in the neighborhood could come out in aggregate and trigger the infection of COVID-19. Furthermore, any misleading news defaming or aggravating such a community association on an individual or a society will use an uncivilized action that takes itself and others at risk into consideration.

In the sense of a global pandemic, even though the substance is harmful in the safest possible case, perpetuating misinformation can have severe and even fatal health implications. Rumors of the potential food crisis have led people earlier in the outbreak to store food and have caused a significant pandemic in many countries. Since reporting on hydroxychloroquine as a potential, though not yet confirmed, cure for COVID-19 care in the United States, a person was killed in consuming a fish tank washing liquid that contains chloroquine [38]. Hundreds died of digital platform stories in the Islamic state after liquor consuming methanol that healed other people from the coronavirus. This is harmful misinformation about which the World Health Organization is most concerned.

4 Methodology

In the online media, multimodality defines information-sharing activities in terms of text-based, anamonic, graphical, geographical, and graphical means—or methods—used to write notifications (or posts) [39]. This chapter concentrates on the combination of both. The methodologies of digital platforms are described and the conventional methods of integrating them into the proposed model. It should be remembered that the models presented are readily extendable to more modalities.

Multimodal Classifiers (MC)

The purpose of supervised learning is to categorize patterns into a set of classes. The ensemble methodology's main idea is to weight several individual classifiers and combine them to obtain a classifier that outperforms individual classifiers, also called late fusion [40] in a multimodal approach. Analytically, ensembles have a tendency to produce improved consequences as soon as there is significant diversity among models [41]. Many ensemble methods, therefore, seek to promote diversity among the models they combine. In the ensemble fusion model, texts and images are first processed separately to provide decision-level results [42]. The observations are then consolidated in two ways: techniques of multimodal and strategies of meta-learning. Weighting procedures are useful for the same purpose and equal performance of the classification models. Semantic-learning methods are perfectly suited for circumstances that certain deep learning models identify or mislabel specific situations consistently.

Support Vector Machines (SVM)

Several research applications utilize a machine learning model to explore the creditworthiness of datasets. SVM is a pair-wise evaluation method that implements SVM [43]. Stopping terms have been omitted from textual information. The characteristics are effectively retrieved. SVM is a method of the equivalent supervised learning algorithms for categorization. SVM classifier then classifies the data after extracting text features and sets fake stories or actual news.

```
Algorithm 1:  Input: D dataset, on-demand functionality,
aggregation-based functionality
2) output: news categorization
a) do news id in D for any application
b) Provide functionality on request and placed on news
   id vector x.
c) x. add (Get(news ide));
d) the termination for the
e) do x vector for any program
f) Retrieve the first element and the characteristics
   placed in b, and w.
g) hw, (x) = (z) z= (wT x + (b))
h) if (z portion 0)
i) allocate g(z)=1;
j) g(z)=-1 otherwise;
k) end if
l) end for
```

Stochastic Gradient Descent (SGD)

An algorithm called gradient descent is utilized for optimizing neural networks, the much more commonly utilized approach to optimize them. Excessive computation is carried out on massive data by the pool gradient descent because before every variable change, it recomputes gradients for relevant instances. By upgrading one at a time, SGD removes this redundancy. SGD is much easier and can be used just for online learning. SGD carries out periodic updates that cause the goal feature to fluctuate significantly.

Gradient Boosting (GB)

Learning ensemble [44] integrates different simple classifiers for a solid forecasting model and typically achieves highly accurate prediction. One of the benefits of ensemble methods is that baseline classifications do not need high precision to achieve the ultimate maximum predictive model accuracy [45]. Learning together will break a complicated issue into several minor problems that are easier to comprehend. Gradient boosting is a machine-literacy technique that generates a prediction model in a series of low prediction models, usually decision trees. Gradient boosting progressively and sequentially trains multiple models. Gradient improvement is a greedy algorithm, which can easily overfit a dataset. It will gain from laws that penalize various portions of the method and typically increase the algorithm's efficiency by minimizing duplication.

Bounded Decision Trees (BDT)

For some types of computational problems and algorithms, decision tree models are instrumental in setting lower limits for complexity theory. Its decision tree complexity or query complexity is called the computational complexity of a problem or an algorithm embodied in the decision tree model.

Random Forests (RF)

Decision Tree (DT) classifier becomes one of the most commonly utilized grades. It is also a mighty classifier. As with SVM, DT can correlate as it can classify. RT is a categorization algorithm that generates a collection of classification trees. Also, every categorization tree is constructed with a bootstrap data sample, with the nominee set of variables being a random subset of variables at each division. Random forests then use bagging (bootstrap aggregation), the efficient mix of insecure learners, and random tree building variable collection. To get low bees, each tree needs to be unpruned. Thus, bagging and random variable collection contribute to the low correlation of each tree. The algorithm generates a population of average over a large group of short, broad, yet short correlation trees.

5 Implementation Results

The framework of Fig. 1 demonstrated the proposed fake media identification scheme. For training, the three corpora, datasets from the news stories, and Kaggle site have been obtained from three different outlets. In the preprocessing stage, the newspapers erase stop words and repeat content. After the next step, missing values are cleaned. The dataset is divided into training and test datasets. To extract essential characteristics from the text files, a function removal process will be performed. The features are taken from the media in this process. Techniques like term frequency is added for extracting features. The recovered functions are then loaded into another process classification algorithm. Different machine learning methods, including SVM, GD, BDT, SGD, and the RF, choose to learn and categorize the trends and effects. The forms are then tested to obtain an accurate classification based on output assessments. The technologies are combined to give a multimodel to achieve high performance based on research.

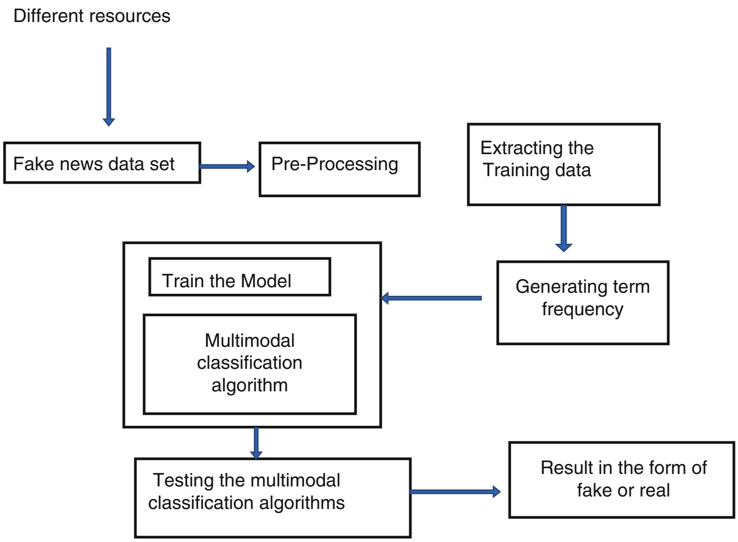


Fig. 1 Framework of multimodal fake detection

The dataset contains fake news.

“1 Chinese scientists created the coronavirus in a secret laboratory"
"2 drinking bleach or eating garlic cures corona infection"
"3 the new 5g technology caused the sickness."
"4 paracetamol cures corona and bleaching powder can kill virus."
"5 "use of coconut oil" ginger garlic to cure the virus."
"6 home treatments can cure or prevent people from contacting the virus."""
"7 eating garlic as part of regular meals are welcomed as antidote for coron a virus."
"8 use of sodium chloride mixture with citric acid"
"9 new research from Harvard university shows that the chemical in chilli p eppers that causes the hot sensation in your mouth reduces the replication rate of coronaviruses. The researchers investigating whether adding more sp icy foods to your diet could help combat covid19 or corona."
“10. corona vaccine being developed by the company causes a high rate of co mplications" but that these concerns were being disregarded in favour of r eleasing the vaccine quickly."

From the dataset, words are retrieved, and they are like

"corona, virus, cure, cause, kill, people, garlic, research, contact, diet, infect, china, covid, chines, create, concern, high, rate, vaccine, conduct”.

Term frequency is a weighting matrix utilized in a misleading information dataset to calculate a word (count + weight). The tokens are used to compile using TF technologies and the weights are allocated to the tokens from the text data. Term Frequency referred to as tf and computed by count (c) and the term (t).

Non/Sparse entries	: 287/5890
Sparsity	: 95%
Maximal term length	: 18
Weighting	: Term Frequency (tf)

Word probabilities:

The immune system of caffeine.: 0.03448276,
Garlic in food: 0.03448276
Chinese Government Murdering Corona Virus Citizens: 0.03448276
Corona virus spread will increase cold diet: 0.03448276
More alcohol intake reduces corona risk: 0.03448276

Table 1 The topics of the selected dataset

	Topic 1	Topic 2	Topic 3	Topic 4
1	China	Kill	Virus	Corona
2	High	Cause	Cure	Research
3	Universe	People	Contact	Infect
4	Design	Garlic	Conduct	COVID
5	Support	Chinese	Govern	Concern
6	Treatment	Create	Reduce	Trace

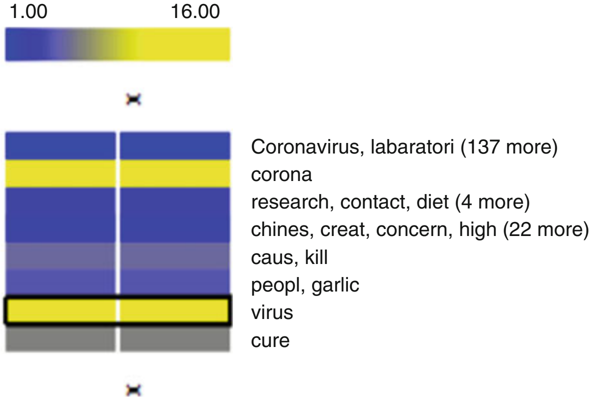


Fig. 2 Heatmaps of fake word tokens for frequency

The company's corona vaccine causes many risks, but these issues have been ignored, and rapid release of the vaccine, Corona, and powder bleaching will destroy the virus: paracetamol cures: 0.03448276

Encouragement of hydroxychloroquine as a coronavirus remedy, mutating to kill the virus by ingesting disinfectant.: 0.0344810

Mouth rinsing will destroy or swallow a fair amount of liquor, the virus is before the body is compromised 0.03448276

Using cocoon milk, ginger, curative garlic: 0.03448276 consuming vitamin c and other diets may cause the virus weak: 0.03448276 (Table 1, Figs. 2, 3, 4, 5, 6, and 7)

6 Conclusion

The concern of misleading information has become a trendsetting concept, and these stories will cognitively manipulate patients and change their actions. It is even more critical than the outbreak everywhere else. To educate the community of what steps are needed to be healthy and vigilant during the COVID-19, mostly in social media, this research provided perspectives that can help predict yellow journalism expectations and how to properly build technology to encourage the appraisal of media coverage and other content trusts. When users can evaluate the truthfulness

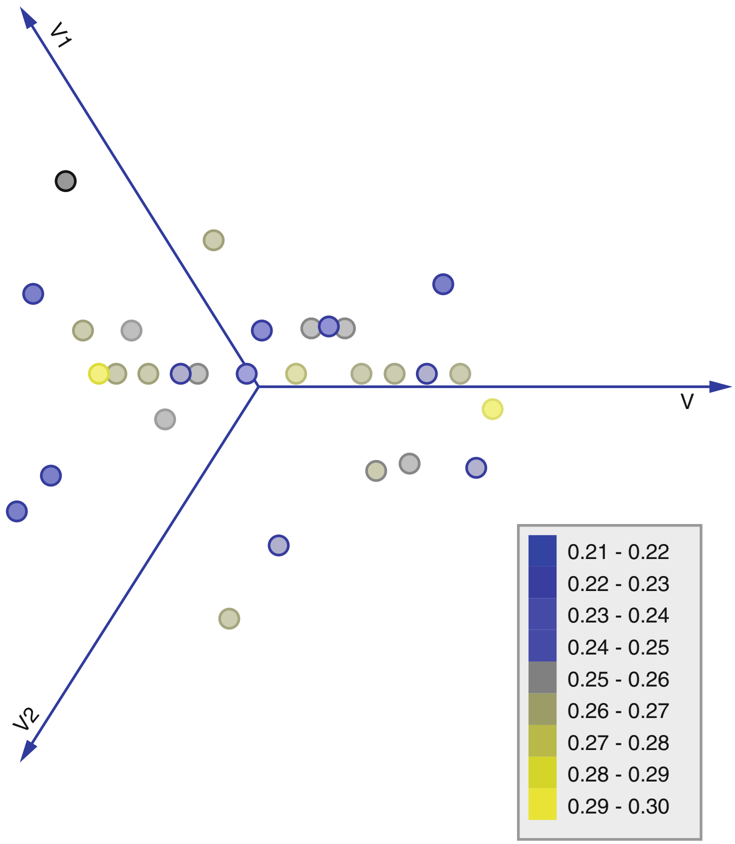


Fig. 3 Linear projection of features of after SVM processing

of different social media platforms in connection with a disaster, they can raise awareness of the issue, trust planned measures to optimize their protection, and reduce casualties and lead to adaptability to fake news from multimodal classification structures, such as SVM, GD, SGD, BDT, and RF. A multimodal framework is sufficient for reliable performance. They assess how counterfeits in spreading information are correlating with socio-psychological elements in engaging various social media platforms to provide some promoters with perspectives. The research findings indicate useful customized solutions, for example, emphasizing digital confidence and reducing frustration to minimize social media behaviors' negative implications.

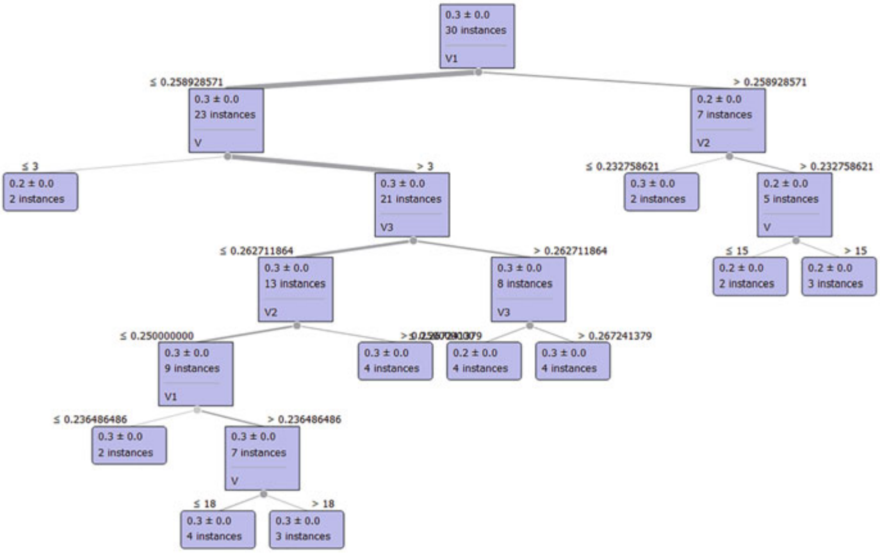


Fig. 4 Bounded decision tress of fake news

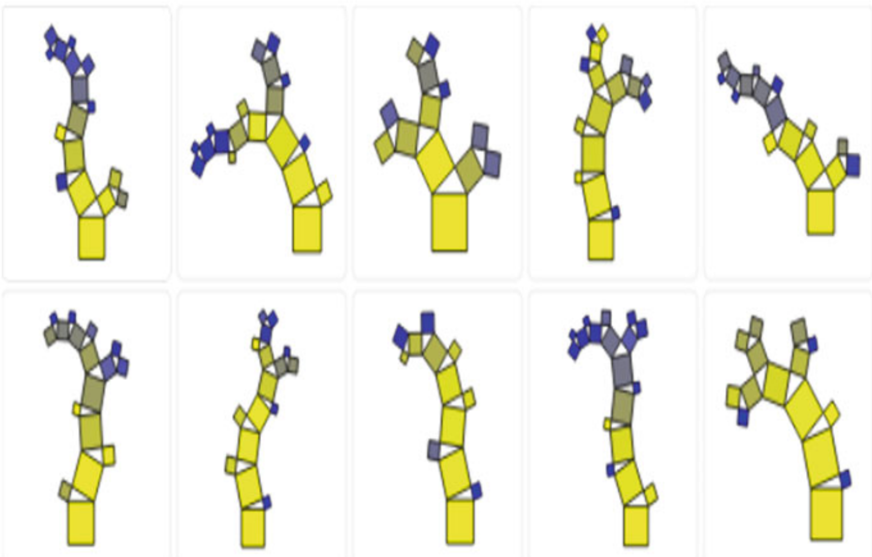


Fig. 5 The Pythagorean display of random forest of fake word probabilities

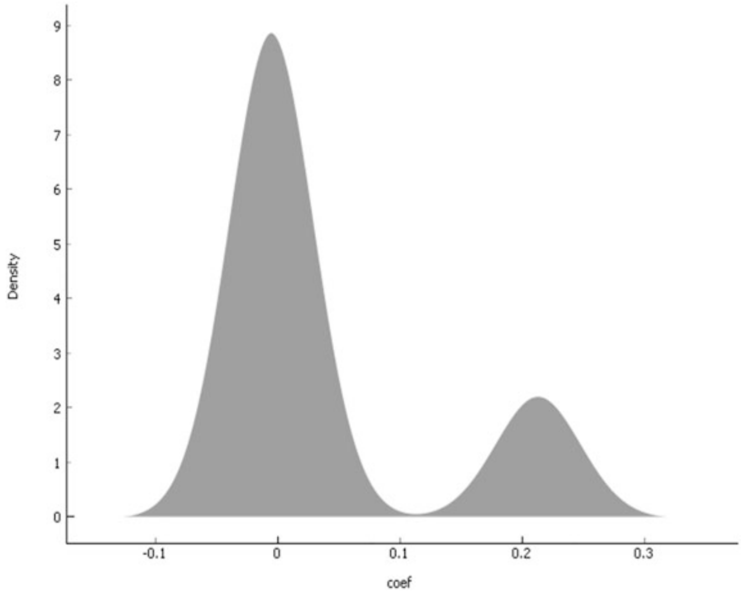


Fig. 6 Stochastic gradient descent distributions

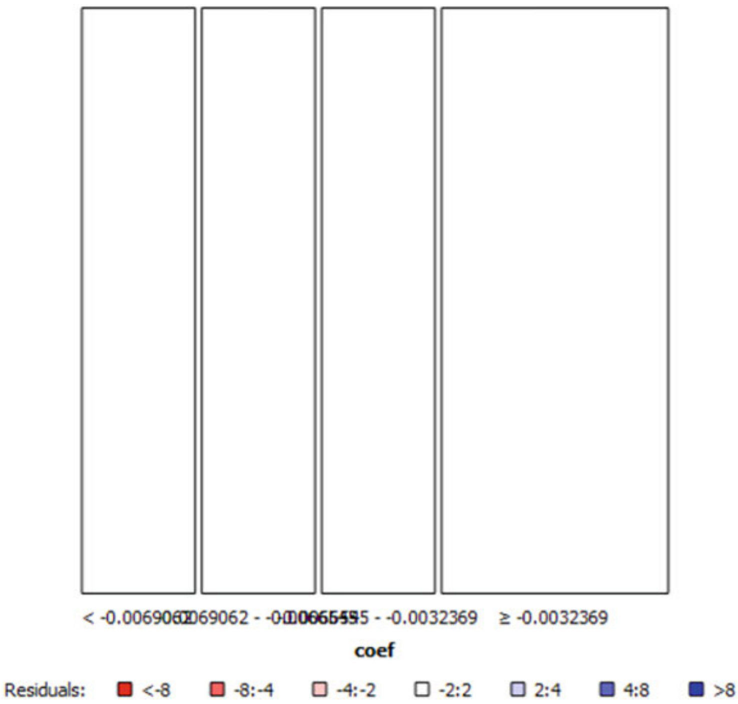


Fig. 7 Stochastic gradient descent mosaic display

7 Future Scope

Misinformation that has significant adverse effects on human person and the broader community has also been utilized in the digital platforms. The false news issue was discussed in this chapter. We intend to develop multimodal hierarchical approaches for incremental learning in future investigation. When new examples arrive, increasing understanding requires restoring those layers. The regeneration time of a structure can be minimized in a categorical methodology through information from the next layers. The samples, validation steps, and exciting potential strategies for misleading analysis into news detection will be addressed in the future.

References

1. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-ansusersrs-hub/q-a-detail/q-a-coronaviruses#:~:text=symptoms
2. L.C. Lukito et al., Social media user personality classification using computational linguistic, in *Proceedings of ICITEE*, (Tangerang, Indonesia, 2016)
3. https://www.indiatvnews.com/news/india/noida-police-fake-facebook-post-coronavirus-rumours-619728
4. https://www.boomlive.in/fact-file/fake-news-in-the-time-of-coronavirus-a-boom-study-8008
5. Z. Jin et al., Multimodal fusion with recurrent neural networks for rumor detection on microblogs, in *Proceedings of ACM Multimedia 2017*, (2017), pp. 795–816. https://doi.org/10.1145/3123266.3123454
6. Y. Wang et al., EANN: Event adversarial neural networks for multi-modal fake news detection, in *Proceedings of KDD*, (2018), pp. 849–857
7. K. Shu et al., beyond news contents: The role of social context for fake news detection, in *Proceedings of WSDM*, (2019), pp. 312–320
8. W.Y. Wang, "Liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv:1705.00648 (2017)
9. Popat et al., Debunking fake news and false claims using evidence-aware deep learning. arXiv:1809.06416 (2018)
10. Rasool et al., Multi-label fake news detection using multi-layered supervised learning, in *Proceedings of the 2019 11th International Conference on Computer and Automation Engineering*, (ACM, Perth, Australia, 23–25 February 2019), pp. 73–77
11. S. Feng et al., Syntactic stylometry for deception detection, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers—Volume 2*, (Jeju Island, Korea, 8–14 July 2012), pp. 171–175
12. A. Abbas et al., "Data set on coping strategies in the digital age: The role of psychological well-being and social capital among university students" in Java Timor, Surabaya, Indonesia. Data Brief, 105583 (30 Jun 2020)
13. K. Lerman et al., *Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks* (AAAI Publications, Washington, DC, 2010)
14. C. Castillo et al., Information credibility on twitter: Association for Computing Machinery, in *The 20th International Conference on World Wide Web 2011*, (New York, 2011), pp. 675–684
15. E. Ferrara et al., The rise of social bots. Commun. ACM **59**(7), 96–104 (24 Jun 2016)
16. C. Baccarella et al., SM? It's serious! Understanding the dark side of SM. Eur. Manag. J. **36**, 431–438 (2018)

17. J. Fox et al., The Dark Triad and trait self-objectification as predictors of men's use and self-presentation behaviors on social networking sites. Pers. Indiv. Differ. **76**, 161–165 (2015)
18. Immordino-Yang et al., Rest is not idleness: Implications of the brain's default mode for human development and education. Perspect. Psychol. Sci. **7**(4), 352–364 (2012)
19. K. Rost et al., Digital social norm enforcement: Online firestorms in SM. PLoS One **11**(6), e0155923 (2016)
20. W. van Zoonen et al., Understanding the consequences of public social media use for work. Eur. Manag. J. **35**(5), 595–605 (2017)
21. H. Allcott et al., Social media and fake news in the 2016 election. J. Econ. Perspect. **31**(2), 211–236 (2017)
22. D.M. Ordway, Fake news and the spread of misinformation. Journalists Resource (9 January 2017)
23. Hughes et al., Twitter adoption and use in mass convergence and emergency events. Int. J. Emerg. Manag. **6**(3–4), 248–260 (2009)
24. A. Gupta et al., Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy, in *Proceedings of the 22nd International Conference on World Wide Web*, (ACM, 2013), pp. 729–736
25. M. Balmas, When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. Commun. Res. **41**(3), 430–454 (2014)
26. M. Banko et al., Open information extraction from the web, in *IJCAI'07 Proceedings of the 20th International Joint Conference on Artificial Intelligence*, (2007), pp. 2670–2676
27. A. Bessi, et al., Social bots distort the 2016 us presidential election online discussion. First Monday **21**(11) (2016)
28. J. Gottfried, et al., News use across social media platforms 2016 (26 May 2016) http://www.journalism.org/2016/05/26/news-use
29. H. Allcott et al., Social media and fake news in the 2016 election. J. Econ. Perspect. **31**(2), 211–236 (2017)
30. Z. Jin et al., Multimodal fusion with recurrent neural networks for rumor detection on microblogs, in *Proceedings ACM on MM 2017*, (Mountain View, CA, USA, 23–27 October 2017), pp. 795–816
31. Y. Wang et al., EANN: Event adversarial neural networks for multi-modal fake news detection, in *Proceedings of the 24th ACM SIGKDD18*, (ACM, New York, NY, USA, 2018), pp. 849–857
32. Y. Kim, Convolutional neural networks for sentence classification. CoRR, vol. abs/1408.5882 (2014)
33. R. Saxena, How the Naive Bayes Classifier Works in Machine Learning. Retrieved 20 Oct 2017 (2017)
34. M. Potthast, et al., A stylometric inquiry into hyper partisan and fake news. arXiv:1702.05638 (2017)
35. https://www.thequint.com/news/usersbqoof/fines-arrests-how-governments-around-the-world-are-curbing-covid-19-fake-news-coronavirus
36. https://economictimes.indiatimes.com/magazines/panache/facebook-working-on-virality-circuit-breaker-to-identify-fake-news-before-it-goes-viral/articleshow/77733541.cms
37. https://www.mondaq.com/india/operational-impacts-and-strategy/942532/misinformation-and-covid-19-an-analysis
38. https://www.who.int/news-room/feature-stories/detail/immunizing-the-public-against-misinformation
39. C. Lutkewitte, *Multimodal Composition: A Critical Sourcebook* (Bedford/St. Martin's, Boston, 2013)
40. P.K. Atrey et al., Multimodal fusion for multimedia analysis: A survey. Multimedia Systems **16**(6), 345–379 (2010)
41. L. Rokach, Ensemble-based classifiers. Artif. Intell. Rev. **33**(1), 1–39 (2010)
42. Y. Peng et al., Multimodal ensemble fusion for disambiguation and retrieval. IEEE MultiMedia **23**(2), 42–52 (2016)

43. S. Gilda, Evaluating machine learning algorithms for fake news detection, in *2017 IEEE 15th Student Conference on Research and Development (SCOReD)*, (IEEE, Putrajaya, 2017)
44. Z. Zhou et al., Deep forest: Towards an alternative to deep neural networks, in *Proceeding of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, (IJCAI, 2017), pp. 3553–3559
45. X. Yuan et al., A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. Pattern Recogn. **77**, 160–172 (2018)

A Review of Machine Learning Approaches in Clinical Healthcare

Pooja Dixit, Manju Payal, Vishal Dutt, and Ankita Tuteja

1 Introduction

Health problems have an adverse effect on human life. Health providers collect clinical data about every specific patient during medical care to determine how to treat that patient using the leverage knowledge from the general population. Thus, data play a basic role in the resolution of health problems, and better information is important to improve patient care [1].

