The Springer Series in Applied Machine Learning

Manju
Sandeep Kumar
Sardar M. N. Islam *Editors*

Artificial Intelligence-based Healthcare Systems

Springer

The Springer Series in Applied Machine Learning

Series Editor: Oge Marques

The Springer Series in Applied Machine Learning focuses on monographs, textbooks, edited volumes, and reference books that provide suitable content and educate the reader on how the theoretical approaches, algorithms, and techniques of machine learning can be applied to address real-world problems in a principled way.

The goal is a series of state-of-the-art books for the libraries of both machine learning scientists interested in applying the latest techniques to novel problems as well as practitioners, technologists, and researchers in various fields interested in leveraging the latest advances in machine learning for developing solutions that work for practical problems. The scope spans the breadth of machine learning and AI as it pertains to all application areas — both through books that address techniques specific to one application domain — and books that show the applicability of different types of machine learning methods to a wide array of domains.

Manju • Sandeep Kumar • Sardar M. N. Islam
Editors

Artificial Intelligence-based Healthcare Systems

Springer

Editors
Manju
Department of Computer Science &
Engineering and Information Technology
Jaypee Institute of Information Technology
Noida, Uttar Pradesh, India

Sandeep Kumar
School of Engineering & Technology
Christ (Deemed to be University)
Bengaluru, Karnataka, India

Sardar M. N. Islam
Victoria University
Melbourne, VIC, Australia

ISSN 2520-1298 ISSN 2520-1301 (electronic)
The Springer Series in Applied Machine Learning
ISBN 978-3-031-41924-9 ISBN 978-3-031-41925-6 (eBook)
https://doi.org/10.1007/978-3-031-41925-6

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Paper in this product is recyclable.

Preface

In the era of this modern age, there is lack of high-quality hospitals and medical experts especially in the remote areas of developing countries. To provide an effective healthcare on time is important to all the living being, and this becomes critical necessity when it comes to rural areas population suffering with chronic disease. Nowadays, the modern concept is to observe the health-related issues with the help of smart technology. This modern system consists of plethora of technologies which includes portable remote health system, wearable sensors, wireless communications and many more. In order to utilize these smart devices, the concept of trending modern technologies such as artificial intelligence, Internet of Things and machine learning are on high demand. Since, artificial intelligence has put tremendous impact on vivid fields of life, health sector too gets benefited from it. The conventional telemedicine is facing many issues due to the existing architecture which is based on store-and-forward concept. Some critical issues are smaller number of local health centers, lack of dedicated staff, lack of manpower to prepare patient reports, time critical diagnosis from the concerned medical expert, and need for uninterrupted Wi-Fi connection. Nowadays, the smart techniques are automating this whole process based on artificial intelligent tools and techniques with smart IoT devices which assists while doing this complete process.

Noida, Uttar Pradesh, India
Bengaluru, Karnataka, India
Melbourne, VIC, Australia

Manju
Sandeep Kumar
Sardar M. N. Islam

Contents

Schedule and Routing in Home Healthcare System Using Clustering Analysis and Multi-objective Optimization

Tarık Küçükdeniz and Beril Şevval Develioğlu

Abstract Home healthcare (HHC) services, delivered by healthcare professionals at patients' homes, have seen a growing demand due to the increasing elderly population. The COVID-19 pandemic has highlighted the significance of medical homes, particularly for high-risk individuals. Implementation experiences show that team planning in HHC services enhances staff satisfaction by promoting balanced workload distribution, while effective planning reduces patient access times and improves overall system efficiency.

This study tackles a multi-objective HHC problem aimed at achieving balanced workloads for HHC teams, minimizing daily travel distances, and reducing appointment delays. By leveraging 2-day patient location data from a state hospital's HHC unit, the problem was addressed using clustering analysis and genetic algorithm.

1 Introduction

Home healthcare (HHC) is the examination, treatment, and rehabilitation services offered at home by professional health teams to patients who have difficulties reaching health institutions due to various diseases [1]. HHC is an alternative solution that is applied in response to the increasing hospital demand around the world. The main goal of such services is to reduce costs by shortening the duration of hospital stays [2–4].

According to the data announced by the World Health Organization (WHO), 60% of health expenditures are due to the treatment of chronic diseases [5]. The increase in chronic diseases, along with the aging population, causes an increase in

---

T. Küçükdeniz (✉) · B. Ş. Develioğlu
Faculty of Engineering, Department of Industrial Engineering,
İstanbul University-Cerrahpaşa, Istanbul, Turkiye
e-mail: tkdeniz@iuc.edu.tr

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_1

health expenditures as well. This shows that the need for home healthcare services will increase in the coming years, and more importance should be given [6].

Home healthcare does not have an internationally determined list of services. Therefore, each country determines a list of services that it can periodically update within its own possibilities and sociocultural structure [7]. Home healthcare can be offered as protective-preventive, supportive-encouraging, therapeutic-curative, rehabilitative, acute-short term, chronic-long term, palliative-mitigating, and terminal care [8].

According to the direction published by the Ministry of Health in 2015, home healthcare is provided by health personnel in line with the diagnosis and planned treatment made by specialist doctors to individuals who have difficulties in reaching the hospital due to their illness and/or old age, which severely impair their quality of life. It includes examination, analysis, treatment, medical care, follow-up, and rehabilitation services provided together with social and psychological counseling services. The basic services provided by the health personnel can be counted as general examination, physiotherapy, prescribing and supplying drugs, providing medical equipment and devices, training the patient and their relatives about the medical home process, patient's illness and the drugs he/she uses, and the use of medical equipment and devices [9].

Clustering the patients to be visited, assigning the patient clusters to the health teams, and determining the routes of the teams are the most important steps that affect the efficiency of the HHC process and handled as the scheduling and routing problem in home healthcare (HHCSRP). HHCSRP is handled in different decision-making levels in HHC process.

In home healthcare, planning is done at different decision-making levels in relation to the length of the planning period: strategic planning, tactical planning, and operational planning. The assignment of patient clusters to healthcare personnel by dividing patients into regions is done at strategic level planning and is related to zoning, classification, and assignment problems in operations research. Determining the number of personnel required for home health services and assigning personnel to regions are usually done in tactical planning and are related to operations research problems of boxing, resource sizing, and assignment. Assigning healthcare personnel to patients and designing routes for each healthcare personnel are done in operational level planning, and these two problems related to scheduling and vehicle routing problems in operations research can be solved simultaneously or independently of each other.

The scheduling and routing problem (HHCSRP) in home health services is generally defined as "determining