In health services, there is enhanced utilization of information systems. Digitization of patient data produces a lot of information. These data are stored in several systems in various formats. In Fig. 1, almost 80% of health data are unformed. In provision of healthcare services, the data have high potential. It is the primary utilization of healthcare services. It is helpful to enhance the quality of secondary utilization. It also improves the management, research, and planning. It consists of different types of the facilities. The first facility consists of huge amounts of data. The second facility is the enhanced computing power. The third facility is the improved methods for machine learning. This facility makes fast and automated production. The ML methods enable to analyze complex data with actual outputs. Here, ML is referred as the machine learning [2] (Fig. 2).

P. Dixit (✉)
Bundelkhand University Jhansi, Jhansi, Uttar Pradesh, India

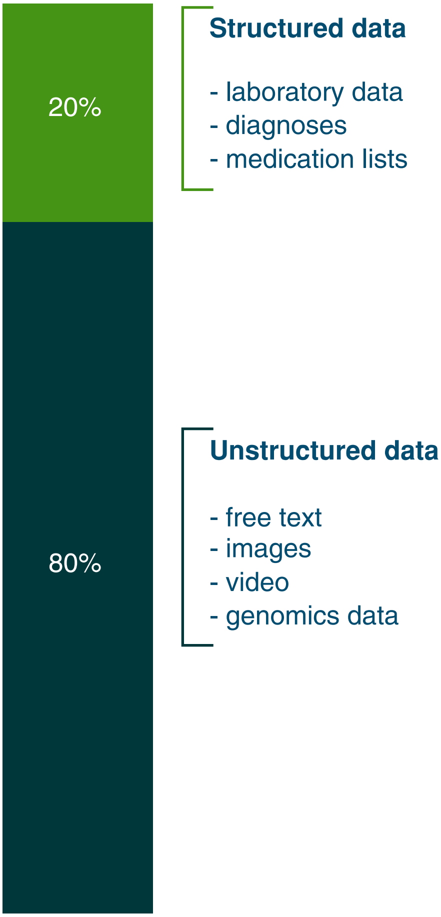
M. Payal
Software Developer, Academic hub Ajmer, Ajmer, Rajasthan, India

V. Dutt
Aryabhatta college, Ajmer, Rajasthan, India

A. Tuteja
Chitkara University School of Engineering and Technology, Chitkara University, Solan, Himachal Pradesh, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_15

Fig. 1 Types of health data



ML methods can analyze huge amounts of several types of data with actual results. There are three types of disruption fields for ML in healthcare which are as follows:

1. First disruption is related to the interpretation of medical images. There are some examples available in radiology, eye diseases, pathology, and so on.
2. Second disruption is the prognostics. There are some types of instances available which are metastatic cancer, dementia, stroke, and so on.
3. Third disruption is related to the interpretation of medical images. There are some examples available which are rare diseases, oncology, pathology, and so on.

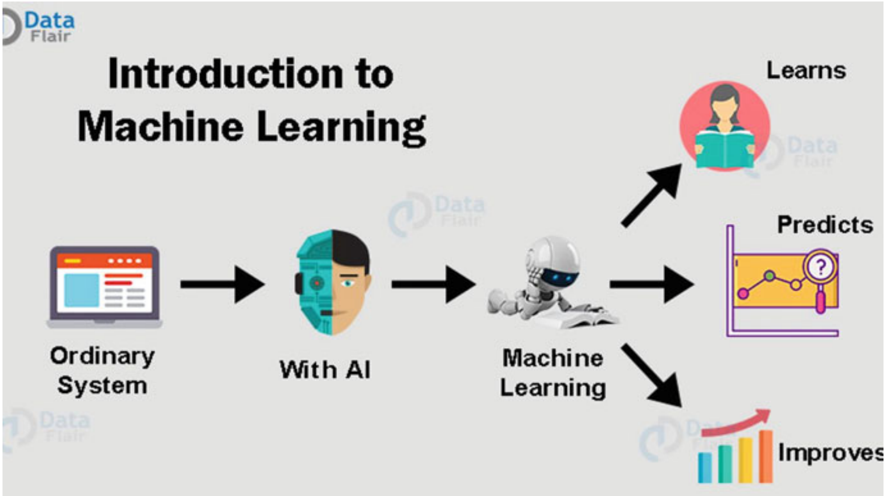


Fig. 2 Introduction to machine learning

2 Applied Machine Learning in Healthcare

ML in medicine has made headlines recently. ML is referred as the machine learning. Google has created a method which is used to recognize cancer on mammograms. This method is developed in machine learning. For identifying skin cancer, there is use of one method which is known as the deep learning method. This method is used by Stanford. A recently published article revealed the output of a deep ML method that had the option to analyze diabetic retinopathy in retinal pictures. It is clear that ML puts other arrows in the quiver of medical decision-making [3].

Nevertheless, ML provides itself to few processes more than others. Methods can give instant advantages with processes that are standardized or reproducible. ML is referred as machine learning. In addition, datasets with the biggest image, for example cardiology, radiology, and pathology, are robust applicants. ML can be accomplished to identify abnormalities, visualize image, and view regions that require responsiveness, thus enhancing the accuracy of the whole process. In the long run, ML will bring in advantage to the family internist or practitioner. ML can provide goals and opinions to enhance reliability, accuracy, and efficiency.

3 The Ethics of Using Algorithms in Healthcare

It's been said before that AI is the simplest tool in clinical treatment. Could there be an inclination for specialists to think about AI to be a bothersome second supposition? At one point, autoworkers expected that mechanical innovation would filter out their positions. So likewise, there could also be specialists who fear that AI is the beginning of a cycle that would convey them outdated. Regardless, it is the art of prescription which will never be displaced. Patients will reliably require the human touch, therefore the careful and sympathetic relationship requires extra care. Neither any techniques of machine learning nor another future advanced technique in medicine will replace this. but the advanced tools are required in clinical, that improve the health care [4].

The attention ought to be on the best way to utilize AI to enlarge understanding consideration. For instance, if doctor test a cancer in a patient, at that point, they need the greatest biopsy results, that can get. An AI that can survey the pathology slides and help the pathologist with a finding is significant. In the event they can get the outcomes in a short time with an indistinguishable level of precision, at that point, at last, this will improve quiet consideration and fulfillment [5].

4 Current Machine Learning Healthcare Applications

The index below is by no means exhaustive, but provides a usable landmark of some of the effects of ML in the healthcare business.

1. **Identifying Disease Diagnosis in Medical Imaging**

   One of the focal ML applications in clinical consideration is that the distinctive confirmation and investigation of diseases and burdens which are for the foremost part seen as hard to dissect. This will fuse anything from dangerous developments which are hard to seek out during the elemental stages to other genetic contaminations. IBM Watson Genomics is an unprecedented portrayal of how organizing scholarly enlisting with genome-based tumor sequencing can help in making a brisk finding. Berg, the biopharma goliath, is using AI to form accommodating medications in regions, for example, oncology. P1vital's redicting Response to Depression Treatment (PReDicT) expects to develop a conceivable way to affect investigate and provides treatment in routine clinical conditions [6].
2. **Treatment Queries and Suggestions**

   Diagnosis is an extremely complex cycle, and includes – at least for now – a bunch of factors (everything from the color of white of a patient's eyes to the food they have for breakfast) of which machines can't presently collate and make sense; in any case, there's little uncertainty that a machine may help in helping doctors to make the correct contemplations in analysis and treatment, essentially by filling in as an expansion of logical information.

   That is the thing that Memorial Sloan Kettering (MSK's) Oncology division is focusing on in its ongoing association with IBM Watson. MSK has reams of information on malignant growth patients and therapies utilized over decades, and it is ready to introduce and propose therapy thoughts or choices to specialists in managing remarkable future disease cases – by pulling from what worked best before. The sort of a knowledge-expanding apparatus, while hard to sell into the hurly-brawny universe of emergency clinics, is now in premier usage [6].

3. **Scaled-Up / Crowd-Sourced Medical Data Collection**

   There is a lot of spotlight on pooling information from different cell phones in sequence to aggregate and comprehend all the more live well-being information. Apple's ResearchKit is intending to do this in the treatment of Parkinson's illness and Asperger's disorder by permitting clients to get to intuitive applications (one of which applies AI for facial acknowledgment) that evaluate their conditions after some time; their utilization of the application takes care of continuous advancement information into a mysterious pool for future investigation.

In spite of the enormous storm of medical services information given by the web of things, the business actually is by all accounts testing in how to understand these data and make ongoing changes to therapy. Researchers and patients can be hopeful that, as this pattern of pooled user information proceeds, specialists will have more ammunition for handling tough illnesses and different cases [7].

4. **Drug Discovery**

   While a significant part of the medical services industry is a swamp of laws and confusing motivations of different partners (emergency clinic CEOs, specialists, attendants, patients, insurance agencies, and so forth . . .), drug disclosure stands apart as a moderately clear monetary incentive for AI medical service application makers. This application additionally manages one moderately clear client who happens to for the most part have profound pockets: drug organizations.

IBM's own well-being applications has had activities in drug revelation since its initial days. Google has likewise bounced into the medication revelation conflict and joins a large group of organizations previously collecting and bringing in cash by chipping away at drug disclosure with the assistance of ML. Here, ML is referred as the machine learning [8].

5. **Robotic Surgery**

   The da Vinci robot has got the majority of consideration in the mechanical medical procedure space, and some could contend it in light of current circumstances. This gadget permits specialists to control dexterous mechanical appendages so as to perform medical procedures with fine details and in restricted spaces (and with less quakes) than would be conceivable by the human hand alone. While not all automated medical procedure methodologies include ML, a few frameworks use PC vision (helped by ML) to recognize separations, or a particular body part, (for example, distinguishing hair follicles for transplantation on the head, on account of hair transplantation medical procedure). Likewise, in some cases ML is used to consistent the movement and development of robotics limbs when taking commands from human regulators. Here, ML is referred as the machine learning [9].

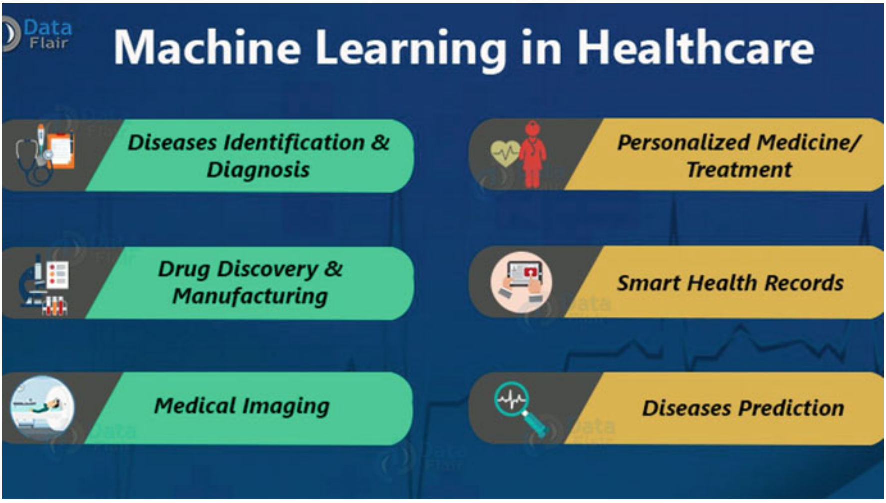


Fig. 3 Machine learning in healthcare

6. **Smart Fitness Reports**
   Keeping up healthiness records is a comprehensive cycle, while technology has played its part in easing the data entry process, actually, even now, a bigger aspect of the process takes an enormous amount of time to conclude. The key capacity of AI in clinical consideration is to ease cycles to save lots of time, effort, and money. Report portrayal methodologies using vector machines and ML-based OCR affirmation methodology are continuously gathering steam, for example, Google's Cloud Vision API and MATLAB's AI-based handwriting affirmation advancement. ML based health system is today at the bleeding fringe of working up the cutting-edge period of astute, splendid prosperity records, which can meld ML-based costs beginning from the soonest stage to assist with investigation, clinical treatment recommendations, etc. [5] (Fig. 3).

5 Machine Learning Algorithms

ML methods mainly separate features from information, for example, patients' "qualities" and clinical results of intrigue (Fig. 4).

For quite a while, AI in medical services was overwhelmed by the calculated relapse, the most basic and normal calculation when it is important to arrange things. It was anything but difficult to utilize, snappy to complete, and simple to decipher. Nonetheless, in the previous years, the circumstance had changed and support vector machine and neural organizations have started to lead the pack [10] (Fig. 5).

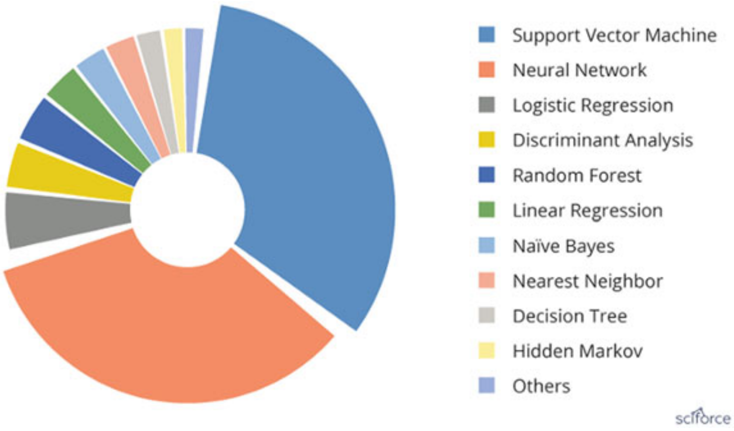


Fig. 4 The most mainstream machine learning calculations utilized in the clinical writing. The information is produced through looking through the machine learning calculations inside medical care on PubMed

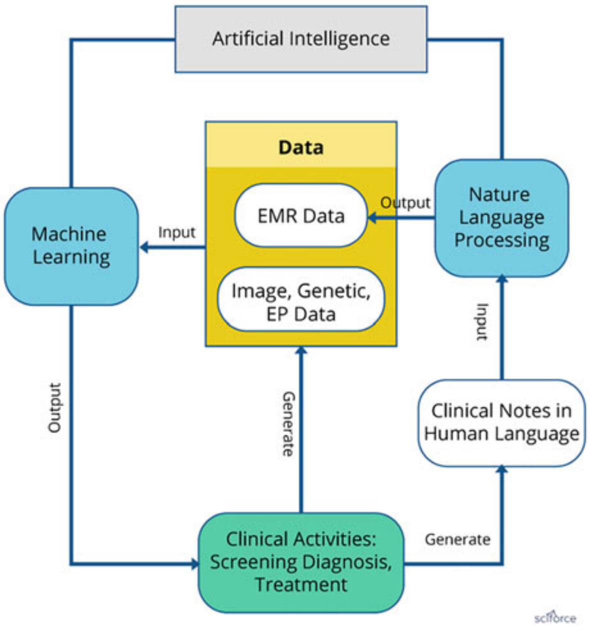


Fig. 5 Machine learning and natural language processing in healthcare

1. **SVM$_{(s)}$**

   SVMs are the most standard AI method that is being utilized by the medical services manufacturing. It utilizes a directed learning model for arrangement, relapse, and location of layouts. Of late, the calculation has been utilized to

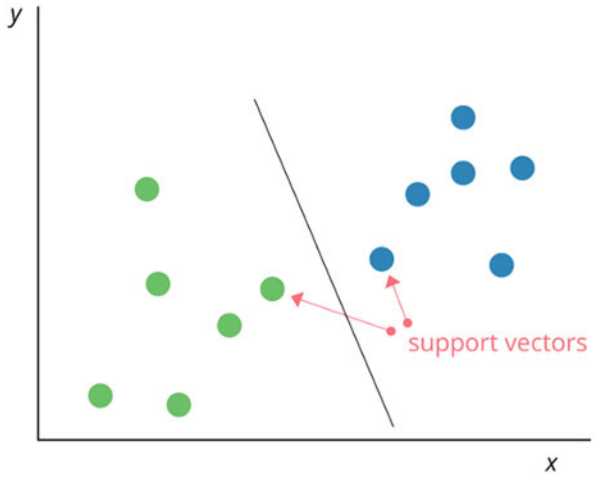


Fig. 6 Support vector machine

anticipate the drug adherence of heart patients that has helped millions to keep away from genuine outcomes, for example, medical clinic readmission and even demise. It is additionally being utilized for protein order, picture isolation, and text arrangement [7] (Fig. 6).

Support vector machines are utilized broadly in clinical exploration, for instance, to recognize imaging biomarkers, to analyze malignancy or neurological illnesses, and all in all for order of information from imbalanced datasets or datasets with missing qualities [11].

2. **Neural Networks**

   In NN, the relationships among the result and the info factors are portrayed through hidden layer blends of prespecified functionals. The objective is to gauge the loads through information and result information so that the normal error among the result and their expectations is limited. Here NN is referred as the neural network (Fig. 7).

NN is effectively applied to different regions of medication, for example, symptomatic frameworks, biochemical investigation, picture examination, and medication improvement, with the common case of bosom malignancy forecast from mammographic pictures [12].

3. **Logistic Regression**

   This ML method is utilized to foresee the current situation of the unmitigated ward variable using indicator factors. It is regularly utilized for characterizing and foreseeing the likelihood of an occasion, for example, infection hazard to the executives, which helps specialists in settling on basic clinical choices. It additionally enables clinical establishments to target patients with more danger

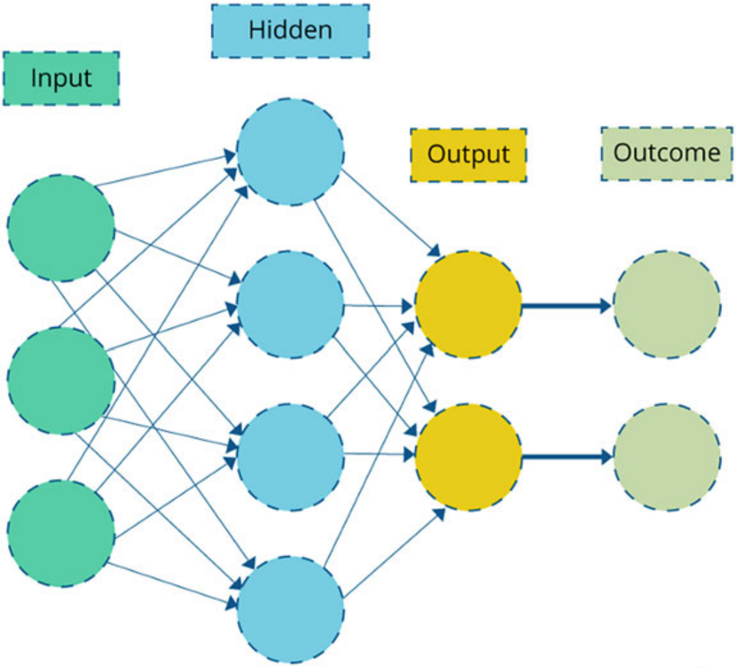


Fig. 7 Neural network

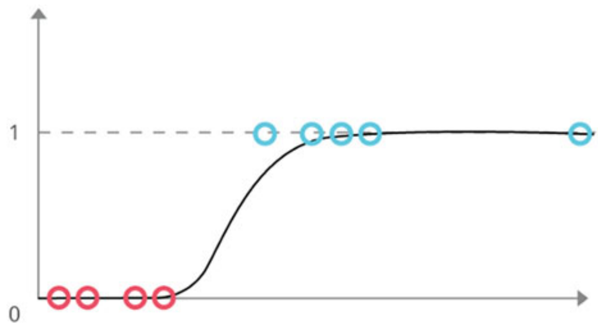


Fig. 8 Logistic regression

and curates behavioral wellbeing intends to improve their day-by-day wellbeing habits (Fig. 8).

In medical services, it is generally used to tackle order issues and to foresee the likelihood of a specific occasion, which makes it an important device for an infection hazard appraisal and for improving clinical choices [13].

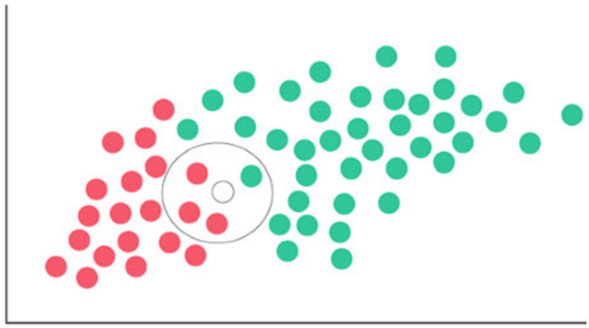


Fig. 9 Naïve bayes classifier

4. **Natural Language Processing**
   In medical services, an enormous extent of clinical data is as account text, for example, physical assessment, clinical research center reports, employable notes, and release synopses, which are unstructured and limitless for the PC program without uncommon strategies for text handling. It tends to these issues, as it distinguishes a progression of illness applicable watchwords in the clinical notes dependent on the verifiable information bases that after approval enter and enhance the organized information to support decision-making.
5. **Naïve Bayes**
   In view of the Bayes hypothesis, this is one of the most proficient AI methods ever known to humankind and is profoundly utilized by the medical care industry for clinical information explanation and infection expectation. With regard to information mining, characterization can be named as information investigation, which is frequently used to remove models portraying information classes. Since the likelihood of conveyance is high, Bayes classifier can accomplish an ideal outcome [14] (Fig. 9).

It stays as one of the best and effective arrangement calculations and has been effectively applied to numerous clinical issues, for example, in order of clinical reports and journal articles.

6. **Deep Learning**
   This is an expansion of the traditional neural network procedure, being, to lay it out plainly, an NN with numerous layers. Having more limits contrasted with traditional ML methods, deep Learning can investigate more perplexing non-straight examples in the information. Being a pipeline of modules every one of them is teachable, and deep learning speaks to a versatile methodology that, among others, can perform programmed highlight extraction from crude information.
   In the clinical applications, deep learning calculations effectively address both machine learning and natural language processing assignments. The regularly

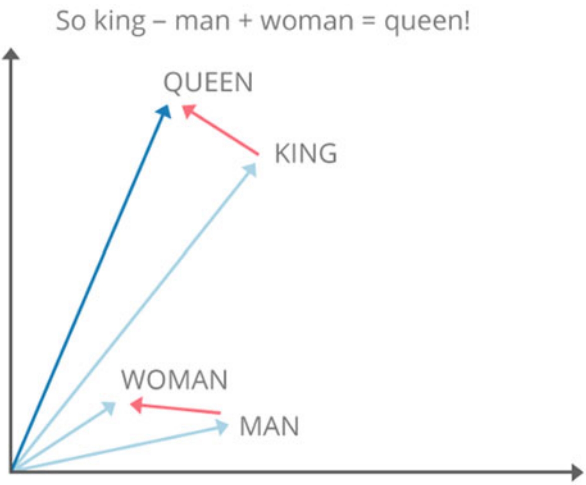


Fig. 10 Word vectors

utilized deep learning methods incorporate convolution neural network (CNN), repetitive neural organization, profound conviction network, and multilayer perceptron, with CNNs driving the race from 2016 on.

7. **Word Vectors**

   Viewed as an achievement in NLP, word vectors, or word2vec, is a gathering of related models that are utilized to create word embeddings. In their essence, word2vec models are shallow, two-layer neural networks that reproduce etymological settings of words. Word2vec produces a multidimensional vector space out of a book, with every extraordinary word having a comparing vector. Word vectors are situated in the vector space such that words that share settings are situated in nearness to one another [15] (Fig. 10).

Word vectors are utilized for biomedical language preparing, including closeness discovering, clinical terms normalization, and finding new parts of ailments.

8. **Recurrent Neural Network**

   The second in prevalence in medical services, RNNs symbolize to neural networks that utilize successive data. RNNs are called intermittent on the ground that they play out a similar errand for each component of a grouping, and the yield relies upon the past calculations. RNNs have a "memory" which catches data about what has been determined a few stages back (additional on this later) [16] (Fig. 11).

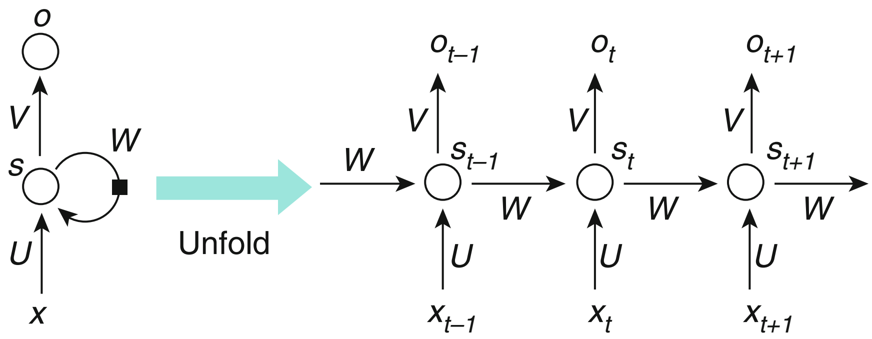


Fig. 11 Amazingly widespread in NLP, RNNs are likewise a potent technique for foreseeing clinical occasions

6 Future Possibilities in General Practice

The genuine medical services advantage later on will probably be the collaborations gotten through consolidating the intensity of AI-related advances over the whole patient excursion. While we cannot foresee the future with 100% exactness, we can consider future situations that are probably going to occur. Take, for example, a future situation of an overweight male patient who is a previous smoker with diabetes and atrial fibrillation. Later on, that patient is probably going to approach wearable gadgets to follow blood glucose levels, pulses, and rhythms, and exercise levels after some time. Those data might be synchronized up to a focal observing framework that utilizes ML to perceive unusual or undesired example changes. At the point when an anomalous example change is perceived, the observing framework can naturally advise the patient's supplier and educate the patient to plan an arrangement (or advise the patient to call the paramedics, or look for crisis care or earnest consideration should the distinguished example change be more basic).

When the patient shows up at the supplier's office, the patient can register at a voice-empowered biometric stand that uses NLP to acquire enlisted information. Accepting that the patient has no funds to pay, the patient does not have to pay at the booth or set aside the effort to see anybody at the front work area to examine the record balance. Preceding the supplier seeing the patient, the supplier can survey the applicable information sent over by the patient's wearable gadgets, alongside the reasons why the alarm was produced, and a rundown of potential conclusions created by the wearable observing framework's AI capacities. When the supplier sees and inspects the patient, the supplier at that point directs the visit note. After the note has been directed, NLP and ML transform the content in that directed note into systematized data in the background, utilize those arranged data to refresh the classified data in the patient's electronic clinical record, and consequently produce the fundamental charging data to send to the patient's insurance agency for payment. Likewise, in view of the data given by the wearable observing framework just as the electronic clinical record, a treatment arranging framework can utilize a mix of NLP and ML to prescribe changes in accordance with the patient's present treatment plan (counting medicine doses and frequencies just as feast and exercise schedules) to address the patient's particular clinical circumstance. During the visit, the patient notices feeling tired and having a hack for as far back as six to about 2 months, so the supplier orders lab tests alongside a chest CT check, given the patient's smoking history. Patient and supplier are told of the outcomes by means of the patient entrance and supplier gateway. The outcomes incorporate an anomalous chest CT with a mass at first distinguished by CAD as having a 75% likelihood of being cellular breakdown in the lungs, a finding that the deciphering radiologist concurs with. The patient is again provoked by a NLP-controlled chatbot to cause a catch-up meeting with the supplier, and afterward with the oncologist after a biopsy affirms the finding of cellular breakdown in the lungs, to examine the outcomes and subsequent stages.

Before meeting with the patient, the oncologist can audit the patient's electronic clinical record and imaging discoveries, and at that point use NLP and ML to survey proof-based clinical writing for potential treatment choices or clinical preliminaries that explicitly apply to the patient. When a treatment choice is picked, ML can be utilized to tailor the treatment plan to the patient's particular clinical condition.

Both while the patient is going through treatment and after treatment, wearable gadgets will screen the patient and again advise both the patient and supplier should the AI-empowered checking framework confirm that mediation is required. ML can then additionally propose the best strategy depending on the patient's clinical condition and prescriptive investigation, and screen persistent conduct so as to give impetuses to fitting conduct. At last, this current patient's information, alongside information from different patients, is de-distinguished and sent into a populace wellbeing data set. That information base, with new data being included day by day, is consistently being mined utilizing NLP and ML for possible relationship over a wide arrangement of information with an end goal to recognize already unfamiliar causes for and medicines of illnesses. This situation embodies huge numbers of possible advantages to the patient, supplier, payer, and drug industry of utilizing the intensity of AI over the patient excursion [17] (Fig. 12).

7 Challenges

Simulated intelligence accompanies various difficulties for the medical care industry, outstandingly:

- Issues of security and protection
- Potential absence of interoperability with different stages
- Need for an elevated requirement of execution in any event outflanking people
- Blindness to enthusiastic signs from patients

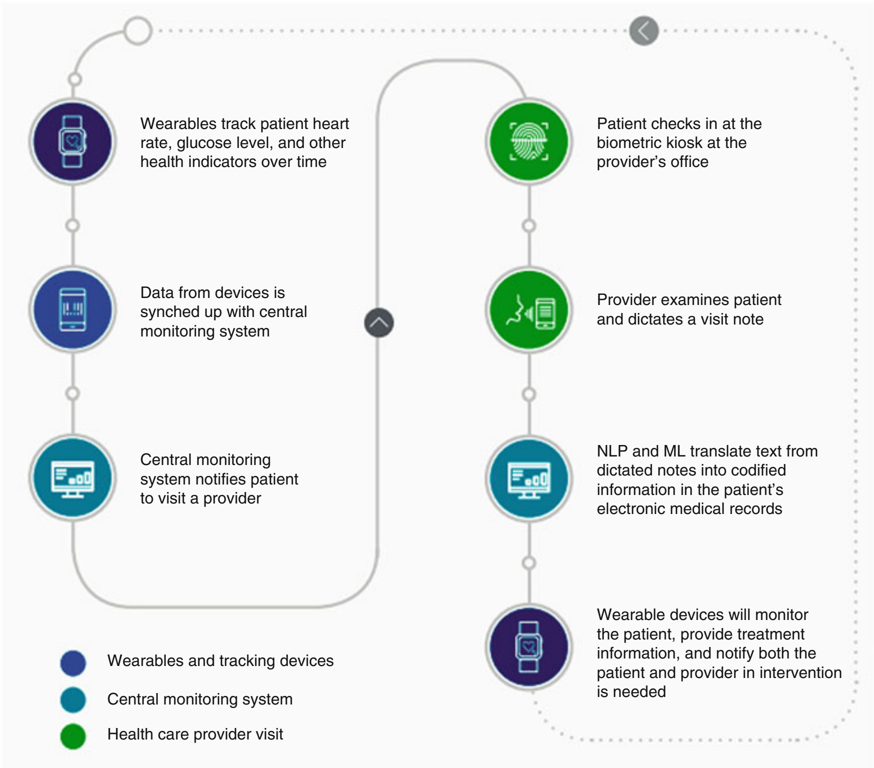


Fig. 12 Patient journey

1. Simulated intelligence requires a lot of preparing information. DeepMind's AlphaGo program, for instance, accomplished authority in the game Go by preparing 30 million recorded human moves. In medical services, as each new "game" conceivably speaks to a human life in danger, imaginative procedures for "preparing" should be found.
2. Fruitful selection of AI in medical care must be accomplished by tending to nerves felt by everyone and by medical care laborers. Consolation is required that AI will not encroach on human self-rule and that it will enlarge instead of supplant people.
3. The medical care industry has been delayed to receive advances, placing it in a feeble position with regard to sending AI. In the United States, one-fourth of medical clinics and over 40% of doctors are not utilizing electronic wellbeing record frameworks as indicated by McKinsey.
4. Without adequate amounts of excellent information in normalized designs, AI reception inside the medical services area will flounder. The test is in uniting the various, awkward vaults of electronic clinical records, lab and imaging frameworks, doctor notes, and medical coverage claims.

Risks/Threats

1. The progress of AI in medical services accompanies various dangers. While centered, AI might be ignorant concerning more extensive setting signals and it might likewise battle to manage the "inborn vulnerability" of medication in reality. Furthermore, dependence on computerization may at last decrease the abilities of doctors.
2. Five likely traps of using AI in medical care:
    - Lack of necessary moral principles over the entire division
    - Rushing the progress of AI, and error potential drawbacks
    - Insufficient preparation of clinical experts
    - Poor patient training and correspondence on advantages and hindrances
    - Unaffordable arrangements neglect to transform AI into the stethoscope of the twenty-first century
3. In spite of the fact that most customers trust medical service suppliers to make sure about their information, it is by and by being taken and information breaks are progressively observed as being unavoidable. Simulated intelligence can make new assault surfaces by sending information to outside suppliers and outer offices, implying that making sure about these channels is basic.
4. Self-governing specialists can act in sudden manners, with conceivably antagonistic results and in opposition to accepted practices (for instance, exhibiting clearly prejudicial conduct). In delicate regions, for example, medical services, it will be significant for professionals to have the option to get to a review trail and to consider whether computerized checking is required [18].

8 Conclusion

Current advances in AI consciousness present an energizing chance to improve medical care. Nonetheless, the interpretation of exploration methods to powerful clinical sending presents another wilderness for clinical and AI research. Strong, imminent clinical assessment will be fundamental to guarantee that AI frameworks are sheltered and powerful, utilizing clinically material execution measurements that go past proportions of specialized precision to incorporate how AI influences the nature of care, the fluctuation of medical service experts, the proficiency and profitability of clinical practice, and, in particular, understanding results. Free datasets that are illustrative of future objective populaces ought to be curated to empower the correlation of various calculations, while cautiously assessing for indications of likely inclination and fitting to unintended confounders. Engineers of AI instruments must be cognizant of the possible unintended results of their methods and guarantee that calculations are planned in view of the worldwide network. Further work to improve the interpretability of calculations and to comprehend human–method cooperation will be fundamental to their future selection and security upheld by the advancement of mindful administrative structures [9, 10].

References

1. Chen, R. Ranganath, in *A Review of Challenges and Opportunities in Machine Learning for Health* (2019) Corpus ID: 208857467
2. R.L. Richesson, W.E. Hammond, M. Nahm, D. Wixted, G.E. Simon, J.G. Robinson, et al., Electronic health records based phenotyping in next-generation clinical trials: A perspective from the NIH Health Care Systems Collaboratory. J. Am. Med. Inform. Assoc. **20**(e2), e226–e231 (2013)
3. E. Corbett, The real-world benefits of machine learning in healthcare. *Health Catal.* (2017, April 25)
4. T. Grote, P. Berens, On the ethics of algorithmic decision-making in healthcare. J. Med. Ethics **46**(3), 205 (2019 November). https://doi.org/10.1136/medethics-2019-105586
5. D. Faggella, in *Machine Learning Healthcare Applications – 2018 and Beyond* (Emerj The AI Research and Advisory Company, 2019, May 19)
6. B. Beaulieu-Jones, S.G. Finlayson, C. Chivers, Trends and focus of machine learning applications for health research. JAMA Netw. Open **2**(10), e1914051 (2019). https://doi.org/10.1001/jamanetworkopen.2019.14051
7. K. Wazny, Applications of crowdsourcing in health: An overview. J. Glob. Health **8**(1), 010502 (2018 Jun). https://doi.org/10.7189/jogh.08.010502 PMCID: PMC5840433
8. C. Réda, E. Kaufmann, A. Delahaye-Duriez, Machine learning applications in drug development. Comput. Struct. Biotechnol. J. **18**, 241–252 (2020). https://doi.org/10.1016/j.csbj.2019.12.006
9. F. Khan, in *Robotic Surgery and Machine Learning* (2019). https://medium.com/ai-techsystems/robotic-surgery-and-machine-learning-7f87824228ea
10. https://analyticsindiamag.com/top-6-ai-algorithms-in-healthcare/
11. P. Janardhanan, L. Heena, F. Sabika, Effectiveness of support vector machines in medical data mining. J. Commun. Softw. Syst. **11**(1), 25–30. Research Gate (2015 April). https://doi.org/10.24138/jcomss.v11i1.114
12. N. Shahid, T. Rappon, W. Berta, Applications of artificial neural networks in health care organizational decision-making: A scoping review. PLoS One **14**(2), e0212356. Research Gate (2019 February). https://doi.org/10.1371/journal.pone.0212356
13. A. Yadav, M. Ali, L. Hui, A. Yadav, Analysis of healthcare data of Nepal hospital using multinomial logistic regression model. Int. J. Manag. Inf. Technol. **11**(2):2718–2728. Research Gate (2016 April). https://doi.org/10.24297/ijmit.v11i2.4864
14. K. Sureskumar, in *Naive Bayesian Classification Approach in Health care Applications.* Research Gate. Project: Genetic algorithm based sensor node classifications in wireless body area networks (WBAN) (2017 November)
15. J.A.M. Sidey-Gibbons, C. Sidey-Gibbons, Machine learning in medicine: A practical introduction. BMC Med. Res. Methodol. **19**(1), 1. Research Gate (2019 March). https://doi.org/10.1186/s12874-019-0681-4
16. A. Haya, N. Radi, Á. MacDermott, in *Recurrent Neural Networks in Medical Data Analysis and Classifications.* Research Gate (2016 December). https://doi.org/10.1016/B978-0-12-803468-2.00007-2
17. A. Kaushal, K. Abrams, D. Sklar, B. Fera, in *The future of artificial intelligence in health care How AI will impact patients, clinicians, and the pharmaceutical industry* (Deloitte, n.d.)
18. S. Shinde, P.R. Rajeswari, Intelligent health risk prediction systems using machine learning: A review. Int. J. Eng. Technol. **7**(3), 1019–1023 (2018 June). https://doi.org/10.14419/ijet.v7i3.12654

Coronavirus Pandemic: A Review of a New-fangled Risk to Public Health

Sunita Sharma, Amit Kumar, Lokesh Chandra Gupta, S. K. Ghoshal, and Deepika Gaur

1 Introduction

The present pandemic of COVID-19 is due to the recurrence of the coronavirus (CoV), wherein a new-fangled virus was recognized and primarily entitled as nCoV. Its name was further changed to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS CoV-2) and the related ailments were termed as COVID-19 [1]. It originates from the Coronaviridae group and is a positive abandoned RNA virus. As the virus splurged globally (on March 11, 2020), it was legitimately labelled as the epidemic [2].

Way back in 1960, the foremost case of CoV was reported as cold. In 2001, around five hundred patients were diagnosed with a flu-identical syndrome. CoV strain via the polymerase chain reaction led to the infection of 17–18 patients. Until 2002, CoV was known to be a meek non-mortal virus. In 2003, several reports were available with the evidences of the dissemination of CoV to several nations. In 2003, numerous cases of SARS triggered by CoV with 1000 deaths were testified. It was declared as the black year for the microbiologists and subsequently dedicated efforts were made to understand CoV. It was concluded that the pathogenesis of the disease was CoV. Until then, totally 8096 patients were inveterate as infected with CoV and

S. Sharma (✉) · D. Gaur
Department of Applied Sciences, The NorthCap University, Gurugram, Haryana, India

A. Kumar
National Institute of Health and Family Welfare, Munirka, New Delhi, India

L. C. Gupta
District TB Officer, Chief Medical Office, Muzzaffarnagar, Uttar Pradesh, India

S. K. Ghoshal
Physics Department, Faculty of Science, Advanced Optical Materials Research Group & Laser Center, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_16

in 2004, it was affirmed as the 'state emergency'. In Hong Kong, about 50 cases of SARS were tested out of which 30 were infested by CoV. Eight years later, Saudi Arabia reported many infected patients and deaths [3–7] which was triggered by the Middle East Respiratory Syndrome CoV (MERS-COV). COVID-19 was the foremost acknowledged virus and isolated from the pneumonia patient in Wuhan (China) [5, 7–9].

CoVs are members of the Coronaviridae group and other Nidovirales wherein α-CoV, β-CoV, γ-CoV and δ-CoV are four classes of the subfamily. α-CoV and β-CoV infect merely animals; however, γ-CoV and δ-CoV contaminate birds. In animals, gastroenteritis is caused by α-coronaviruses and β-coronaviruses. SARS-CoV and MERS-CoV is responsible for the spartan breathing disorder in people. However, the supplementary coronaviruses persuade merely slight upper breathing ailments in immunocompetent masses, while the infants, kids and senior citizens are more vulnerable to these viruses [10].

2 Virion Structure

The shape of CoV virions is spherical with a diameter of approximately 125 nm [11, 12]. A main projecting feature of CoV was the weapon-contour point prognoses stemming on an exterior of the virion and named coronavirus due to its resemblance with the solar corona. CoVs have helically symmetrical nucleocapsid that lies inside the virion envelope. CoV encompasses 4 key essential proteins such as the spike (S), membrane (M), envelope (E) and nucleocapsid (N). The S protein is the strongly N-allied glycosylated. The unique spike assembly on the surface of the virus is due to the homotrimers-fixed S protein [13, 14] which facilitates the connection to the host receptor [15]. The S protein is sliced by a host cell into S1 and S2 wherein S1 builds the huge receptor-binding area, whereas S2 makes up the branch of the spike molecule. The M protein is considered responsible for providing the virion its outline due to its trifling size having 3 transmembrane regions. In the virion, the M protein possesses binary unlike conformations permitting it to endorse the curvature of the membrane and binding to the nucleocapsid. The E protein is present in lesser amounts in the virion which is enormously contrary, possessing a similar structural design. The E protein expedites association and discharge of the virus [16]. The N protein is found in the nucleocapsid. It has two different realms including the N-terminal domain (NTD) and C-terminal domain (CTD) which are the binder for the RNA in vitro, and each adopts diverse methods to bind RNA. The finest RNA binding occurs due to the offerings from both fields [17–19]. The N protein is profoundly phosphorylated [20] and recommends initiating a change in the structure [20, 21]. A genomic wrapping signal binds explicitly to the C-terminal RNA binding area [22]. The N protein impasses nsp^3 [18, 23], a key constituent of the replicas multipart and M protein [24]. The hemagglutinin-esterase (HE) is a subsection of β-CoV which fixes the sialic acids on the shallow glycoproteins and embraces the acetyl-esterase action [25]. This in turn enhances the S protein-facilitated cell entrance and virus blowout via the mucosa [26]. In addition, the HE boosts the Murine Hepatitis Virus [27] despite its contradictory designation in the tissue culture for unidentified explanations [28, 29].

3 Objective

The primary objective is to overview the available information regarding the virus responsible for COVID-19, its origin, epidemiology, spreading, diagnosis, prevention of the transmission, quarantine and treatment.

4 Methodology of Review

Revisiting the available information in the Google and PUBMED.

5 Epidemiology of COVID-19

An outburst of the pneumonia instigated by n-CoV arose in Wuhan (China) with the enduring threat of a pandemic in December 2019. The epidemic of SARS-CoV-2 became a Public Health Emergency of International Concern on 30 January, 2020. Certainly, SARS-CoV-2 possesses a sturdier transmission capability compared to the SARS-CoV. The rapid escalation in the inveterate cases makes the deterrence and control of COV-19 exceptionally solemn [30].

6 Total Coronavirus Cases

2019-nCoV was less severe, more transferrable, and different from both SARS-CoV and MERS-CoV; nonetheless, they were meticulously related. The illness inception in the people confirmed the rapid transmission of 2019-nCoV from person to person [31, 32] and the first mortality was reported on 11 January 2020.The substantial movement of the Chinese people during the Chinese New Year fired this endemic. Soon after, the cases in other provinces of China and different nations were identified. The spread to the healthcare workers who were treating the patients was first reported on 20 January 2020. The lockdown was imposed in Wuhan till 23 January 2020 and extended to other metropolises. The cases of COVID-19 were reported in people who did not have any travel history to China, thereby confirming the native transmission among people [1, 33]. To prevent further transmission and spread of this lethal virus, the airports introduced the screening, isolation (14 days) and testing mechanisms for detecting the symptomatic persons travelling in China. Shortly, it was found that the asymptomatic people are also responsible for the transmission of this disease. Thus, all the people returning from China were kept under isolation for 14 days and tested to confirm for the symptoms of COVID-19. Despite all efforts, the confirmed cases of Coronavirus continued to have an exponential growth and the cases kept doubling in every 1.8 days [34]. As of 20 August 2020, over 22.5 million infected cases of COVID-19 with nearly 791 thousand deaths have been affirmed globally. India has reported over 2.8 million cases of confirmed COVID-19 and about 54,000 deaths till date (third place in the world after USA and Brazil) wherein all the contact persons of the CoV cases were put under quarantine. However, these data may not give the exact numbers of the cases owing to the restrictions of scrutiny and testing. Although the SARS-CoV-2 originated from the bats, the intermediate animal through which it got transmitted to the humans still remains ambiguous [35].

7 Transmission

Researches revealed that the population of all ages is susceptible to the Coronavirus infections. Mainly the big droplets produced due to coughing and sneezing by the symptomatic people are responsible for the spread of this infection. However, it can also spread through the asymptomatic people and prior to the start of the symptoms. The virus-related loads are more in the nasal cavity than in the oesophagus. The patients infected from Coronavirus are infectious to others not only for the entire duration of the symptoms but even during the process of the clinical recovery. The infected droplets can spread easily up to 1–2 m and if these gets deposited on the surface then they remain infectious there for several days under suitable ambience. However, these germs can be destroyed within 60 s using common disinfectants. In addition, the infection can also be spread if the virus enters the body through nose, mouth or eyes. It exists in the stool and adulteration of water wherein the growth period differs from 2 to 14 days. Various model studies showed that the expected reproduction rate of this virus can range from 2 to 6.47 [36, 37].

The common clinical features of COV-19 are fever, cough, sore throat, headache, tiredness, myalgia, and breathlessness. Up to 7 days, the disease may lead to pneumonia, breathing failure, and demise which is caused due to the increase in the provocative cytokines including interleukin. The complications may comprise severe tremor, lung and kidney injuries where the reclamation of the patient starts in the second or third week. The high-risk population is the senior citizens and those with low immunity. The fatality rate of the infected persons admitted in the hospital ranges from 4% to 11% [35].

8 Diagnosis

The incubation period for SARS-CoV, MERS-CoV and COVID-19 is 4–10 days, 5–6 days and 3–7 days, respectively [38]. The corresponding death period for the above mentioned diseases is reported to be 20–25 days, 11–13 days and 3–7 days, respectively. The symptoms of COVID-19 are high temperature, cough, dyspnoea, muscle ache, confusion, headache, sore throat, rhinorrhoea, chest pain, diarrhoea, biliousness, vomiting, anosmia and dyspepsia. In the self-limited infection, it is not necessary to diagnose the CoV because it can go away on its own. However, it is essential in the epidemiological studies to detect an etiological mediator and the diagnosis becomes vital in the red zones (places with more COVID-19 patients). The detection of the confirmed cases aids to control the epidemics. Real-time polymerase chain reaction (RT-PCR) technique has been established for the analysis of CoV in human. The multiplex real-time RT-PCR assays can analyse all CoVs [39, 40]. In addition, serologic assays can detect the RNA that is difficult to isolate for epidemiological studies [29, 30]. CoVs can enter into the cell either via the conformist endosomal paths of entry or through the non-endosomal passage or via both. The interferon-inducible transmembrane proteins (IFITM) type of action remains unexplained wherein the cell-to-cell blend assays proposed that it wedges the enclosed virus entry by stopping the synthesis of the viral cover with the plasma membrane, thereby modifying the host membrane fluidity.

9 Prevention of Transmission

So far, no antiviral therapeutics has been developed that precisely mark human coronaviruses and hence the treatments are merely helpful. The interferons (IFNs) are somewhat effective in in vitro in contrast to CoV. The IFNs in combination with the ribavirin have improved action in vitro than the IFNs alone when dealing with a few coronaviruses. The efficiency of this blend in vivo needs additional investigation. The SARS and MERS epidemics have encouraged conducting researches on such viruses, thereby recognizing diverse and apt anti-viral targets. Intensive researches must be carried out for the drugs development targeting such mechanisms, enabling to constrain the viral duplication [41, 42].

10 Quarantine

The International Health Regulations (IHR) is a lawful mechanism that binds all countries on the globe and the Member States of the World Health Organisation (WHO). Its role is to aid the global public to inhibit and answer to the acute public fitness hazards that can be transmitted across the borders and threating to the people.

In harmony with the principles delineated in IHR, the travel advisories were issued on a regular basis during the surge of COVID-19 cases in China. According to the travel advisory, Indian travellers were instructed to desist from travelling to China. In fact, the current visas (even eVisa previously dispensed) were now invalid for any overseas national itinerant from China. The passengers who had travelled to China have to be quarantined on return. Although the intermediate- and protracted-term influence of the travel ban persists to be observed, the model studies recommend that for the short span of time. Thus, it was improbable to have a significant influence on the comprehensive spread of SARS-CoV-2 except that unremitting 90% transportable constraints were employed along with extra 50% decline in native spread. Such prohibitions could solitarily build a powerful armour in order to curtail the current outburst.

China enforced a lockdown on 23 January 2020 in Wuhan to quarantine the intense public health surveillance system combined with swift analytical tests and quarantine whenever essential. In order to control the international epidemics, the teamwork of the government, medical staff and common people remains critical. As an alternative to the forced top-down quarantine approaches, the self-isolation and self-inspection appear an added ecological and implementable strategy in a prolonged epidemic like COVID-19 [43].

11 Treatment

At present, merely few options are available for the prevention of this infection. The vaccines have got only approval, but they cannot be always used because of their inadequate effectiveness. It has been reported that in some cases they were connected in the assortment of the new pathogenic CoVs through the recombination of the mingling worries. Although many potential vaccines have been developed, none of them is approved for SARS-CoV. Therapeutic SARS-CoV counterbalancing antibodies have been produced and could be salvaged. The healthcare providers need to be protected using such antibodies. The development of the vaccine for coronaviruses faced several encounters [44]. The vaccines should either provide better immunity in comparison to the novel virus or it must reduce the disease experienced in the course of a secondary infection. The viruses may become adaptable to the vaccine. Moreover, several approaches have been established for the progress of the vaccine to diminish the possibility of the recombination.

Worldwide, approximately 40 teams have been working for developing the injections against CoV infection. The effort is principally encouraged by the extraordinary achievement and quick growth of the Ebola virus vaccines. There is a possibility that COV injections can be based on the viral RNA. The merits of the RNA and DNA vaccines are that they can quickly be developed with high safety; however, no vaccine based on RNA or therapeutic has got regulatory approval. The research on SARS vaccine for COVID-19 is in progression in The Institute Pasteur in France. Despite intensive studies on coronaviruses during last 20 years predominantly relating to the SARS-CoV-1 epidemic, no vaccine is accessible for either SARS-CoV-1 or MERS so far. Nevertheless, few probable vaccines have lately been invented for phase I clinical trials [29].

12 Application of Nanotechnology to Combat Against Coronavirus

The nanotechnology and nanomedicine have outstanding prospects when dealing with various serious and challenging health problems. For the treatment or disinfection of viruses, nanobiotechnology plays a vital role.

Diagnosis and Treatment of CoV by Nanomaterials

To detect diverse coronaviruses, silver nanoparticles (Ag NPs), MoS_2 nanosheets, Zirconia NPs (Zr NPs) and gold NPs (Au NPs) have been exploited [45–48]. Meanwhile, the coupling of the nanomaterials with the colorimetric sensing, electrochemiluminescence, immunesensing, photoluminescence and chiroimmunosensing techniques became potential for the coronaviruses detection. In near future, the electrochemical devices will also be used for the detection of the new kind of coronaviruses due to their good capacity to couple with the nanomaterials. The use of nanomaterials in these techniques imparts high sensitiveness and leads to quick analyses. To determine the feasibility of the nanomaterials' inclusion in the development of highly effective vaccine for COVID19, materials including the Au, Ag, silver sulphide, TiO_2 (titania), zirconium, graphene and some biopolymers have been used for diagnostic and therapeutic purposes. The vaccines based on the nanoparticles have shown strong latent to persuade advanced defensive immunity retort compared to conformist antigen-based injections. For the rapid detection of the viral infection at an early stage, nanoassays provide superior specificity and sensitivity compared to the existing state-of-the-art techniques.

Nanomaterials for Facemasks Production

The saggy and disposable device called facemask is used to cover the mouth and nose of a person to protect against the potential contaminants in the surroundings. The respirators have been designed to shield the person from gasping hazardous chemicals and contagious particles wherein these respirators minimize the respiratory exposure to aerial toxins. A facemask comprises a filter made up of nanofibres wherein these nanofibres possess very high surface area per unit mass, leading to an increase in the filter performance by capturing the naturally occurring nanoparticles like viruses and bacteria. Single-fibre filtration theory can be employed to explain the filtration performance of these filters for the particles with a size of 100 nm. The most penetrating particle sizes range from 30 to 100 nm and may vary depending on the test conditions. However, no indication for thermal recoil effects is found for particles with 4 nm in size. In-depth studies are required to measure the total inward leakage for the respirators protecting against nanoparticles [49].

Nanomaterials for Disinfectants

Generally, the patients are at high risk of infections in the hospitals. Several studies have been carried out to develop medications to restrain the spread of germs from various medical equipment surfaces to the environment of the hospitals. The nanotechnology emerged as an outstanding weapon to fight against such infections. Antimicrobial nanomaterials can regulate the infections and destroy bacteria. Nanoparticles are mostly nontoxic compared to other antibiotics and detergents. On top, these nanomaterials are steady and can be produced using simple methods. So far, such a disinfectant that possesses good effectiveness for a wide variety of pathogens has been deficient wherein each pathogen needs the most suitable disinfectant. The suitability of the disinfectant depends on many issues as well as concentrations, time of activity and the types of the surfaces and microbial. In fact, an improper conservation method may lead to contamination of many detergents, causing the stiffening of polymers and enduring corrosion of the treating materials. In this regard, nanomaterials enable in overpowering all such shortcomings. In the recent past, more research was directed on the expounded antimicrobial coatings where the graphene-based nanomaterials revealed outstanding bactericidal activity. The common surfaces in contact with the patients need effective disinfection [50–52] and thus the deposited thin films of graphene-based nanomaterials on top of such surfaces can be greatly useful. A Pune-based startup (Weinnovate Biosolutions in India) has developed a non-alcoholic aqueous-based colloidal silver solution (using NanoAg) as the hand sanitizer and disinfectant. This solution is non-combustible and devoid of harmful chemicals. Therefore, it can emerge as a potential sanitizer for the prevention of infection transmission.

Anti-COVID-19 Nanocoating

The polymers enclosed with copper nanoparticles (Cu NPs) are painted or sprayed as coating on the surfaces. After the deposition of the coating, the nanoparticles pledge the release of metal ions onto the surface. The Cu ions are formed when electrons were stripped from the atoms of a particular element. The Cu atoms became electrically neutral when the negatively charged electrons balance the positively charged protons. The ions have a reduced negative charge and thus carry an effective positive charge. The Cu ions display an effective antiviral activity against the influenza, herpes simplex and vaccinia virus, eliminating the viral particles that can cling to the surfaces. The ability of the nanoparticles to slowly release the ions indeed provides the solution for the long-lasting protection against COVID-19, other viruses and bacteria [53]. Thus, the coating can actually be effective in reducing the infectivity of the viruses tenfold for weeks, or even months. Such coatings are free from toxic or heavy metals, and thus safe for humans. In short, it may enable the users to minimize the use of harmful chemicals as cleansing agents.

13 Conclusions

Unquestionably, many lessons can be learnt since the international retort to the SARS-COV-2 danger. Maximum of such retorts appear volatile, with minute preparation for the venture to health schemes, involvement of the public and empowerment. The primary intension of the epidemic, intermediator, its treatment, early diagnosis of the asymptomatic patients all remains elusive. Medical judgements have commenced to recognize the injections and active treatment routines. However, exertions to detect medicine at later stage of COVID19 need to be emphasized more. The role of nanotechnology in different facets towards the inhibition and management of COV is remarkable. This infectious disease threat at present time is long lasting. In order to have lesser scales of harm to human life and economy, there is an urgent necessity for investments in setting up people-centric health systems.

Acknowledgement This work has been reported under Statement of Co-operation between Department of Applied Sciences, The NorthCap University, India, and Universiti Teknologi Malaysia, Malaysia.

References

1. Centers for Disease Control and Prevention: Coronavirus disease 2019 (COVID-19) – Situation summary (2020, March 11). Retrieved from: https://www.cdc.gov/coronavirus/2019-ncov/summary.html
2. K. Ramphul, S.G. Mejias, Coronavirus disease: A review of a new threat to public health. Cureus **12**(3), e7276 (2020) Retrieved from: https://doi.org/10.7759/cureus.7276
3. Centers for disease control and prevention (CDC), Outbreak of severe acute respiratory syndrome—Worldwide. MMWR Morb. Mortal. Wkly Rep. **52**(12), 241–246 (2003)
4. World Health Organization, *Coronavirus Never before Seen Inhumans Was the Cause of SARS–Update 31* (The Organization, Geneva, 2003)
5. World Health Organization, in *Summary of Probable SARScases with Onset of Illness from 1 November 2002 to 31 July 2003*. (14 February 2020a).Retrieved from: http://www.who.int/csr/sars/country/table2004_04_21/en/index.html

6. J.S. Peirwas, S.T. Lai, L.L. Poon, Y. Guan, L.Y. Yam, W. Lim, W. Lim, J. Nicholls, W.K.S. Yee, W.W. Yan, M.T. Cheung, V.C.C. Cheng, K.H. Chan, D.N.C. Tsang, R.W.H. Yung, T.K. Ng, K.Y. Yuen, Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet **361**, 1319–1325 (2003)
7. World Health Organization, in *WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China Geneva (9 January 2020 and 14 January 2020)* (2020b). Retrieved from: https://www.who.int/china/news/detail/09-01-2020-whostatement-regarding-cluster-of-pneumoniacases-in-wuhanchina
8. N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China. New Engl. J. Med. **382**(8), 727–733 (2020)
9. D. Kumar, R. Malviya, P.K. Sharma, Corona virus: A review of COVID-19. Eur. J. Med. Oncol. **4**(1), 8–25 (2020)
10. C. Jie, L. Fang, L.S. Zheng, Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. **17**, 181–192 (2019)
11. M. Bárcena, G.T. Oostergetel, W. Bartelink, F.G.A. Faas, A. Verkleij, P.J.M. Rottier, A.J. Koster, B.J. Bosch, Cryo-electron tomography of mouse hepatitwas virus: Insights into the structure of the coronavirion. Proc. Natl. Acad. Sci. U. S. A. **106**, 582–587 (2009)
12. B.W. Neuman, B.D. Adair, C. Yoshioka, J.D. Quispe, G. Orca, P. Kuhn, R.A. Milligan, M. Yeager, M.J. Buchmeier, Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. J. Virol. **80**, 7918–7928 (2006)
13. D.R. Beniac, A. Andonov, E. Grudeski, T.F. Booth, Architecture of the SARS coronavirus prefusion spike. Nat. Struct. Mol. Biol. **13**, 751–752 (2006)
14. B. Delmas, H. Laude, Assembly of coronavirus spike protein into trimers and its role in epitope expression. J. Virol. **64**, 5367–5375 (1990)
15. A.R. Collins, R.L. Knobler, H. Powell, M.J. Buchmeier, Monoclonal antibodies to murine hepatitwas virus-4 (strain JHM) define the viral glycoprotein responsible for attachment and cell–cell fusion. Virology **119**, 358–371 (1982)
16. L. Jose, N. Torres, L. Marta, B.C.V. DeDiego, J.M. Jimenez, A. Jose, R.N. Raul, F. Delgado, R.C. Castaño, C. Alcaraz, J. Torres, V.M. Aguilella, L. Enjuanes, Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogeneswas. PLoS Pathog. **10**, 1–19 (2014) e1004077
17. C. Chang, S.C. Sue, T.H. Yu, C.M. Hsieh, C.K. Tsai, Y.C. Chiang, S.J. Lee, H.H. Hsiao, W.J. Wu, W.L. Chang, C.H. Lin, T.H. Huang, Modular organization of SARS coronavirus nucleocapsid protein. J. Biomed. Sci. **13**, 59–72 (2006)
18. K.R. Hurst, C.A. Koetzner, P.S. Masters, Identification of in vivo-interacting domains of the murine coronavirus nucleocapsid protein. J. Virol. **83**, 7221–7234 (2009)
19. S.A. Stohlman, M.M. Lai, Phosphoproteins of murine hepatitwas viruses. J. Virol. **32**, 672–675 (1979)
20. S.A. Stohlman, R.S. Baric, G.N. Nelson, L.H. Soe, L.M. Welter, R.J. Deans, Specific interaction between coronavirus leader RNA and nucleocapsid protein. J. Virol. **62**, 4288–4295 (1979)
21. R. Molenkamp, W.J. Spaan, Identification of a specific interaction between the coronavirus mouse hepatitwas virus A59 nucleocapsid protein and packaging signal. Virology **239**, 78–86 (1997)
22. L. Kuo, P.S. Masters, Functional analyswas of the murine coronavirus genomic RNA packaging signal. J. Virol. **87**, 5182–5192 (2013)
23. K.R. Hurst, C.A. Koetzner, P.S. Masters, Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. J. Virol. **87**, 9159–9172 (2013)
24. L.S. Sturman, K.V. Holmes, J. Behnke, Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. J. Virol. **33**, 449–462 (1980)

25. A. Klausegger, B. Strobl, G. Regl, A. Kaser, W. Luytjes, R. Vlasak, Identification of a coronavirus hemagglutinin-esterase with a substrate specificity different from those of influenza C virus and bovine coronavirus. J. Virol. **73**, 3737–3743 (1999)
26. L.A. Cornelwassen, C.M. Wierda, F.J. Van der Meer, A.A. Herrewegh, M.C. Horzinek, H.F. Egberink, R.J. de Groot, Hemagglutinin-esterase, a novel structural protein of torovirus. J. Virol. **71**, 5277–5286 (1997)
27. L. Kazi, A. Lwassenberg, R. Watson, R. Groot, S.R. Weiss, Expression of hemagglutinin esterase protein from recombinant mouse hepatitis virus enhances neurovirulence. J. Virol. **79**, 15064–15073 (2005)
28. A. Lwassenberg, M.M. Vrolijk, A.L. Van Vliet, M.A. Langereis, J.D.F. de Groot-Mijnes, P.J.M. Rottier, R.J. de Groot, Luxury at a cost? Recombinant mouse hepatitIs viruses expressing the accessory hemagglutinin esterase protein dwasplay reduced fitness in vitro. J. Virol. **79**, 15054–15063 (2005)
29. R.F. Anthony, P. Stanley, Coronaviruses: An overview of their replication and Pathogeneswas. Methods Mol. Biol. **1282**, 1–23 (2015)
30. Y.Y. Zheng, Y. Tong, J.Y. Zhang, X. Xie, COVID-19 and the cardiovascular system nature reviews. Nat. Rev. Cardiol. **17**, 259–260 (2020)
31. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet **395**, 497–506 (2020)
32. Coronavirus disease (COVID-19) pandemic (21 March 2020). Retrieved from: https://www.who.int/emergencies/mers-cov/en/
33. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China (2020). Retrieved from: https://www.ourphn.org.au/wp-content/uploads/20200225-Article-COVID-19.pdf
34. C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, T. Zimmer, V. Thiel, J. Christian, W. Guggemos, M. Seilmaier, C. Drosten, P. Vollmar, K. Zwirglmaier, S. Zange, R. Wölfel, M. Hoelscher, Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N. Engl. J. Med. **382**, 970–971 (2020)
35. N. Jain, A. Choudhury, J. Sharma, V. Kumar, D. De, R. Tiwari, A review of novel coronavirus infection (coronavirus dwasease-19). Glob. J. Transfus. Med. **5**, 22–26 (2020)
36. M.L. DeDiego, E. Alvwerez, F. Almazan, M.T. Rejas, E. Lamirande, A. Roberts, W.J. Shieh, S.R. Zaki, K. Subbarao, L. Enjuanes, A severe acute respiratory syndrome coronavirus that lacks the E gene was attenuated in vitro and in vivo. J. Virol. **81**, 1701–1713 (2007)
37. J.L. Nieto-Torres, M.L. Dediego, C. Verdia-Baguena, J.M. Jimenez-Guardeño, J.A. Regla-Nava, R. Fernandez-Delgado, C. Castaño-Rodriguez, A. Alcaraz, J. Torres, V.M. Aguilella, L. Enjuanes, Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogeneswas. PLoS Pathog. **10**, e1004077 (2014)
38. A.O. Docea, A. Tsatsakwas, D. Albulescu, O. Crwastea, O. Zlatian, M. Vinceti, S.A. Moschos, D. Tsoukalas, M. Goumenou, N. Drakoulwas, J.M. Dumanov, V.A. Tutelyan, G.G. Onwaschenko, M. Aschner, D.A. Spandidos, D. Calina, A new threat from an old enemy: Re-emergence of coronavirus (review). Int. J. Mol. Med. **45**, 1631–1643 (2020)
39. S.L. Emery, D.D. Erdman, M.D. Bowen, B.R. Newton, J.M. Winchell, R.F. Meyer, S. Tong, B.T. Cook, B.P. Holloway, K.A. McCaustland, P.A. Rota, B. Bankamp, L.E. Lowe, T.G. Ksiazek, W.J. Bellini, L.J. Anderson, Real-time reverse transcription-polymerase chain reaction assay for SARS-associated coronavirus. Emerg. Infect. Dis. **10**, 311–316 (2004)
40. E.R. Gaunt, A. Hardie, E.C. Claas, P. Simmonds, K.E. Templeton, Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. J. Clin. Microbiol. **48**, 2940–2947 (2010)
41. J. Cinatl, B. Morgenstern, G. Bauer, P. Chandra, H. Rabenau, H.W. Doerr, Treatment of SARS with human interferons. Lancet **362**, 293–294 (2003)

42. L.J. Stockman, R. Bellamy, P. Garner, (2006). SARS: Systematic review of treatment effects. PLoS Med. **3**, e343 (2006)
43. P. Chatterjee, N. Nagi, A. Agarwal, B. Das, S. Banerjee, S. Sarkar, N. Gupta, R.R. Gangakhedkar, The 2019 novel coronavirus disease (COVID-19) pandemic: A review of the current evidence. Indian J. Med. Res. **151**, 147–159 (2020)
44. L.J. Saif, Animal coronavirus vaccines: Lessons for SARS. Dev. Biol. (Basel) **119**, 129–140 (2004)
45. X. Huang, M. Li, Y. Xu, et al., Novel gold nanorod-based HR1 peptide inhibitor for Middle East respiratory syndrome coronavirus. ACS Appl. Mater. Interfaces **11**(22), 19799–19807 (2019)
46. A. Loczechin, K. Seron, A. Barras, E. Giovanelli, S. Belouzard, Y.T. Chen, N. Metzler-Nolte, R. Boukherroub, J. Dubuisson, S. Szunerits, Functional carbon quantum dots as medical countermeasures to human coronavirus. ACS Appl. Mater. Interfaces **11**(46), 42964–42974 (2019)
47. T. Kato, Y. Takami, D.V. Kumar, E.Y. Park, Preparation of virus-like particle mimetic nanovesicles displaying the S protein of Middle East respiratory syndrome coronavirus using insect cells. J. Biotechnol. **306**, 177–184 (2019)
48. G. Nikaeen, S. Abbaszadeh, S. Yousefinejad, Label-free, electrical detection of the SARS, application of nanomaterials in treatment, anti-infection and detection of coronaviruses. Nanomedicine (London) **15**(15), 501–1512 (2020)
49. C. Akduman, E.P. Akçakoca, Kumbasar, Nanofibers in face masks and respirators to provide better protection. IOP Conf. Ser. Mater. Sci. Eng. **460**, 012013 (2018)
50. O. Akhavan, E. Ghaderi, Toxicity of graphene and graphene oxide nanowalls against bacteria. ACS Nano **4**(10), 5731–5736 (2010)
51. M. Saccucci, E. Bruni, D. Uccelletti, A. Bregnocchi, M.S. Sarto, M. Bossù, G.D. Carlo, A. Polimeni, Surface disinfections: Present and future. J. Nanomater. **2018**, 1–9 (2018) Article ID 8950143
52. S. Liu, T.H. Zeng, M. Hofmann, E. Burcombe, J. Wei, R. Jiang, J. Kong, Y. Chen, Antibacterial activity of graphite, graphite oxide, graphene oxide, and reduced graphene oxide: Membrane and oxidative stress. ACS Nano **5**(9), 6971–6980 (2011)
53. C. Poggio, M. Colombo, C.R. Arciola, T. Greggi, A. Scribante, A. Dagna, Copper-alloy surfaces and cleaning regimens against the spread of SARS-CoV-2 in dentistry and orthopedics: From fomites to anti-infective nanocoatings. Materials **13**(3244), 1–12 (2020)

Impact of COVID-19 Pandemic on Obese and Asthma Patients: A Systematic Review

Nenavath Srinivas Naik and Cheruku Ramalingaswamy

1 Introduction

COVID-19 was declared as a "Global Pandemic" by WHO (World Health Organization) on 11 March 2020 [1]. The COVID-19 pandemic challenges the health systems across the world to an unprecedented stage. The SARS-CoV-2 virus first emerged in the Wuhan city of China from December 2019 and has spread throughout the world with the initial model of transmission to humans as shown in Fig. 1 [2]. It is vulnerable to severely comorbid patients causing higher mortality rates. Various comorbidities are found with COVID-19 infection. Some of the comorbidities include obesity, diabetes, asthma, cardiovascular diseases, migraine, and throat infections. About 20% of the positive cases do not record any symptoms [3]. The most sensitive patients are the older age groups, children, and adults with inadequate health conditions.

There is a serious concern about whether the prevailing treatments for asthma and obesity may worsen the immunology of the COVID-19 patient. COVID-19 generally results from respiratory infections and other respiratory diseases. Its ill effects and severity are being studied by scientists for providing prevention and control. Patients with long-term respiratory diseases seem to be more vulnerable to COVID-19 infection [5, 6]. For the progression of COVID-19, old age, obesity, chronic cardiac disease, and high blood pressure are some of the reported symptoms

N. S. Naik (✉)
Department of Computer Science and Engineering, International Institute of Information Technology (IIIT), Naya Raipur, India
e-mail: srinu@iiitnr.edu.in

C. Ramalingaswamy
Department of Computer Science and Engineering, National Institute of Technology, Warangal, Telangana, India
e-mail: rmlswamy@nitw.ac.in

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_17

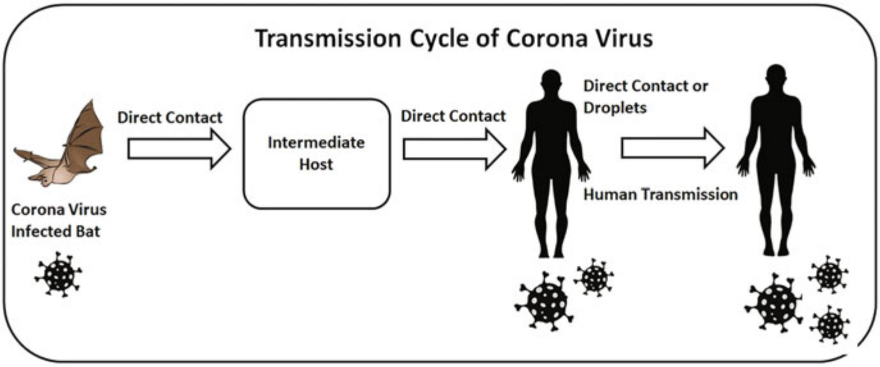


Fig. 1 Mode of transmission to humans

for identifying the risk. However, the licensed biologics for asthma treatment may target add-on experiments to maintain and reduce steroid uses. This study can help in determining the present status of the correlation of asthma and obesity with other commodities for the prevalence of contagious COVID-19. Evidence from several works of literature [9] supports that air pollution, lung inflammations, asthma, and obesity are predisposing factors for the progression of coronavirus infections.

2 Conjunction of COVID-19 and Asthma

Asthma is a respiratory viral disease with acute exacerbations and it needs hospitalizations. Asthma patients are more alert and conscious in fear of having high risk from the pandemic. According to the Global Asthma Report [10] 2018, about 1150 deaths have occurred due to asthma, and no measures were taken to address this disease before the COVID-19 outbreak. It is a state from which an individual's airway becomes narrowly swollen and produce. Asthma might be less or it could restrict day-to-day tasks. Sometimes, it might result in a life-threatening stroke.

Asthma can lead to chest pain, difficulty breathing, and cough. The symptoms can appear severe. Infection can usually be handled with saving inhalers to deal with ailments (salbutamol) and control inhalers (steroids) which avert symptoms. Severe cases might call for longer acting inhalers that help keep the airways open (e.g., formoterol, salmeterol, and tiotropium), in addition to inhalant steroids. Asthma can be a minor nuisance for some people. It is sometimes a problem that could result in a life-threatening asthma strike and disrupts our daily tasks. Infection cannot be cured, but its symptoms may be manipulated. Asthma patients have to consult a physician to track symptoms and signs, as asthma usually affects time.

Effects of COVID-19 in Asthma Patients

Asthma patients are continuing with the exacerbations in this COVID-19 epidemic. Studies say that the inhalation of corticosteroids to treat asthma attacks reduces the ability of the SARS-CoV-2 virus to some extent. It is also reported that people with asthmatic inflammation are not at high risk from the epidemic COVID-19 [21]. As steroids degrade the immunity of the patients and can worsen the cause, it is not clarified whether the steroids have a positive or negative impact on the COVID-19. Asthma patients are more inflammatory to other severe asthma inflammation – eosinophilic asthma. This raises the level of white blood cells causing nasal, sinus, respiratory, and airway inflammation which lead to more severe risk of COVID-19. Additionally, for the ingress of SARS-CoV-2 virus, various enzymes present in the lung cells are beneficial for respiratory viruses to expand. However, the role of enzymes in aiding the ability of these viruses to infect asthma patients is unclear [17].

The coronavirus 2019 outbreak (COVID-19) is frightful for all individuals, however, for all those with asthma, there is a tremendous fear that they will have severe side effects or may get SARS-CoV-2 disease. It is important to be aware that there is no evidence of an escalation in the range of infections in people with asthma. Although the Centers for Disease Control and Prevention (CDC) [10] says that patients having asthma may be at greater risk for serious illnesses, no published statistical data aid this decision currently. There has been a report indicating that asthma might increase the threat of hospitalization due to COVID-19 for adults aged 18–49 years. Yet, this is contingent upon a few of the patients. In contrast, data from New York by which asthma had been expressed were lower (more protective) in those exterminated by COVID-19. It is important to stay informed about the epidemic that is emerging, as advice could change the situation in the future.

Considering the changing ideas on using steroids and also on COVID-19, the majority are planning about exactly what to do when their prescription can be just actually a steroid (inhaled or oral). The data show that the SARS-CoV-2 might raise hails from damaging patients using steroids for viral disease [16]. Using steroids for curing different disorders (such as asthma) had not been considered. Individuals with asthma have been put to track their asthma. From the existing outbreak, the very best thing that someone with asthma could perform (about asthma) will be always to get and track asthma. By halting a regulator medicine, the person will be set in peril of building asthmatic conditions. From the outbreak, treatment will require heading to urgent care or the emergency department, where the patient has a threat to become vulnerable with COVID-19 from someone. Thus, as it were, by going to track asthma, the patient with asthma is diminishing the chance of vulnerability to COVID-19 [20].

It is very crucial to not forget that there are assortments. The SARS-CoV-2 illness does not appear to bring about asthma exacerbations. No matter it is always vital for asthma patients to maintain their asthma under the very greatest potential control. Similarly, their lungs will probably be prepared if an allergen or any disorder results in a worsening of asthma [25]. The main concern for people with allergies in this outbreak is to keep on doing exactly what you are doing from first – maintain taking your controller medication and inform your health provider regarding any indications which you could develop. Furthermore, make sure to maintain social distancing and wash your hands.

Symptoms of COVID-19 in Asthma Patients

COVID-19 spreads rapidly from person to person through contact with contaminated droplets. An individual may likewise have the option to pass onto the disease before symptoms develop. Others may remain asymptomatic yet pass on the infection. According to the CDC, manifestations of COVID-19 might appear 2–14 days after contact with the illness. Symptoms may include:

- Fever
- Dyspnea
- Dry cough
- Aching muscles
- Pharyngitis (sore throat)
- Headache
- Chills
- Anosmia and hypogeusia (loss of smell or taste)

As stated by the United Nations, the majority of people recover with no special cure from COVID-19. They also judge that one in every six people who contract COVID-19 will become gravely sick and can suffer difficulty in breathing. Individuals with asthma should watch out for the following symptoms:

- An increase in chest tightness or wheezing
- Dyspnea
- Early morning or nighttime coughing
- More repeated use of rescue inhaler

Precautions and Measures to Be Taken by Asthma Patients Against COVID-19

People with allergies might possess concerns regarding the way they are influenced by COVID-19. The way to reduce the probability of fabricating asthma is by keeping illness avoidance habits up and curbing the illness. People with allergies have to play it safe once any form of the disease is dispersing inside their area. Precaution measures are described in the following subsections.

Taking All Asthma Medications as Directed Individuals should have all asthma prescriptions, for example, steroid inhalers, saving inhalers pills, and biologics as coordinated. Uncontrolled asthma can be a very severe medical hazard for those who have asthma. The Asthma and Allergy Foundation of America (AAFA) implies that people will have a 14–30-day distribution of their prescriptions [18]. An asthma task program is just really a personalized plan which individuals may apprehend to regulate their asthma. This includes the following:

– Having a good supply of medication
– Knowing how to use an inhaler correctly
– Avoiding asthma triggers
– Disinfecting and cleaning touched surfaces, such as door countertops and handles
– Avoiding any cleaning products that could trigger asthma
– Following measures to decrease tension, which might cause asthma attacks

Individuals ought to oversee intense asthma scenes with an inhaler, for example, albuterol. A report in the Journal of Allergy and Clinical Immunology debilitates and points out that the utilization of nebulizers is recommended only when it is an emergency. The reason is that nebulizer can build the danger of spreading infection droplets noticeable all around, conceivably circulating the infection to others close by.

Avoiding Potential Asthma Triggers Usual asthma triggers include:

– Smoking tobacco
– Pollen, pets, dust mites
– Air pollution
– Extreme weather conditions
– Intense exercise
– Mold
– Acid reflux
– Stress
– Strong odors
– Food additives or alcohol such as sulfites

Following COVID-19 Infection Prevention Recommendations

– Cleaning hands regularly by utilizing an alcohol-based hand sanitizer when water and soap are not available
– Avoid touching the eyes, nose, and mouth with unwashed hands
– Covering the nose and mouth with tissue paper along with perhaps even a sleeve after coughing or sneezing
– Washing hands after throwing used tissues in the garbage
– Avoid touching surfaces which others have touched
– Disinfecting and cleaning touched surfaces, such as door countertops and handles
– Avoid contact with individuals who are sick especially if they have a cough, fever, or both
– Practicing social distancing from others in public places
– Getting vaccinated for the flu if at all possible

Keeping the Immune System Strong An individual desires a healthier immune system to fight any illness, notably COVID-19. Practicing these policies helps fortify the immune system:

– Aspire for not less than 7 h sleep per night
– Reduce stress levels as much as possible
– Eat a diet rich in vegetables and fruits
– Get regular physical exercise
– Maintain a healthy weight

From the observational studies in Spain, exploration of 2,034,921 patient's data using natural language processing (NLP) and artificial intelligence (AI) techniques found 71,192 (about 1.4%) patients suffering from asthma and it shows that the frequency of COVID-19 infections is low in asthma patients [19]. Hospitalization of the suspected case is needed only for the older age groups.

A recent analysis of 1590 COVID-19 patients in China shows the absence of asthma patients which speculates that the use of TH2 asthma medication minimized the susceptibility to COVID-19 in patients. Contrarily, over 12% of hospitalized 393 COVID-19 cases were chronically documented for asthma. A study on 1298 patients ailing from coronavirus declares that both non-asthmatic and asthmatic patients were observed to be nonassociated with the adverse effects of COVID-19 irrespective of obesity, age, and various comorbidities [9]. Among 376 patients in Spain who were found SAR-CoV-2 infected on RT-PCR tests, some of the infected patients did not require to be hospitalized with previous records of mild asthma symptoms and were under medications [13]. A larger data sample can yield the correlation decision on dependencies of asthma and its susceptibility to COVID-19 [2].

Angiotensin-converting enzyme (ACE) 2 can somewhat contribute to the treatment against the widespread impact of COVID-19 inflammation since it acts as an inhibitor of lung cells [2]. Two mediators of asthma like transmembrane protease serine 2 (TMPRSS2) and angiotensin-converting enzyme 2 (ACE2) were treated with IL-13, and it is found that inflammation decreases ACE2 and increases TMPRSS2 in asthma [7, 8].

Based on Poisson regression models of classification, records of 1526 COVID-19-positive patients were classified to diagnose the asthma medications in them (until April 2020). Only 14% of positive patients are found to be using corticosteroids for treating asthma. It is noted that mild asthma patients are not subjected to hospitalization in COVID-19 prospects [3].

During the epidemic, while keeping a space of 2 meters has been proposed, regular visits to health care and medical centers might be postponed or handled through telemedicine. But patients with asthma need to carry facial visits even during the COVID-19 outbreak to maintain control of asthma. However, there is no consensus on how to prioritize medical services for patients. South Korea's government has designated 'Public Relief Hospitals' throughout the length of the COVID-19 epidemic to stop the virus from spreading and also to ensure overall

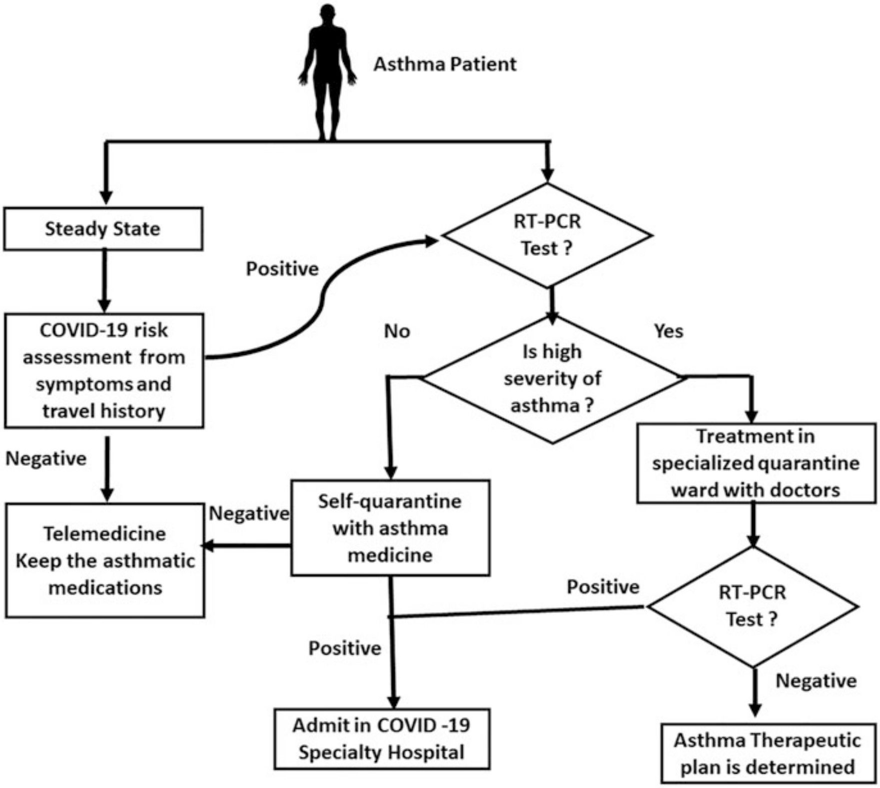


Fig. 2 Specialized clinical plan for the management of asthmatic patients during the COVID-19 pandemic [43]

services. The goal of this system is to separate areas for general patients from COVID-19 patients to prevent infection in the hospital as shown in Fig. 2 [43]. Therefore, general patients without risk of COVID-19 or some other suspicious symptom of COVID-19 might be permitted to visit health institutions. Ironically, under the Public Relief Hospitals system, although patients with asthma should be prioritized for face-to-face care, optimal medical services cannot be given until a negative result of the COVID-19 RT-PCR evaluation is confirmed since they generally pose respiratory symptoms like cough, sputum, and dyspnea, which are difficult to tell apart from COVID-19.

Asthma is considered to be one of the chronic diseases that can increase the critical symptoms with COVID-19 [14]. Though the symptoms of asthma and COVID-19 seem to converge in terms of shortness of breath and dry cough, there exists a distinction in the occurrence of discomfort in chest, and fever which is considered a common symptom for COVID-19 to occur. However, the behavior of the Severe Acute Respiratory Syndrome Coronavirus 2 virus is found to be different from that of asthma respiratory viruses.

Assessing the risk factors, the prevalence of asthma is lower in COVID-19 adult patients. However, various studies on the group of pandemic-affected patients suggest that asthma is common in several patients [15]. No incidence of severe asthma occurrence is seen in COVID-19-positive cases [16]. According to the American College of Allergy, Asthma, and Immunology (ACAAI), there is no evidence of upsurge of deaths, attacks, or infection rates from COVID-19 in individuals facing severe asthma. Also, there is no proof that allergic medication, especially inhaled corticosteroids, increases the risk of COVID-19 contamination. Overlapping of the symptoms can mislead to the diagnosis of COVID-19.

The question arises of managing the COVID-19-infected individual with serious asthma background. The Global Initiative for Asthma recommends continuing of medications for asthma control [16]. American Academy of Allergy, Asthma & Immunology (AAAAI) declares that there is no clear proof that treating asthma patients with anti-*IL4/IL13*, anti-*IgE* or anti-*IL5Ra* medicines will make the asthma patients somewhat immune to COVID-19 reactions. American College of Allergy, Asthma, and Immunology (ACAAI) states that inhalation of corticosteroids does not escalate the risk of COVID-19 [17]. Also, other national societies such as British Thoracic Society (BTS) and the Italian Society of Allergy, Asthma & Clinical Immunology (SIAAIC), European Respiratory Society (ERS) and European Lung Foundation (ELF) are rigorously aiming and working toward the eradication of this epidemic [18].

According to the Children's Hospital of Philadelphia (CHOP), there is a drastic decline in requirements for emergency departments for treating asthma patients as observed because of the COVID-19 pandemic [19]. The order of stay at home, work from home, mandating covering of mouth and nasal with masks, and social distancing has dropped the average number of people visiting the emergency department or hospitalizing on acute asthma attacks [20]. Since there is no clear evidence that asthma patients are more prone to COVID-19 infection. There is a need for an epidemiologic study for analyzing the risk of COVID-19 in severe asthma patients.

3 A Conjunction of COVID-19 and Obesity

Obesity is considered to be a silent pandemic in many countries. In western countries, obesity is at pandemic levels. Adults of age above 20 years with obesity are detrimental [5]. 34% of the US population being obese, one ceases to be at higher risk of infection. Obesity and overweight are defined as immoderate fat accumulation, so they represent a hazard to health [16].

Obesity is a perplexing disease including an unnecessary amount of fat. It is not merely a cosmetic concern. It is an issue that builds the risk of ailments and issues, for example, heart disease, hypertension, diabetes, and certain cancers. Most of the world's population lives in countries where overweight and obesity kill a larger number of people than underweight. One gets fat because of the intake of an excessive number of calories to be utilized by the body. This happens as a result of eating foodstuffs abundant in sugar and fat and reduction in athletics. Overweight disrupts the physical activity causing variations in Renin Angiotensin Aldosterone System (RAAS) [22, 23].

There are many reasons why some people find it difficult to avoid obesity. Often, obesity is caused by a composition of genetics, weaved with nature and a personal diet, and a choice of exercise. The fantastic thing is that moderate weight loss can improve or prevent issues related to obesity. Physical exercise, dietary modifications, and changes in eating behavior will allow you to shed weight. Drug and weight loss procedures are extra means to deal with obesity.

Effects of COVID-19 in Obese Patients

Dyslipidemia, cold, cough, fever, fatigue, diarrhea, and hyperuricemia are some of the SARS-CoV-2 viral infections in obese children [4]. Obese people are more susceptible to the current pandemic leading to the two pandemics in coalition [24]. Adiponectin which has high anti-inflammatory properties gets reduced in obesity turning to a higher risk of developing pneumonia which is near to the symptoms of COVID-19. The variations in adipokines disturb the immune cells causing airway injury, dysregulating the metabolism creating a bidirectional link for the decline in lung functioning [1]. Higher angiotensin-converting enzymes 2 (ACE2) in adipose tissue increases vulnerability to COVID-19 infection [25].

The comorbidities associated with obesity place a challenge to verdict the answer based on evidence. The dysregulation of immune hormones by lower adiponectin or higher cytokines gives the impaired prediction of whether obesity is highly sensible to COVID-19 or not [26]. In literature, there are no data reporting that patients with obesity are having a higher risk of infection with COVID-19. However, the Intensive Care National Audit and Research Center (ICNARC) reports that about 72% of 776 COVID-19 confirmed cases in the UK were obese [27].

A Case study in Mexico on datasets from Health Ministry Database comprising the verbal records responded by the hospitalized 32,583 COVID-19 patients reveal that about 35% of patients with a history of obesity (males of average age 38 and females of average age 32) and associated comorbidity are more likely to have affected by the COVID-19 infection [28]. Comorbidities of obesity and Type 2 diabetes contribute to the mortality and severity of COVID-19 risk [29]. Investigations from 1,77,133 patients predicted that the mortality rate is high for severely obese patients of age above 65 years [30]. According to the CDC, severe obesity is the risk of coronavirus diseases. It increases the risk of type 2 diabetes, dysglycemia, metabolic syndrome (MetS), hypertension, blood pressure, heart failure, and arterial disease [31].

Alterations in lifestyle and diet are usually because of environmental modifications, which are apparent during this age of the epidemic, curfew, and quarantine [32]. Quarantine may lead to changes in adolescents and obese children and result in insulin resistance, hypertension, low nutritional food intake, dysregulation of hormones, disrupting the functioning of organs of the body. This contributes to the increased factors related to the severity of COVID-19 risk [33]. The obesity was neglected in consideration of the risk factors of COVID-19. But the quarantine and lockdown ailments lead to obesity generating two pandemic conjunctions [34, 35].

Governmental actions under COVID-19 in different nations included a curfew until further notification. This prompts an adjustment in the way of life and a decline in exercise practice in their population. In this way, aside from embracing work-from-home concepts, the individuals in these nations ceased visiting parks and gyms to exercise. Many people may turn out to be nervous, stressed, or unable to do any form of physical work along with many others becoming obese and consuming more foodstuffs without any sort [11]. As the COVID-19 outbreak keeps reaching worldwide, it appears that it is not only having an impact on health. It has an impact on mental health through the panic of coming down with the virus, worrying about rumors about the disease, social isolation, and financial pressure, family, obsessive beliefs, and information overload on mental health. These may lead to stress and discomfort levels, which will cause physical health problems including cardiovascular disease. A study reported a relationship between incessant worry with obesity and energy intake and diet quality. Human body weight can be influenced by stress through behavioral mechanics that are psychological [12].

Biological Mechanisms During Obesity

– Actuation of the hypothalamic-adrenal-pituitary axis, which contributes to discharging cortisol, that can impact mass by simply boosting the stimulating ingestion by itself, eating brain sensitivity into potentiating reward.
– Inciting reward centers in brain-like striatum and nucleus accumbent, which increase the inclination to eat up food having a high content of sugars, that is, fat.
– Stress influences brain areas responsible for self-regulation, which is crucial to restrain an individual's functions like eating and exercising which are necessary to get a handle on fat loss.

Behavioral Mechanisms During Obesity

– Stress may lead individuals to consume higher amounts of food that is exceptionally palatable with a greater trend.
– Stress reduces the propensity for physical exercises.
– Stress can interrupt sleep schedule leading to shorter intervals of sleep accompanied by high chances of obesity.

Regrettably, obesity in a person with COVID-19 disease is certainly not a decent signal. Obesity, for this situation, can seed serious symptoms and inconveniences. Obese patient experiences a lot of health issues. It is inclined to be much more difficult to acquire diagnostic imaging (like on imaging machines, as you can find weight limits) [29]. It is more tedious for the medical team to move obese people or place them. Also, researches indicate that obesity disturbs the immune system through various functions. A portion of these components is diminished, such as altered monocyte, lymphocyte function, natural killer cell dysfunction, cytokine production, reduced dendritic cell function, macrophage, and reduced reaction to antigen/mitogen stimulation.

In defending a virus, low degrees of resistance are not accepted. It has been found that 50% of people having COVID-19 are related to hypercytokinemia. Considerable scenarios have leukocytosis, lymphopenia, especially T cells, and growth in neutrophil lymphocyte levels (NLR), in addition to lower levels of basophils, monocytes, and eosinophils. Inflammatory cytokines were also raised in extreme cases, including *IL-6, IL-10, IL-2R, TNF-*α. The excessive production of those proinflammatory cytokines brings about what has been depicted as a cytokine storm causing an increased risk of multi-organ failure and vascular hyperpermeability with COVID-19 [30].

Besides, obesity is related to different comorbidities that are no less risky than obesity itself like type 2 diabetes mellitus, coronary artery diseases, hypertension, cerebrovascular strokes, osteoarthritis, and atherosclerosis. These diseases, without anyone else or nursing, influence body fitness. They make a person more defenseless to contract COVID-19 disease. Luckily, these comorbidities need quite a while to happen, that long period is not anticipated by specialists working for COVID-19 to disappear, thanks to the medical and scientific development in the future.

From a cardiovascular view, genetic and trial evidence conclusively demonstrates that obesity (and body fat mass) is causally linked to hypertension, diabetes mellitus, cardiovascular disease, stroke, diabetes, atrial fibrillation, cardiovascular disease, and cardiovascular failure. Obesity potentiates adverse cardiorenal outcomes, the premature development of cardiovascular disease, and multiple risk factors. There is also a metabolic difficulty. In individuals with diabetes mellitus, or at high risk of diabetes mellitus, obesity and excess ectopic fat cause disability of insulin resistance and diminished β-cell function. Both the latter limit the capacity to elicit an appropriate response to an immunologic challenge causing some patients with diabetes mellitus to require substantial amounts of insulin during illnesses. The regulation of the metabolic process required for host defense and also for the complex cellular interactions is lost leading to operational immunologic deficit. COVID-19 may also directly disrupt pancreatic β-cell performance through interaction with all angiotensin-converting receptor two. Furthermore, obesity enhances thrombosis, which is pertinent given the association between high rates of venous thromboembolism and severe prothrombotic and COVID-19-disseminated intravascular coagulation [36].

Beyond thrombotic along with cardiometabolic consequences, obesity has more detrimental impacts on lung function, decreasing forced vital capacity, and forced expiratory volume [37]. Higher relative fat mass can be connected to such adverse adjustments, perhaps pertinent to emerging reports of bigger serious disease from COVID-19 in some specific ethnicities such as Asians. Asians frequently display lower cardiorespiratory fitness and transmit more fat tissue in lower BMIs. Together with extreme obesity (e.g., BMI > 40 kg/m^2), maintenance for men confessed to intensive therapy components can be impeded since these patients tend to be more complicated to picture, solidify, nurse, and even rehabilitate.

Related to the response, there is a distinct association between basal and also obesity inflammatory status characterized by higher circulating C-reactive protein amounts and also interleukin. Adipose tissue in obesity has been proinflammatory with increased expression of cytokines and especially adipokines. There is additionally dysregulated tissue leukocyte expression. Also, inflammatory macrophage (and inherent lymphoid) subsets replace tissue regulatory (M2) phenotypic cells. Obesity can be an independent and causal risk factor for the development of a disease, for example, psoriasis, suggesting that adipose state may have consequences on other environmental provocation. In terms of host defense, obesity wreaks adaptive immune responses to the flu virus and conceivably could do in COVID-19 infection. Individuals may exhibit greater viral shedding, suggesting the prospect of great viral vulnerability. This could be aggravated in families, which are more common and more predominant in the socioeconomically deprived communities. These observations cause a possibility of obesity to give rise to a more adverse virus–host immune response relationship in COVID-19. Poorer supplements and hyperglycemia may further aggravate the situation in some individuals [5].

Much of COVID-19's attention was around elderly people. It is crucial to keep in mind that muscle reduction and weight tissue start to decline in people who have comorbid diseases like cardiovascular and respiratory ailments especially at old age relative mass gains. Age can be correlated with diabetes mellitus and hypertension as a result of metabolic efficacy and vessels. Individuals who are elderly (>70 years), very similar to younger obese individuals, have a less cardiorespiratory reserve to deal with COVID-19 illness. Immune senescence is recognized, as is the idea of inflammation, and virus–host dynamics at the outcomes that are older and infection may be influenced by both [41].

It is preferable to avoid obesity as much as possible, particularly in this time of the COVID-19 epidemic. Trying to keep smart dieting food habits like numerous salads and green nourishments. Day-to-day moderate intensity athletics at home is prudent. Watching online recordings of exercise can help in motivation for doing gym and yoga from home. It is smarter to start decreasing long hours of using cellphones and increase doing different physical exercises. Getting adequate rest is vital for the circadian rhythm of hormones and the immune system. Attempting to have an ordinary sleep cycle by zoning out early and getting up right on time each day and not switching rest hours.

Obesity can cause alterations in the renin-angiotensin-aldosterone system (RAAS) that promotes derangement. Adipocytes might substantially result in the creation of circulating angiotensinogen which, later through metabolic process from renin and angiotensin-converting enzyme 2(ACE2), produces angiotensin II (Ang-II). Therefore, obesity may lead to hyperactive RAAS [42]. In a smaller study, patients having COVID-19 illness were proven to possess Ang-II levels related to the severity of lung injury. High Ang-II levels within the gut may cause damage in addition to pulmonary vasoconstriction resulting in ventilation/perfusion mismatch along with hypoxemia, boosting severe kidney disease. In people who have type 2 diabetes mellitus, Ang-II levels were proven to correlate with body fat loss reduction. The baseline Ang-II levels in the fat can overtake COVID-19-caused Ang-II amount to grow, which may lead to lung disease. Ang-II levels decreased in response to fat reduction. Therefore, physical activity and dietary modification could be effective in reducing this mechanism of disease in obesity [9].

Obesity is related to diminished pulmonary function together with expiratory reserve volume, operational breathing compliance, and capacity. Increased abdominal fat impairs pulmonary role at supine patients by the diminished trip, whereas the bottom of this lung venting can be diminished leading to low oxygen-saturated blood flow by degrees. Additionally, long-term inflammation and elevated amounts of circulating proinflammatory cytokines related to obesity, like leptin, tumor necrosis factor α, and interleukin, can hamper immune reaction and influence both the lung parenchyma and bronchi, consequently leading to the increased morbidity related to obesity in COVID-19 disease [38]. Ultimately, obese individuals require exceptional supervision in their ordinary life and also in this uncommon time of the epidemic. Future researches about the connection between Body Mass Index (BMI) and COVID-19 disease are expected to be announced if obese individuals are at a higher risk of contracting the infection or not. Furthermore, a study assesses whether obesity intensified the distress in hospitals during the epidemic is expected to avoid potential risk.

Symptoms of COVID-19 in Obese Patients

In two cohorts of Chinese grown-ups with COVID-19, those with obesity were at least three times more likely to have a serious instance of the infection than those with typical weight, as indicated by two studies published in Diabetes Care. Furthermore, the increase in obesity was related to increased chances of severe COVID-19 and the relationship between obesity and symptom seriousness was stronger for men than for women.

– Fever
– Dyspnea
– Dry cough
– Aching muscles
– Pharyngitis (sore throat)
– Headache
– Chills
– Anosmia and hypogeusia (loss of smell or taste)

Precautions and Measures to Be Taken by Obese Patients Against COVID-19

It is not astonishing that individuals with a preexisting condition of obesity may feel more restless than normal. It is common and reasonable. The infection quickly grasped control of the world, leaving specialists scrambling to see how it functions, who can be in most peril, and how to control its spread. It is just right that individuals have the data they need to remain healthy and safe. Many people struggle with weight management. Being diagnosed as overweight or to have clinical or hereditary obesity is certainly not a shameful thing, nor does it mean that assistance is not available. Obesity can be hereditary or caused by disease or medication, yet it is likewise regularly connected to the absence of instruction in healthy eating, financial confinements, or psychological trauma, and mental issues. It is a widespread issue that deserves consideration and empathy.

CDC claims that patients with obesity who practice good physical conditioning and are metabolically healthy could be about the much lower risk end. People still need to practice distancing. Taking self-isolation and CDC precautions like cleaning hands well, especially after being with people, avoid touching the eyes, mouth, nose, or whatever else having unwashed hands, and next obstruction within 6 feet space out of people or in any parties of any type. Family members or housemates who have any traveling history must isolate them, as they may carry the virus if they do not have symptoms. This might appear difficult to display, but the spread is both rapid and highly contagious. The CDC believes that droplets created by coughs or sneezes easily pass from person to person [39].

Exercise In the meantime, it may be worthwhile to review the lifestyle to help a person stay fit even during the COVID-19 epidemic. Start a daily exercise routine or start a healthy diet while living alone, for example, these small changes in daily routine practice can benefit after a quarantine. It might be a good idea to look for a food-shipping agency or to begin adding foods. Also, you should not leave your house. The CDC pressurizes the continuation of all current drugs – including ACE inhibitors – that can carry other risks.

Recalibrate Patient's Diet If people are excited about exercise and diet, now is the time to reexamine your relationship with well-balanced meals. Easy alterations? Since one cannot go to the gym, they can make sure to spend more time walking rather than sitting. Spending 15 min on the stairs, doing push-ups, watching televisions, CDC offers. "Think about muscles as little factories one needs to keep churning and burning."

An increase in death rate and no success in vaccination have worsened the clinical dependencies. Contagiousness of infection when inhaled is more since the capacity to produce BMI gets reduced. Thus, the presence of obesity above the age of 60 years is now recognized as an independent risk for admission of a new infection [41]. Critical monitoring of severely obese patients is recommended [39]. Obesity is a poorly documented comorbidity in the COVID-19 epidemic and now identified as a major risk component for serious COVID-19 infections, including those under 60 years of age. Obese people should be closely monitored because of the risk that the COVID-19 virus may increase. People should be concerned with regular excess weight check-up, mental and health care advice, nutritive food supplements, moral support, and guide for the isolation of obese and adolescent children with COVID-19 suspects [40].

4 Conclusion

With the coincidence of the allergy season with COVID-19, recently CDC has illustrated that muscle pain, loss of taste, headache, and sore throat are additional symptoms for COVID-19. Global Cooperation is mitigating the threat of COVID-19 Pandemic. World Health Organization is contributing as lead to a large extent with national actions playing the acute role in disease control. For now, as per the CDC guidelines, using sanitizers that contain 60% ethanol for at least 20–30 s, avoiding touching of eyes, nose, mouth, disinfecting frequently touched areas, maintaining social distancing of 6 feet are some of the preventive measures being followed for safety. People older than 65 years with underlying asthmatic medications and obesity (40 BMI or higher) are at higher risk of facing COVID-19 attacks. Strictly following the government's actions, caring for a daily exercise, a healthy lifestyle, nutritional food intake, optimized sleep, anxiety management, mindfulness and stress-relieving activities, moderating smoke and alcohol can supplement the prior solutions to boost up the immune system at a personal level [1].

The continuous study on affected communities is on an evaluation by pediatricians to diagnose the critical risk factors associated with the pandemic patients diseased with asthma or obesity. The occurrences of COVID-19 are higher in patients with comorbidities of asthma and obesity. COVID-19 causes other respiratory diseases to occur. Patients with chronic respiratory ailments seem to be more defenseless to COVID-19 infection [42]. For the progression of COVID-19, old age, obesity, chronic heart, and high blood pressure are some of the reported symptoms for identifying the risk. At present, there is no special vaccine for patients suffering from any disease. Many organizations have claimed that they have created vaccines but those vaccines are not particular for heart patients, but they are for other general diseases. Patients with asthma can lessen their odds of having a severe disease by keeping their asthma controlled by following disease control tips and by exercising. Individuals with asthma might well not be at risk of obtaining a COVID-19 contract. However, they could be at higher risk of complications.

There is no evidence that more issues are faced by individuals with asthma. A small quantity of evidence shows that people with COVID-19 and asthma are currently still also recovering. People who have asthma should proceed to carry their medication prescribed by their physician. Unchecked asthma may put individuals at an increased risk of complications and respiratory difficulties. There is a need to examine the effects of asthma and obesity patients being caught by the COVID-19 virus. However, the licensed biologics for asthma treatment may target add-on experiments to maintain and reduce steroid uses. This survey can help in determining the present status of the co-relation of asthma and obesity with other comorbidities for the prevalence of contagious COVID-19. Evidence from several pieces of literature supports that air pollution, lung inflammations, asthma, and obesity are predisposing factors for the progression of coronavirus infections.

References

1. A. Hancu, F. Mihaltan, Two pandemics coalition: Covid-19 and obesity. Intern. Med. **17**, 37–45 (2020)
2. E. Hernandez-Garduno, Obesity is the comorbidity more strongly associated for Covid-19 in Mexico. A case-control study. Elsevier Obes. Res. Clin. Pract. **14**(4), 375–379 (2020)
3. K.D. Chhiba et al., Prevalence and characterization of asthma in hospitalized and non-hospitalized patients with COVID-19. Elsevier Am. Acad. Allergy Asthma Immunol. **146**, 307–314 (2020)
4. C.A. Nogueira-de-Almeida et al., COVID-19 and obesity in childhood and adolescence: A clinical review. Elsevier-Pediatria, 1–13 (2020). https://doi.org/10.1016/j.jped.2020.07.001
5. F. Sanchis-Gomar et al., Obesity and outcomes in COVID-19: When an epidemic and pandemic collide. Mayo Found. Med. Educ. Res. **95**(7), 1445–1453 (2020)
6. G. Muscogiuri, Commentary: Obesity: The "Achilles heel" for COVID-19? Elsevier Metab. Clin. Exp., 1–2 (2020). https://doi.org/10.1016/j.metabol.2020.154251
7. H. Kimura et al., Type 2 inflammation modulates ACE2 and TMPRSS2 in airway epithelial cells. J. Allergy Clin. Immunol. **146**(1), 80–88.e8 (2020)
8. L. Wang, D. Foer, D.W. Bates, J.A. Boyce, L. Zhou, Risk factors for hospitalization, intensive care and mortality among patients with asthma and COVID-19. Elsevier J. Allergy Clin. Immunol. **146**(4), 808–812 (2020)
9. S. Lovinsky-Desir, D.R. Deshpande, A. De, L. Murray, J.A. Stingone, A. Chan, N. Patel, N. Rai, E. DiMango, J. Milner, M. Kattan, Asthma among hospitalized patients with COVID-19 and related outcomes. J. Allergy Clin. Immunol. **146**(5), 1027–1034.e4 (2020)
10. P. Halvatsiotis, Demographic and clinical features of critically ill patients with COVID-19 in Greece: The burden of diabetes and obesity. Elsevier Diabetes Res. Clin. Pract. **166**, 108331 (2020)
11. C. Lubrano et al., Is obesity the missing link between COVID-19 severity and air pollution? Environ. Pollut. **266**, 115327 (2020)
12. M.Q. Almerie et al., The association between obesity and poor outcome after COVID-19 indicates a potential therapeutic role for montelukast. Elsevier Med. Hypotheses **143**, 109883 (2020)
13. E. Garcia-Pachon et al., Asthma prevalence in patients with SARS-CoV-2 infection detected by RT-PCR not requiring hospitalization. Elsevier Respir. Med. **171**, 106084 (2020)
14. J.-S. Yeom, Patterns of viral clearance in the natural course of asymptomatic coronavirus disease 2019 (COVID-19): Comparison with symptomatic nonsevere COVID-19. Elsevier Int. J. Infect. Dis. **99**, 279–285 (2020)
15. Morais-Almeida et al., COVID-19, asthma, and biological therapies: What we need to know. World Allergy Organ. J. **13**(5), 100126 (2020)
16. E. Haroun-Díaz et al., Severe asthma during the COVID-19 pandemic: Clinical observations. Elsevier Am. Acad. Allergy Asthma Immunol. **8**(8), 2787–2789 (2020)
17. Reflecting on World Asthma Day in the era of COVID-19. Lancet Respir. Med. **8**, 423 (2020)
18. Asthma in children during the COVID-19 pandemic: Lessons from lockdown and future directions for management. Lancet Respir. Med. (2020)

19. J.L. Izquierdo, The impact of COVID-19 on patients with asthma. Eur. Respir. J. (2020). https://doi.org/10.1183/13993003.03142-2020
20. COVID-19 associated with dramatic decline in ED use by pediatric asthma patients. Med. Express (2020). https://medicalxpress.com/news/2020-06-covid-decline-edpediatric-asthma.html
21. M. Morais-Almeida, Asthma and the coronavirus disease 2019 pandemic: A literature review. Int. Arch. Allergy Immunol. **181**(9), 680–688 (2020)
22. A. Beurnier et al., Characteristics and outcomes of asthmatic patients with COVID-19 pneumonia who require hospitalisation. Eur. Respir. J. **56**, 2001875 (2020)
23. S. Garg et al., Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 – COVID-NET, 14 States, March 1–30, 2020. Morbidity and Mortality Weekly Report
24. Mahdavinia et al., Asthma prolongs intubation in COVID-19. Elsevier J. Allergy Clin. Immunol. Pract. **8**, 2388–2391 (2020)
25. G.P. Nikolaos et al., Impact of COVID-19 on pediatric asthma: Practice adjustments and disease burden. Elsevier J. Allergy Clin. Immunol. Pract. **8**(8), 2592–2599.e3 (2020)
26. M. Morais-Almeidaa et al., COVID-19 and asthma: To have or not to have T2 inflammation makes a difference? Pulmonology **26**, 261–263 (2020)
27. U. Radzikowska et al., Distribution of ACE2, CD147, CD26 and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors. European. J. Allergy Clin. Immunol. **75**(11), 2829–2845 (2020)
28. D. Rastogi, Quantifying the contribution of obesity to incident childhood asthma: It's about time. J. Am. Acad. Pediatr. **142**(6), e20182979 (2018)
29. A. Simonnet, High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. Obes. Res. J. **28**(7), 1195–1199 (2020)
30. C. Finelli, Obesity, COVID-19 and immunotherapy: The complex relationship. Immunotherapy (2020). https://doi.org/10.2217/imt-2020-0178
31. S. Chiappetta, COVID-19 and the role of chronic inflammation in patients with obesity. Int. J. Obes. **44**, 1790–1792 (2020)
32. Y.B. Omar, Predicting mortality due to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. J. Clin. Endocrinol. Metab. **105**(8), dgaa346 (2020)
33. D. Petrakis, Obesity a risk factor for increased COVID 19 prevalence, severity and lethality (Review). Mol. Med. Rep. **22**, 9–19 (2020)
34. N.N. Pettit, Obesity is associated with increased risk for mortality among hospitalized patients with COVID-19. Obesity (2020). https://doi.org/10.1002/oby.22941
35. H.H. Mohammad, Asthma in COVID-19: An extra chain fitting around the neck? medRxiv (2020)
36. I. Akoumianakis, T. Filippatos, The renin–angiotensin–aldosterone system as a link between obesity and coronavirus disease 2019 severity. Obes. Rev. **21**(9), e13077 (2020)
37. K. Markos, Association of obesity with disease severity among patients with coronavirus disease 2019. Obes. Res. J. **28**(7), 1200–1204 (2020)
38. A.A.D. Albashir, The potential impacts of obesity on COVID-19. Clin. Med. **20**(4), e109–e113 (2020)
39. J. Wang et al., Overweight and obesity are risks factors of severe illness in patients with COVID-19. Obesity **28**(11), 2049–2055 (2020)
40. A.T. Akenroye, R. Wood, C. Keet, Asthma, biologics, corticosteroids, and COVID-19. Ann. Allergy Asthma Immunol. **125**(1), 12–13 (2020)
41. Sattar et al., Obesity is a risk factor for severe COVID-19 infection: Multiple potential mechanisms. Circulation **142**(1), 4–6 (2020)
42. F. Sanchis-Gomar, C.J. Lavie, M.R. Mehra, *Brandon Michael Henry* (Mayo Foundation for Medical Education and Research, Giuseppe Lippi, 2020)
43. S. Kim, H.J. Jin, S.R. Kim, Management of severe asthma during the COVID-19 pandemic in Korea. Allergy, Asthma Immunol. Res. **12**(5), 897–901 (2020)

Case Study on COVID-19 Scenario over Highly Affected States of India

Jayshree Kharche, Shubhangi Kharche, and Mrinal Kharche

1 Introduction

COVID-19 pandemic has affected 213 countries and territories with a total of 10,583,932 confirmed cases and a death toll of 513,861 deaths as on July 1, 2020 [1]. It is an infectious disease caused by Corona virus 2019. The commonest symptoms of COVID-19 are dry cough, fever, and tiredness. Other symptoms are body ache, nasal congestion, runny nose, loss of taste sensation, sore throat, or diarrhea [2]. As per the latest report, backache, nausea, and rashes may also be COVID warning signs [3]. It is observed that symptoms begin gradually and are mostly mild. Even after getting infected, few patients do not develop any symptoms and are asymptomatic; 80% of infected population recover from the disease without special treatment. Approximately one out of every six COVID patients becomes seriously ill and develops breathlessness. Development of serious illness is observed in older people, and in patients with high blood pressure, heart problems, or diabetes. Patients having fever, cough, and breathlessness should seek medical attention [2]. According to WHO, primary human-to-human transmission occurs via respiratory droplets. When a person with infection coughs, sneezes, or talks, virus is released in the respiratory secretions. Also, if a person touches an infected surface and then

J. Kharche (✉)
Physiology Department, Symbiosis Medical College for Women (SMCW), Symbiosis International (Deemed University), Pune, Maharashtra, India
e-mail: jayshree.kharche@smcw.siu.edu.in

S. Kharche
Electronics and Telecommunication Department, SIES Graduate School of Technology, Navi Mumbai, Maharashtra, India
e-mail: Shubhangi.kharche@siesgst.ac.in

M. Kharche
Self Employed, SPACE, Pune, Maharashtra, India

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_18

touches facial parts, the person may get infected. Droplets can travel maximum 6 ft or about two meters. The coronavirus disease 19 (COVID-19) is a highly transmittable disease. With 18,256 COVID-19 fresh cases as on July 1, 2020, India's confirmed cases rose to 585,792, including 17,410 deaths. The number of active COVID-19 cases are 220,546, and 347,836, patients have recovered [1].

Following states are worst affected in India:

Maharashtra state is found to be the worst-affected state with 6364 new cases taking the total to 192990 cases, with 104687 recovered and 8376 deaths.

Tamil Nadu has 102721 with 1385 deaths and 58378 recovered cases.

Delhi has 94695 cases with 2923 deaths and 65624 recovered cases.

Gujarat 34600 has total cases with 1904 deaths and 24933 recovered.

Rajasthan has a total of confirmed 19052 cases with 15281 recovered cases and 440 deaths [4]. In India, the case doubling time is 15 days. Also, case fatality rate for COVID-19 is reduced from 2.83% in contrast to 6.19% worldwide. In India, the recovery rate for COVID-19 patients is 48.19% [5]. The recovery rate in India has increased from 11.42% on April 15 to 26.59% as on May 3 and to 38.29% on May 18 to the present 48.19% [6].

2 Specific Healthcare Problems/Difficulties in Indian States

High infectivity of corona virus and large population (about 1.2 billion) of India has posed a great challenge to Indian healthcare system. As compared to other countries, India spends about 3.6% of GDP on healthcare which is very low. The average for other countries for the same was 8.8% of GDP in 2018 [7]. National Health Profile–2019 data showed that in India, in total, there were 7, 13,986 government hospital beds available. This shows only 0.55 beds per 1000 population [8, 9]. The population aged 60 and above is vulnerable and there are higher chances of more complications in the age group. For elderly population, beds available in India are 5.18 beds per 1000 population. [8].

Low hospital beds in India and a pandemic-like coronavirus can lead to hospital bed crisis and may complicate the situation. Also, a major challenge will be that one COVID-19 positive admitted patient will occupy the bed for minimum 10 days. As per prediction 5–10% of total patients may require critical care and also ventilator support. India also lacks severely on ventilator support which may be needed hugely in this pandemic [8].

In one of the studies by Public Health Foundation of India, it was found that in the last two decades, free medicine availability in public healthcare facilities has reduced from 31.2% to 8.9% for inpatient department. Also, it declined from 17.8% to 5.9% for outpatient department [10].

As per the Indian government directives, COVID-19 patients are treated in three types of COVID dedicated facilities viz: confirmed clinically mild cases are admitted to COVID care centers (CCC), moderate cases are hospitalized to dedicated COVID health center (DCHC) and dedicated COVID hospital (DCH) for clinically assigned severe cases [11].

Therefore, for Indian scenario, prevention stands to be the best available option to fight the battle against corona. With growing counts of coronavirus cases, Indian policymakers are disseminating knowledge about preventive practices such as frequent hand washing and not touching the face. It is suggested that social distancing is an effective tool to "flatten the curve." These measures will prevent the health system from being overburdened. To delay community spread in India, Indian prime minister imposed lockdown in India in four stages starting from 25 March 2020. Lockdown is one of the important nonpharmacological interventions to reduce spread of infection and delay community outbreak. During lockdown, along with personal protective measures, transmission reduction strategies [12] include:

- Strict social distancing
- Stay home orders
- Social places and nonessential services closure
- No public gatherings
- Interstate travel restriction with screening at exit and/or entry
- Extensive case identification and isolation
- Contact tracing and quarantine
- Face covering with mask
- Work from home

As per ICMR scientist statement about peak in India, "We are very far away from the peak. Our preventive measures to curtail the disease are very effective and we are better positioned in comparison with other countries [13]."

For the assessment of the extent of spread of COVID-19, ICMR is conducting a sero-survey and almost 34,000 people are being tested as a part to study community transmission and will put up findings of it shortly [13].

Healthcare Facilities and Challenges in Maharashtra to Take Care of COVID-19 Scenario in Maharashtra

It was observed that Maharashtra lies below the national average of 0.55 beds per 1000 population. Maharashtra has only 11587 ICU beds. Maharashtra being the worst hot state by COVID-19, Mumbai and Pune have a maximum number of COVID cases. Mumbai has run out of hospital beds with increasing number of cases reported by health section in science wire news bulletin [14].

As reported in Indian express on May 30, 2020, Maharashtra is the first state to offer 100% free treatment to COVID-19 patients by bringing all residents of the state under the Mahatma Jyotiba Phule Health Insurance scheme. Also, an agreement with General Insurance Public Sector Association (GIPSA) for Mumbai and Pune city patients had been reached to help patients [15].

As on June 20, 2020, Mumbai recorded 1190 new cases, taking the total count to 65329. Mumbai also recorded 136 more deaths pushing the death toll to 3561. Considering bed scarcity, huge population, and increasing cases in Mumbai, a 30,000-bed Covid-19 care center for self-isolation has been created for those who do not have facility to self-isolate. Also, in a fortnight, the first open hospital in the country has been created by the Mumbai Metropolitan Region Development Authority (MMRDA) at Bandra Kurla complex. This open hospital is with 1,000 beds and 200 ICU beds. Similar initiatives to increase isolation facility have been implemented in cities like Pune and Nagpur [16].

Healthcare Facilities and Challenges in Tamil Nadu to Take Care of COVID-19 Scenario

Tamil Nadu has an average of 1.1 hospital bed per 1000 population which is better than the national average [9]. In the state, it was observed that, 10% of COVID patients require hospitalization, 20% require health centers, another 50% can be managed at care centers, and the rest can be put under home quarantine. Still, to arrange these many beds is a challenge for the state government. Many public places have been converted to isolation center and measures are taken by the state government to augment hospital beds on a daily basis, depending on the daily count of positive cases [17].

The Tamil Nadu government has announced that Chief Minister's Comprehensive Health Insurance Scheme (CMCHIS) will cover the treatment for COVID-19 in nongovernment hospitals. According to the government, under this scheme, treatment for COVID-19 can be availed in private hospitals that are empanelled under the scheme without making any payment [18].

Healthcare Facilities and Challenges in Delhi to Take Care of COVID-19 Scenario

Delhi has 1.05 hospital beds per 1000 population which is above the national average [9]. Even after additional funding, there is inadequacy of medical investment and healthcare infrastructure which is posing challenge to combating an effective response against the pandemic in Delhi. In addition, there is lack of health insurance coverage for more than 80% of the population, and approximately 68% of Indians have shortage or no access to essential medicines which may further worsen the pandemic [10].

Healthcare Facilities and Challenges in Gujarat to Take Care of COVID-19 Scenario

Gujarat has less number of hospital beds compared to the national average which is 0.3 per 1000 population [8]. For better treatment and hospital services to Corona patients, the Gujarat High Court directed the state authority that 50% of the total bed should be reserved to treat COVID patients in all the nongovernmental and private hospitals. Also, the court directed the Gujarat government to make the healthcare facility cheaper, available, and reasonable [19]. This move has improved the availability of hospital beds for the affected people.

Healthcare Facilities and Challenges in Rajasthan to Take Care of COVID-19 Scenario

Rajasthan has 0.6 hospital beds per 1000 population [8]. In the state, "Bhilwara model" won applause across the country by showing good control on the spread of pandemic. In the first phase of COVID-19 outbreak, Bhilwara district was among the most affected places in India. Rajasthan government imposed a curfew in the district. Curfew restricted essential services, facilitated substantial screening and house-to-house surveys to find out possible cases. Detailed contact tracing of each positive was carried out. This resulted in converting Bhilwara from most number of coronavirus cases to only one positive case since March 30 [20].

In addition to it, hospitals from the private sector in the state volunteered to join the fight against COVID-19 [21].

Other Measures to Control Spread of COVID-19

Aarogya Setu App is COVID-19 tracker launched by the Government of India. This app, depending on user's location, predicts risk for COVID-19. Also, it informs about COVID positive cases in the proximity of user [22]. For containment of COVID-19, new guidelines were laid down by the Government of India by which districts were divided into red, orange, and green zones depending on risk profiling of districts [23]. Further details of the zones are as follows:

1. Green zones: Districts with zero confirmed cases till date or districts with no confirmed case in the last 21 days.
2. Red Zones or Hotspot Districts: Depending on count of active cases, doubling rate of confirmed cases, efficiency of testing, and surveillance feedback, districts will be labelled as red zone.
3. Orange Zones: Districts which are neither red nor green are orange zones.

Based on this zonal classification, activities were permitted or restricted in the zones. This has helped to track the cases and contacts in the highly infected zone. Also, zoning of districts has reduced the spread of infection.

3 Challenges in Controlling COVID-19 Cases

Large numbers of COVID-19 cases are found in six states of India viz; Maharashtra, Tamil Nadu, Delhi, Gujarat, Uttar Pradesh, and Rajasthan. Fig. 1 shows the number of confirmed cases, recovered cases, and active COVID cases in the highly affected states of India as on May 31, 2020 [24]. Figure 2 shows the number of deaths and the number of new COVID cases in the highly affected states of India as on May 31, 2020. As seen, the highest number of COVID cases are found in Maharashtra (Figs. 1 and 2), as it shares 9.23% of India's total population (Fig. 3) [25] and 42% of its population resides in slums. As on May 31, 2020, Tamil Nadu reported second highest COVID cases (Figs. 1 and 2) Tamil Nadu shares 5.93% of India's total population (Fig. 3), that is, 8, 22, 47, 613 as seen in Fig. 4. In the initial stages, large number of COVID cases emerged in Tamil Nadu due to the people gathering at the market places. It was a challenging task for the frontline workers to maintain social distancing at such places. Delhi reported third highest COVID cases as on May 31, 2020 as seen in Figs. 1 and 2. It shares 1.38% of total India's population (Fig. 3). The major challenge for social distancing in Delhi is its highest population density as seen in Fig. 5. Gujarat and Rajasthan reported a large number of COVID cases as on May 31, 2020, occupying the fourth and

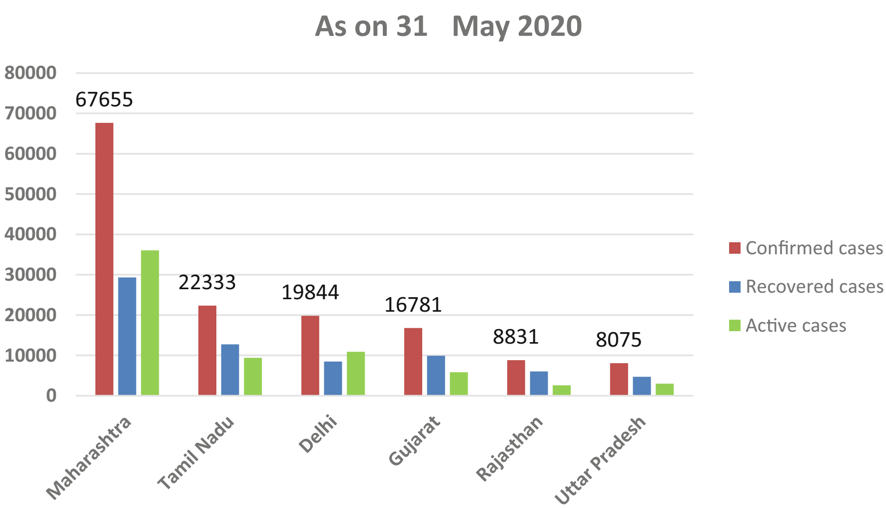


Fig. 1 Confirmed, recovered, and active cases in the six states of India

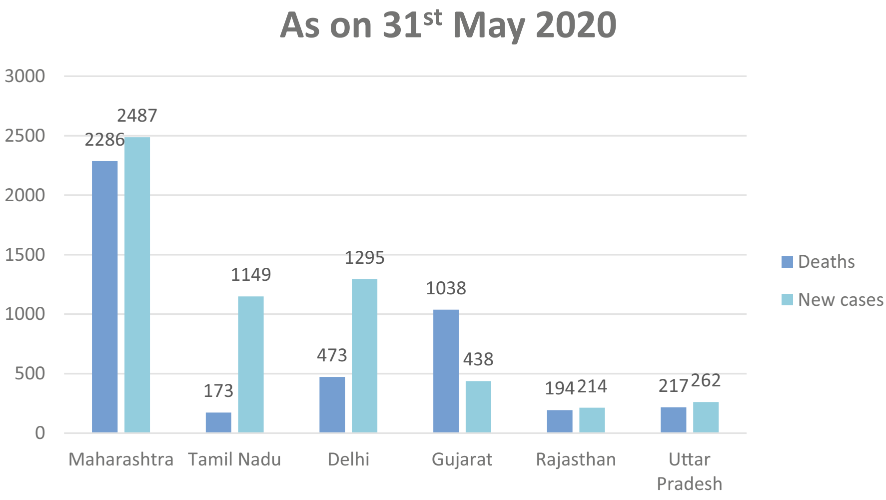


Fig. 2 Deaths and new cases in the six states of India

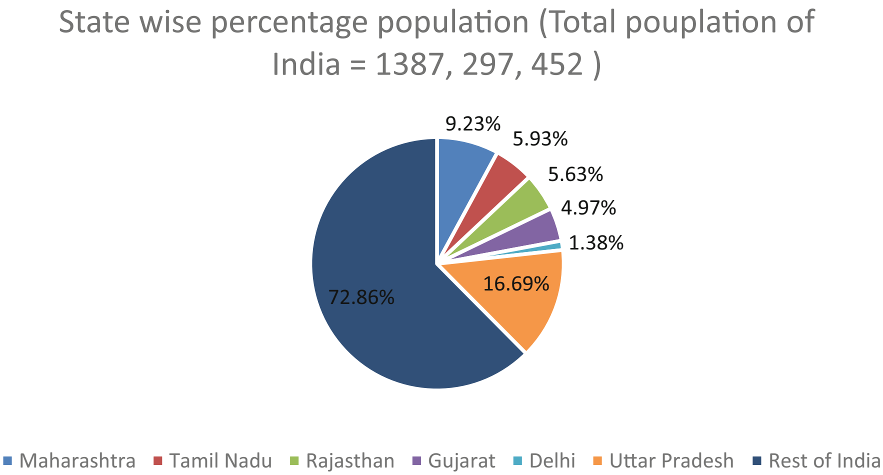


Fig. 3 Percentage share in total population of India

fifth highest positions, respectively, in India. Population wise, Rajasthan is ahead of Gujarat (Figs. 3, 4 and 5), but higher numbers of cases were reported in Gujarat due to foreign country migrants in the initial stages. Social distancing is a challenging task in the nonnotified slum population of 3.84 lakhs in Gujarat. In the initial stages, Uttar Pradesh (U.P.) reported comparatively a lower number of cases. But as on May 31, 2020, U.P. occupied the sixth highest position in COVID cases in India because its large number of migrant laborers returned, giving rise to social distancing challenges in the state.

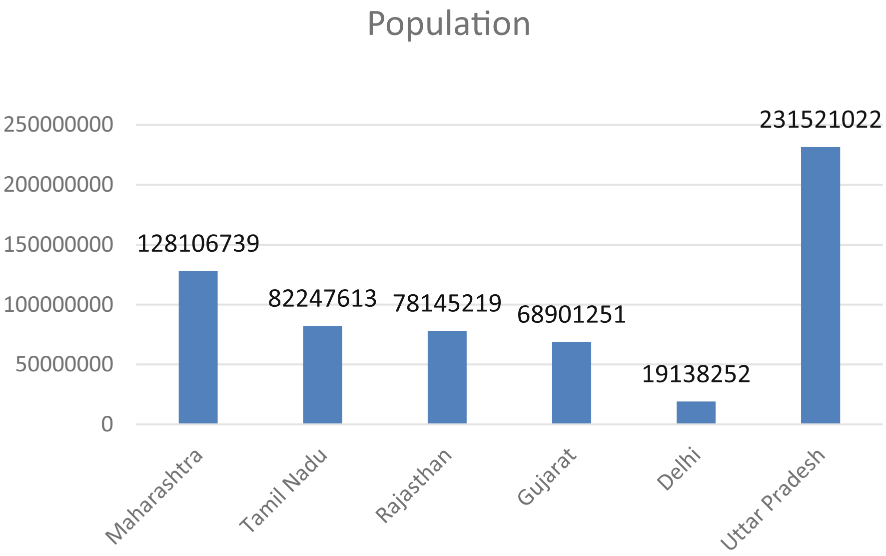


Fig. 4 Population in the six states of India

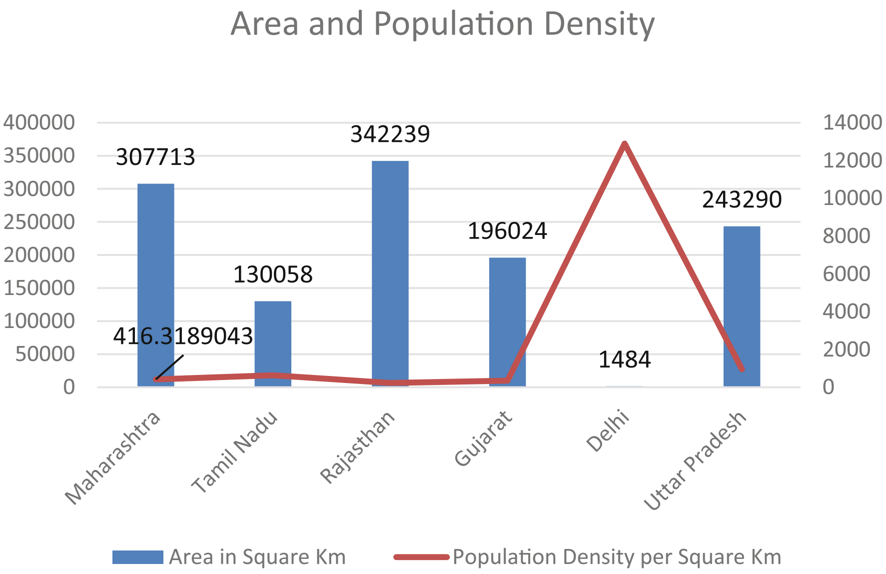


Fig. 5 Area and population density in the six states of India

Reasons and Challenges for Large Number of COVID-19 Cases in Mumbai, Maharashtra

Mumbai, the capital of Maharashtra, is the economic hub of India. So, it has huge number of migrants and has high population density of 32303 persons per square kilometer. There is huge inflow of lower economic class workers from other states in Mumbai. Near about 42% of its population belongs to the slum area. The middle-class and high-class businessmen, traders, and IT professionals from other states and abroad keep visiting Mumbai. So, during the initial phase of COVID-19 scenario, there were social distancing challenges at the airports, railway stations, bus stations, malls, shops, etc. Social distancing norms were not followed at vegetable markets and state transports. The COVID-19 cases increased gradually in Mumbai due to these social distancing issues. Few people from Mumbai were found attending the Tablighi Jamat gathering at Nizamuddin in Delhi. Social distancing is a big challenge in Dharavi which is part of Mumbai and is the biggest slum in Asia. Large number of shops in Dharavi are without toilet facilities. Many households are without proper drainage systems. The overall atmosphere in Dharavi is unhygienic. In Dharavi, the houses are of maximum 10×10 ft with common toilet facilities and no ventilation. About 10–12 people live in a small-sized room. To maintain social distancing in such densely populated slum is a great challenge. Initially, when police tried to maintain social distancing in Dharavi and tried to stop community transmission, people pelted stones at them. Most of the policeman tested positive for COVID-19 due to not maintaining social distancing in congested slums like Dharavi. Police also faced challenges in solving domestic violence cases occurring due to the lockdown, in order to maintain social distancing. The Worli Koliwada is another slum in Mumbai where social distancing is a challenge and most of the COVID-19 cases are found here. Other reasons for COVID-19 spikes in Mumbai include not following social distancing at airports, bus transports, railway stations, vegetable markets, malls, shops etc. during the initial stages when lockdown was not declared. After declaration of the first lockdown in March 2020, proper norms were not followed by the people. This was observed in all subsequent lockdowns in Mumbai. Low testing rate is also one of the reasons for COVID-19 spikes in Mumbai. Cases increased but the tests did not increase at that rate.

Reasons and Challenges for Large Number of COVID-19 Cases in Tamil Nadu

The initial spikes in COVID-19 cases in Tamil Nadu were due to the Tablighi Jamat returnees who tested positive after the gathering at Nizamuddin event in Delhi. About 1500 people attended the meeting out of which 1130 people returned to Tamil Nadu, 515 people were traced, 615 people were untraceable, and 50 people were infected with COVID. Erode from Tamil Nadu was one of the hotspots for COVID due to the Tablighi Jamat gathering. These Tablighi Jamat returnees infected other people, as they travelled by road, train, and air. Opening of liquor shops (TASMAC beer shops) in Tamil Nadu was the reason for COVID-19 spikes during the second week of April 2020. Insufficient social distancing at Koyambedu market was another reason for COVID-19 spikes in Tamil Nadu where huge number of people gathered on April 25, 2020 during lockdown 2.0. About 1867 COVID positive cases were reported by May 9, 2020 [26]. Multiple religious clusters were found during lockdown 3.0 in Tamil Nadu which led to COVID-19 outbreak. Asymptomatic cases encountered in house-to-house-screening in every containment zone and aggressive and extensive testing in Tamil Nadu are the reasons for COVID-19 spikes. To maintain social distancing at Koyambedu market and at the liquor shops was a major challenge for avoiding COVID-19 spread in Tamil Nadu.

Reasons and Challenges for Large Number of COVID-19 Cases in Gujarat

One of the reasons for spikes in COVID-19 cases in Gujarat is that it is a globally connected state from India. So there is huge inflow of people into Gujarat from abroad. Other reasons are the mismanagement over lockdown and laxity in identifying COVID-19 cases. Initially, the major reason for COVID-19 cases was the gathering of 1500 people from Gujarat at Tablighi Jamat Markaz Nizamuddin in Delhi. One more reason is the high-density, low-income population areas in Ahmedabad, Gujarat, where social distancing is challenging task.

Reasons for Large Number of COVID-19 Cases in Delhi

COVID cases initiated with two Chinese nationals tested positive in month of January 2020 in Delhi; 16 Italian tourists and large number of flyers from Delhi were reported at the airport [27]. Despite thermal scanning of passengers at airports, people having high fever could have been missed out. Almost 50% of the initial infected cases in the month of April 2020 were from Tablighi Jamat meeting at Markaz Nizamuddin, Delhi [28]. Markaz is a religious place where preaching takes place. People from different states of India particularly Hyderabad, Maharashtra, Tamil Nadu, Gujarat, and Karnataka attended the Tablighi Jamat meeting at Markaz. On identifying the COVID cases at Markaz, 1584 people were evacuated as on march 31, 2020. Aggressive testing is also the reason for the large number of COVID cases in Delhi. The number of tests per million is put at around 5200 in Delhi (highest in India) [29].

Reasons for Large Number of COVID-19 Cases in Uttar Pradesh

Uttar Pradesh state of India is well known for its highest population of 231, 521,022, that is, 23.15 crores, determined in year 2020. It has the population density of 2100 people per square mile or an average density of 828 persons per square kilometer calculated over an area of 243290 square kilometers or 93930 square miles. Around two lakh migrant workers returned to UP from other states in May 2020, leading to sudden spikes in COVID cases.

Reasons for Large Number of COVID-19 Cases in Rajasthan

There are large asymptomatic cases (around 80%) in a mild form with no outward manifestation of virus like cold, fever, or shortness of breath. The asymptomatic super-spreaders such as daily booth managers, vegetable vendors, milk suppliers, provision store owners, and health workers are the cause for COVID spread in Rajasthan. Foreign (Oman) returned people, lack of social distancing at Tablighi Jamat meeting, and other hidden travel histories [30] are the major reasons for COVID spread in Rajasthan.

4 Use of Technology to Handle Challenges During COVID-19 in India

Robots are used in many hospitals across a number of cities in India to protect the healthcare workers from the COVID-19 infections and intensify the screening process. The robots perform the screening for each and every visitor entering the hospital including the nurses, patients, and doctors, and medical and nonmedical staff. Robots are designed to deliver food and medicines to COVID-19 patients. Some robots are designed to help doctors treat COVID-19 patients remotely using the Internet. The robots first ask questions to the patients related to their body temperature, status of cough, cold, respiratory problems, etc. and depending on the answers, direct the patients to enter the consulting rooms. Other robots in the consulting rooms interact with the doctors using the Internet. Some robots can answer general questions which are asked by the patients like Google Alexa and also can serve water to them. Some robots are used to spread awareness on COVID-19 and display ways to contain the pandemic while some others are used to distribute sanitizers and masks.

Drones which are the unmanned aerial vehicles are used to combat COVID in India. Drones are used for surveillance to ensure that the lockdown is followed and social distancing is maintained. Drones are used to monitor densely populated areas and crack down on lockdown violators. Drones are preferred for sanitization, as they are efficient, fast, and less labor intensive. It also avoids exposing sanitation workers to possible infection. Drones can also be fitted with loudspeakers to issue instructions to the general public. A 28 KG drone with 10 l capacity can disinfect the 500 m $\times$ 500 m area in 10–11 min or 1.5 km–2 km straight line with 10-m periphery to disinfect the whole area. Drones are used in disinfecting government buildings, tall structures, and hospitals almost in 26 cities of India. Some drones have a payload capacity of 15–20 l with a flight duration of 40–45 min and maximum ceiling height of 450 ft. The drones are actually used for spraying from 6.00 am to 6.00 pm every single day. Human beings can cover 3–4 km distance to spread disinfectants, whereas drones can cover 20 km distance every single day. A fleet of 300 drones covers 20 km a day, covering a total distance of 6000 km. The drones also come with thermal imaging which helps to identify and manage crowds and detect body temperature.

In Mumbai city (Maharashtra state), drones are used to monitor Dharavi area which is one of the biggest slums and a containment zone where police cannot enter small streets. The police worked with the members of drones federation to handle this monitoring task. The police use drones to monitor the overall situation during lockdown. The police used the drones to make announcements about the lockdown and used the footages from drones to keep vigil on the lockdown situation. The police also used the drones to spray disinfectants in slums, railway stations, markets, bus terminals, busy streets, and government hospitals which are the most publically used.

In Gujarat, drones are used to make announcements for the norms to be followed during the lockdown. They are also used to spray disinfectants in the most crowded areas of Ahmedabad and in the un-notified slums. Drones are used for surveillance during the lockdown and take appropriate actions from the observed footages.

In Tamil Nadu, the Cuddalore police used drones to curb the illegal liquor sales under the lockdown period. The drones kept vigilance on the villagers of Alagianatham, Chavadi, Commandanmedu, Marundhadu, and Vaanpakkam located near the interstate border of Puducherry from where the tipplers have been sourcing the liquor illegally during the lockdown. In Chennai city (Tamil Nadu state), drones are used to spray disinfectants on hospitals and crowded public places. This helped to avoid the spread of COVID and reduced the work burden on the frontline workers: doctors, nurses, sanitization workers, and policeman during the fight against COVID-19.

Technological Challenges During COVID-19

Heavy Load on Internet

Due to the wide spread of COVID-19, initially, a complete lockdown was declared at the national level giving rise to new challenges in management of IT, connectivity and security. Millions and millions of people started creating heavy Internet traffic.

Online workers, Information Technology (IT) professionals, doctors, school and college teachers and students, private coaching classes, YouTubers and many other Internet Protocol-based APP users started using the Internet on a larger scale. This became the cause of trouble for high-technology companies to offer online gaming, video streaming services, and offline remote conferencing services, etc.

Information Technology

The Information Technology (IT) professionals faced lot of problems to work from home during the COVID-19 lockdown period. Reputed companies usually provide the infrastructure required by the professionals to work from home (WFH), but not all employees have it at sufficient level. In most of the companies, freshers are not given the facilities to work from home. Some companies provide high-end laptops loaded with required software and the broadband dongles to work from home. The employees in such cases can use the set Virtual Private Networks to complete the work from home. But all these facilities are not made available to all employees by IT Company nor have all IT companies made such prior arrangements giving rise to too many challenges to work from home during the lockdown periods. It is challenging to work with very old laptops, computers, poor connectivity, and outdated and unsupported software. Some professionals used tablets with lack of security patches to complete the tasks. The WFH includes the use of equipment which is shared with family members. This leads to challenge in securing company's data. WFH also leads to Wi-Fi connectivity issues, as the bandwidth is shared among multiple devices at home. Many employees are not familiar with the use of computing devices, connecting to core systems, using videoconferencing systems collectively and securely, leading to massive requirements in user training; IT professional lost lot of their fruitful time to set requests for Virtual Private Networks (VPNs), setting up laptops, and training people for the proper and secure use of the resources instead of working on new things. The remote work involves the use of VPNs, VoIP (Voice over IP) virtual meetings, cloud technology, work collaboration tools, facial recognition technologies, and many more. Another challenge for IT professionals is to provide sufficient cloud storage to tackle the dramatic increase in online orders. Increase in online purchases posed a new challenge of saving jobs of shopkeepers and vegetable vendors. Online shopping during COVID-19 situation poses a new set of challenges. In online shopping, there can be in-person delivery of goods which is not virus-free and needs proper sanitization. Another option for delivery of goods is the use of robots which can ensure sufficient level of safety against the coronavirus pandemic. Remote work for IT professionals provides advantages of avoiding COVID-19 spread, saving commute time, and increasing work flexibility, but on the other hand, it leads to lack of work–life balance and loneliness.

Distance Teaching-Learning

Distance teaching-learning is a big challenge in COVID-19 situation. At the initial spread of coronavirus in India, the teaching-learning was in progress by adopting the usual traditional ways in all the educational Institutes except for the online courses offered by IIT like National Program on Technology Enhancement and Learning (NPTEL). As the COVID spread reached its first spike in mid of March 2020, all of a sudden, the educational institutes were brought to closure due to the national-level lockdown. In order to complete the syllabus, various educational institutes including schools, colleges, and coaching classes had to adopt the online teaching learning environments. It was a big challenge for many schools to conduct the year-end examinations. Some schools that had already completed the year end exams before the lockdown could manage to declare the results online. But the schools which were lagging behind had a great challenge ahead for completing the syllabus, conducting the exams, and declaring the results. Many schools promoted the students to the next higher class failing to adopt the sudden change in the system and requirements. Many professional educational institutes and universities offering undergraduate, post graduate, doctoral, and post-doctoral courses had to adopt to the distance teaching learning techniques to complete the syllabus. In this process, the faculty and students faced lot of challenges with respect to the use of online tools and software. The immediate and heavy demand of online platforms and tools for educational purposes posed a challenge ahead of big IT companies. Google provided free use of Google meet for stipulated time, Cisco allowed free use of Cisco Webex platform, and Microsoft offered the use of Microsoft teams. Many use Zoom platform for conducting meetings and conducting lectures in spite of its security issues and use for limited time and for limited number of users.

Rural Area Distance Education Problems

- Lack of smart phones, laptops, computers, and tablets
- No Wi-Fi connectivity
- Frequent power cuts
- Difficulty-to-use technology
- No proper earphones, mikes, and speakers
- No silent sitting place for attending online class

Urban Area Distance Education Problems

Pre-primary school children are unable to understand use of technology, and every time, parents' assistance is required. If both parents are working, then it becomes problematic for them, as they have to take time out of their office work. Students are unable to concentrate on the PowerPoint presentations due to distractions at home. Students cannot cope with the online teaching speed. Many times, there exists a time

Table 1 Impact of COVID on frontline workers

State	Frontline workers infections	Date
Maharashtra	42% of infected doctors, 70.5% of infected nurses and 84% of infected medical workers	April 23, 2020 [31]
	At least 500 doctors infected as per Indian Medical Association (IMA)	June 05, 2020 [32]
Delhi	106 total employees of hospital were tested positive including doctors and nurses.	May 12, 2020 [33]
	Delhi reported 69 infected doctors from 548 total in India	June 05, 2020 [34]
Chennai City (Tamil Nadu state)	At least 37 doctors were infected.	May 10,2020 [35]
	112 health workers (including doctors, nurses and emergency management technicians, hospital staff, and ambulance drivers)	June 05, 2020 [34]
	18 nurses and 57 paramedics	

delay between the content delivery and the moment of listening. Some delay due to poor Internet connections. When the teacher projects the PowerPoint presentations while online teaching, the students are unable to understand the teachers' power of expressions nor can the teacher read students face to get real feedback on teaching-learning. Another issue in online teaching is that the teachers are really unaware whether the students are listening to the content delivery or not. Teacher cannot continuously monitor the video screens of all students. In an online class, let us say, of 45 min, the teacher is hardly able to monitor the video screens for 2 to 3 times at the maximum, otherwise the teaching content is left aside.

Impact of COVID on Frontline Workers in India

Higher number of COVID infections are found among the frontline workers, which include cops, doctors, nurses, other medical staff (ward boys), and cooks in hospitals, sanitization workers, vegetable vendors, and pharmacy. A total of 548 doctors, nurses, and paramedics were infected as on May 6, 2020 across India [36] including ward boys, field workers, security guards, sanitization workers, lab attendants, laundry, peons, and kitchen staff (Table 1).

COVID Cases Forecasting

The forecasting for COVID cases is performed for India and its highly affected states from the Kaggles dataset [37] by considering the data up to June 16, 2020 as seen in Figs. 6, 7, 8, 9, 10, 11, and 12. The Auto Regressive Integrated Moving Average (ARIMA) model [38] in Python is used for predicting the COVID cases till end of July 2020 and till the end of October 2020 as seen in Table 2.

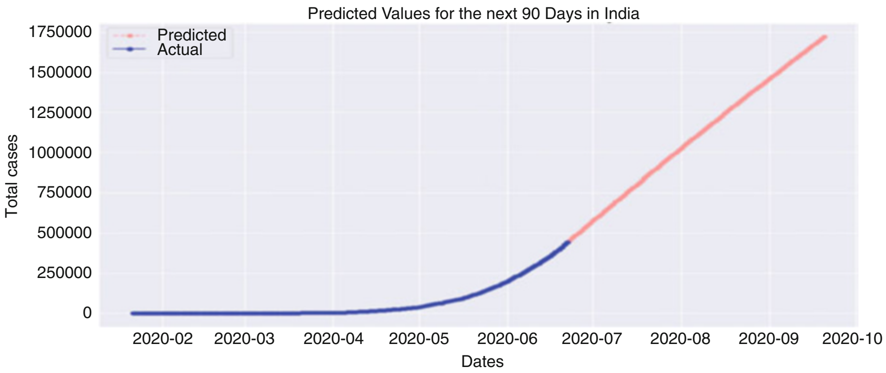


Fig. 6 Forecasting for COVID cases in India by October 2020

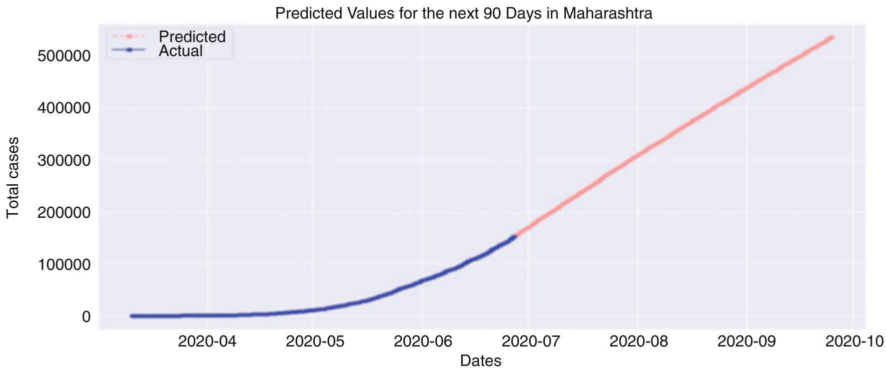


Fig. 7 Forecasting for COVID cases in Maharashtra by October 2020

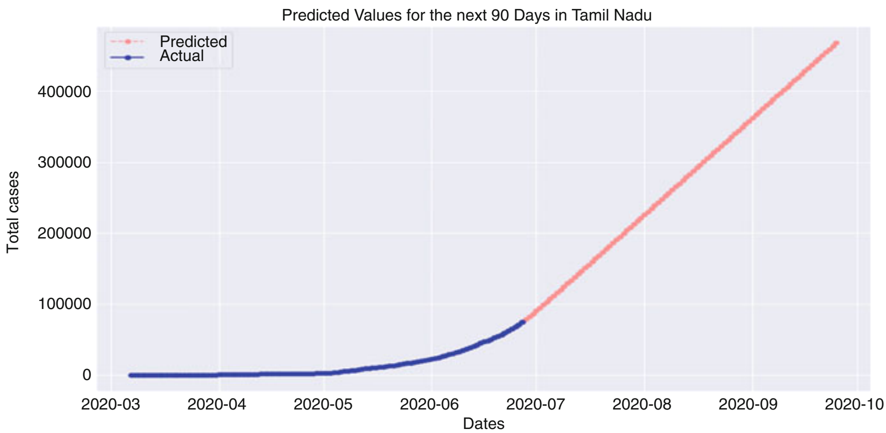


Fig. 8 Forecasting for COVID cases in Tamil Nadu by October 2020

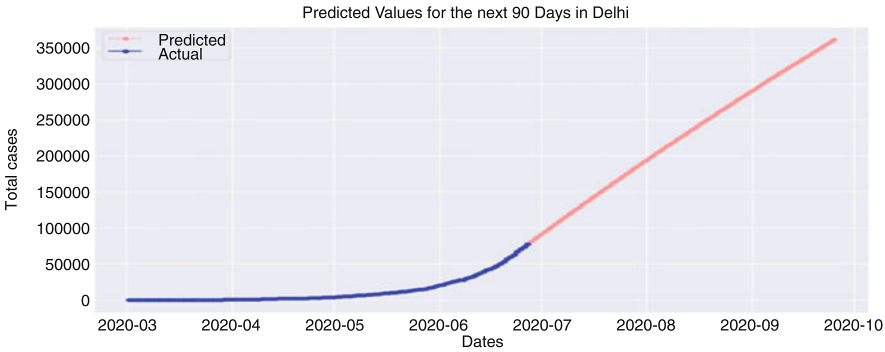


Fig. 9 Forecasting for COVID cases in Delhi

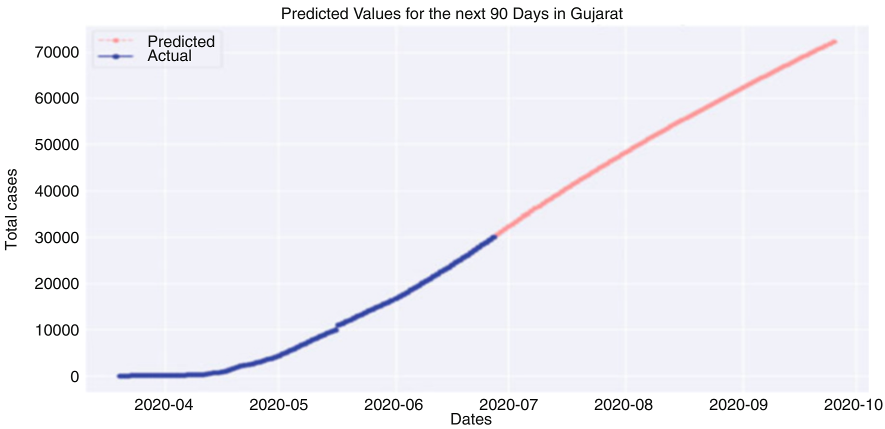


Fig. 10 Forecasting for COVID cases in Gujarat by October 2020

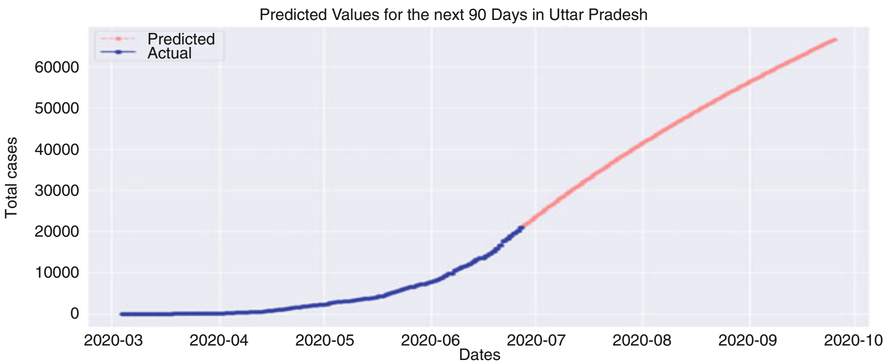


Fig. 11 Forecasting for COVID cases in Uttar Pradesh by October 2020

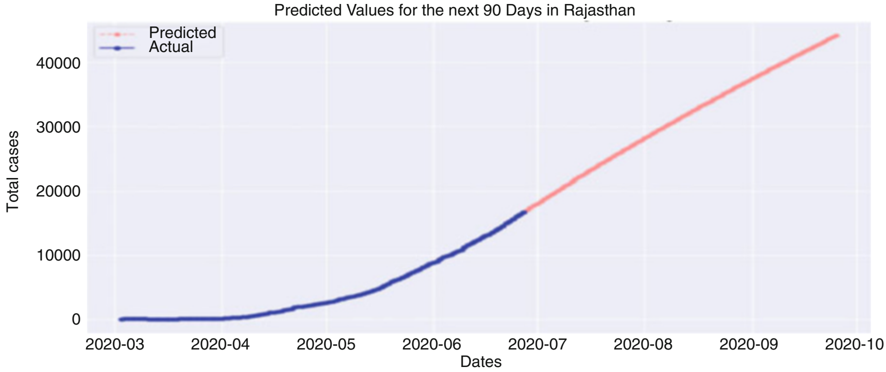


Fig. 12 Forecasting for COVID cases in Rajasthan by October 2020

Table 2 Forecasting for COVID cases based on data available till June 16, 2020

Place	Predicted COVID cases July 15, 2020	COVID cases as on July 22, 2020	Forecasting for COVID cases by end of October 2020
India	80,000	11,90,000	17,50,000
Maharashtra	2,00,000	3,19,000	5,00,000
Tamil Nadu	1,00,000	1,76,000	4,00,000
Delhi	1,00,000	1,24,000	3,50,000
Gujarat	35,000	50,465	70,000
Rajasthan	25,000	30,390	70,000
Uttar Pradesh	25,000	51,160	66,000

5 Architectural Design and Planning Challenges Related to COVID-19

Types of infrastructure primarily affected due to COVID-19 are:

1. Hospital/healthcare infrastructure
2. Public gathering spaces like cinema halls, malls, religious places, institutes, etc.
3. Work places like offices, industries, factories, etc.
4. Transportation infrastructure:
   (a) Airplane, bus, cabs, trains
   (b) Airports and railway stations
5. Hotels and restaurants
6. Tourism places

Following are the suggestions which can be adopted / incorporated:
General recommendations:

- To have separate entry and exit.
- Each entry point should facilitate proper sanitization of people.
- Thermal scanners should be installed at entry points.
- Touch-me-not type of infrastructure like doors with automatic opening / closing sensors, voice-operated elevators
- Wider stairs/ escalators to minimize use of railing contact in public buildings.

Planning Aspects for Slums Related to COVID-19

The informal settlements or slum areas of the India mostly in Mumbai and Pune in Maharashtra are the least prepared for the pandemic of COVID-19 since basic needs such as water supply, toilets, drainage, and waste collection are to be shared and inadequate. There is shortage of adequate and secure housing. Inadequate space causes overcrowding in slums which is responsible for inability for physical distancing. Due to all these factors, it is practically impossible to advice self- or home quarantine. These infrastructure challenges in urban slums cause rapid spread of an infection [39].

Scientific policy suggestions have to be made to dampen the spread of COVID-19, to improve availability of medical care to urban poor irrespective of the infection status. To improve long-term well-being of urban slums, provision of social and physical improvements in infrastructure is needed which comes with better economy.

Immediate measures to protect urban poor from COVID-19 include:

1. To improve social distancing to set up slum emergency planning committee
2. Ensure payment to poor and provision of things of daily living like food
3. Training and deployment of community health workers
4. Improvement in sanitation and hygiene
5. Implementation of a solid waste collection strategy

Planning Aspects for Buildings Related to COVID-19

Organisms get transmitted between rooms on different floors of a building, carried within the system airflow. Contamination of surfaces in rooms and systems by droplets results in spread of infection. This is of particular concern in high-risk transmission settings such as hospitals and healthcare buildings. One important factor identified was the interconnectedness of all parts of the building by the wastewater plumbing system and, therefore, the potential for contaminated air to travel throughout the building unhindered. Also it is identified the short-duration burst of contaminated air from the wastewater plumbing system that caused the cross-contamination.

It is recommend that the following steps be taken to ensure that transmission through the wastewater plumbing system is minimized: (1) do not ignore unexplained foul smells in bathrooms, kitchens, or wash regions; (2) It is good if water appliances in kitchens and bathrooms are fitted with a functioning U-bend; (3) to prevent the loss of the water trap seal within a U-bend, open a tap on all water appliances for at least 5 s twice a day (morning and evening), paying special attention to floor drains in bathrooms and wet rooms; (4) in case wastewater pipe from an appliance seems to be damaged or open, it should be sealed immediately (5) if there appears to be any crack or leak in pipework, seal with tape or glue; and (6) continuous monitoring of system performance is needed especially for large or tall buildings [40].

6 Conclusion

With all the challenges like vast population, limited healthcare infrastructure, limited healthcare workers, India has achieved a better control over COVID-19 as seen from the current and forecasted statistics. India has improved the cure rate by effective strategies. Still, India has to go a long way to achieve an effective control on the spread of COVID-19.

Acknowledgments The authors would like to offer thanks to the anonymous editors and reviewers of this manuscript for spending their valuable time and providing valuable reviews and insightful remarks.

References

1. Coronavirus update live. 1 July 2020. Retrieved from https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/
2. Detail Question and Answers on COVID-19 for Public. 17 March 2020. Retrieved from https://www.mohfw.gov.in/pdf/FAQ.pdf
3. New red flags backache nausea rashes may also be covid warning signs. 1 July 2020. Retrieved from https://economictimes.indiatimes.com/industry/healthcare/biotech/healthcare/new-red-flags-backache-nausea-rashes-may-also-be-covid-warning-signs/articleshow/76722174.cms
4. COVID 19 statewise status. 1 July 2020. Retrieved from https://www.mygov.in/corona-data/covid19-statewise-status/
5. India's COVID-19 recovery rate improving, fatality rate declining: Health Ministry. 2 June 2020. Retrieved from https://economictimes.indiatimes.com/news/politics-and-nation/indias-covid-19-recovery-rate-improving-fatality-rate-declining-health-ministry/articleshow/76140228.cms?from=mdr
6. India far away from peak much better positioned in covid -19 fight than other nations: Govt. 2 June 2020. Retrieved from https://economictimes.indiatimes.com/industry/healthcare/biotech/healthcare/india-far-away-from-peak-much-better-positioned-in-covid-19-fight-than-other-nations-govt/articleshow/76156812.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst

7. P. Mehra, in *India's Economy Needs Big Dose of Health Spending* (2020 April 8). Retrieved from https://www.livemint.com/news/india/india-s-economy-needs-big-dose-of-health-spending-11586365603651.html
8. P. Singh, S. Ravi, S. Chakraborty, in *COVID-19-Is India's Health Infrastructure Equipped to Handle an Epidemic?* (2020 March 24). Retrieved from https://www.brookings.edu/blog/up-front/2020/03/24/is-indias-health-infrastructure-equipped-to-handle-an-epidemic/
9. Hospital beds (Per 1000 people) India, 2011. Retrieved from https://data.worldbank.org/indicator/SH.MED.BEDS.ZS?locations=IN
10. Lack of medical investment, healthcare infra big challenges for India's COVID-19 fight: Fitch. 14 May 2020. Retrieved from https://economictimes.indiatimes.com/industry/healthcare/biotech/healthcare/lack-of-medical-investment-healthcare-infra-big-challenges-for-indias-covid-19-fight-fitch/articleshow/75731604.cms?from=mdr
11. Guidance document on appropriate management of suspect/confirmed cases of COVID-19. 7 April 2020. Retrieved from https://www.mohfw.gov.in/pdf/FinalGuidanceonMangaementofCovidcasesversion2.pdf
12. Govt of India, Ministry of home affairs circular dated 1 May 2020. Retrieved from https://www.mha.gov.in/sites/default/files/MHA%20Order%20Dt.%201.5.2020%20to%20extend%20Lockdown%20period%20for%202%20weeks%20w.e.f.%204.5.2020%20with%20new%20guidelines.pdf
13. India far away from peak, much better positioned in COVID-19 fight than other nations: Govt. 2 June 2020. Retrieved from https://www.indiatvnews.com/news/india/india-far-away-from-peak-much-better-positioned-in-covid-19-fight-than-other-nations-govt-622808
14. A. Roy, D. Ghoshal, A. Kalra, in *Mumbai Runs Out of Hospital Beds as COVID-19 Cases Continue to Rise* (2020 May 25). Retrieved from https://science.thewire.in/health/mumbai-coronavirus-hospital-beds-staff-shortage/
15. S. Suryawanshi, in *Maharashtra to Offer 100% Free Treatment for COVID-19 Patients in Private Hospitals Too* (2020 May 1). Retrieved from https://www.newindianexpress.com/nation/2020/may/01/maharashtra-to-offer-100-free-treatment-for-covid-19-patients-in-private-hospitals-too-2137991.html
16. A. Laskar, in *Maharashtra Creates Jumbo Medical Centres as Covid 19-Count Crosses-50000 Mark* (2020 May 25). Retrieved from https://www.livemint.com/news/india/maharashtra-creates-jumbo-medical-centres-as-covid-19-count-crosses-50-000-mark-11590425578800.html
17. D.S. Jesudasan, in *T.N. Government Begins Identifying Places to Set Up COVID-19 Care Centres* (2020 June 15). Retrieved from https://www.thehindu.com/news/national/tamil-nadu/tn-govt-begins-identifying-places-to-set-up-covid-19-care-centres/article31829206.ece
18. TN to cover COVID-19 treatment in private hospitals under CM's health insurance scheme, 4 June 2020, retrieved from https://www.newindianexpress.com/states/tamil-nadu/2020/jun/04/tn-to-cover-covid-19-treatment-in-private-hospitals-under-cms-health-insurance-scheme-2152269.html
19. Reserve 50% Beds Of All Private Hospitals Beds for COVID-19 Patients: Gujarat HC Tells State. 28 May 2020. Retrieved from https://medicaldialogues.in/state-news/gujarat/reserve-50-beds-of-all-private-hospitals-beds-for-covid-19-patients-gujarat-hc-tells-state-66130
20. D. Mukherjee, in *Explained: The 'Bhilwara Model' of 'Ruthless Containment' to Stop Coronavirus*. 11 April 2020. Retrieved from https://indianexpress.com/article/explained/explained-bhilwara-model-ruthless-containment-stop-coronavirus-6350395/
21. A. Sharma, in *After Bhilwara Success, Rajasthan Enlists Half of Hospitals for Dealing with COVID from Private Sector* (2020 April 8). Retrieved from https://economictimes.indiatimes.com/news/politics-and-nation/after-bhilwara-success-rajasthan-enlists-half-of-hospitals-for-dealing-with-covid-from-private-sector/articleshow/75039118.cms?from=mdr
22. A. Sengupta, in *Government Launches Aarogya Setu COVID-19 Tracker App on Android, iOS* (2020 April 2). Retrieved from https://gadgets.ndtv.com/apps/news/aarogya-setu-covid-19-tracker-app-coronavirus-launch-indian-government-android-ios-2204804

23. A. Banerjea, in *Covid-19: Lockdown Extended by Two Weeks with Effect from 4 May, says MHA* (2020 May 1). Retrieved from https://www.livemint.com/news/india/covid-19-lockdown-extended-by-two-weeks-with-effect-from-4-may-says-mha-11588337761963.html
24. India corona count, 2020, May 31. Retrieved from https://www.google.com.
25. State census 2011, 2020, June 06. Retrieved from https://www.census2011.co.in/states.php
26. Coronavirus Live India 2020, May 09. Retrieved from https://www.nationalheraldindia.com/india/coronavirus-live-updates-covid-19-india-wuhan-us-italy-who-alert-novel-coronavirus-latest-news-09th-may
27. The Tribune, in 14 Italians Isolated, 2020, March 05. Retrieved from https://www.tribuneindia.com/news/nation/delayed-restriction-spurts-virus-cases-50905
28. Coronavirus in India: Tablighi Jamaat Meet Turns Covid-19 super spreader, 2020, April 1. Retrieved from https://www.indiatoday.in/mail-today/story/coronavirus-in-india-tablighi-jamaat-meet-turns-covid-19-super-spreader-1661958-2020-04-01
29. Coronavirus outbreak, 2020, May 13. Retrieved from https://www.hindustantimes.com/delhi-news/in-50-days-of-lockdown-number-of-covid-19-cases-in-delhi-increase-254-times-report/story-kVAhn9kVIp2BcLpyM4XhnO.html
30. Tablighi Jamaat's mass gatherings across Asia spark coronavirus clusters, 2020, April 21. Retrieved from https://timesofindia.indiatimes.com/articleshow/75264112.cms
31. Data | How many doctors and nurses have tested positive for coronavirus in India?, 2020, April, 23. Retrieved from https://www.thehindu.com/data/how-many-doctors-and-nurses-have-tested-positive-for-coronavirus-in-india/article31410464.ece
32. At least 500 doctors infected in state, says IMA, 2020, June 04, retrieved from https://indianexpress.com/article/india/at-least-500-doctors-infected-in-state-over-140-in-mumbai-6441533/
33. Coronavirus | From care to crisis, how a Delhi government hospital lost the plot, 2020, May 12. Retrieved from https://www.thehindu.com/news/cities/Delhi/from-care-to-crisis-how-a-delhi-government-hospital-lost-the-plot/article31561548.ece
34. Corona cases in India rise at alarming rate after BSF, police personnel test positive, 2020, May 08. Retrieved from https://www.deccanchronicle.com/nation/current-affairs/080520/corona-cases-in-india-rise-at-alarming-rate-after-bsf-police-personne.html
35. Tamil Nadu: Doctors and cops fall as Covid-19 breaches frontline resistance, 2020, May 10. Retrieved from https://timesofindia.indiatimes.com/city/chennai/tamil-nadu-doctors-and-cops-fall-as-covid-19-breaches-frontline-resistance/articleshow/75656659.cms
36. Coronavirus Outbreak, 2020, May 10. Retrieved from https://www.hindustantimes.com/india-news/548-docs-nurses-paramedics-infected-with-covid-19-across-india-report/story-o2pM3w2adM4g3PXI6TBlkN.html
37. Dataset on Novel Corona Virus Disease 2019 in India. https://www.kaggle.com/sudalairajkumar/covid19-in-india.
38. S. Makridakis, M. Hibon, ARMA models and the Box–Jenkins methodology. J. Forecast. **16**(3), 147–163 (1997)
39. J. Corburn, D. Vlahov, B. Mberu, L. Riley, W.T. Caiaffa, S.F. Rashid, A. Ko, et al., Slum health: Arresting COVID-19 and improving well-being in urban informal settlements. J. Urban Health **97**, 1–10 (2020)
40. M. Gormley, T.J. Aspray, D.A. Kelly, COVID-19: Mitigating transmission via wastewater plumbing systems. Lancet Glob. Health **8**(5), e643 (2020)

Index

S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1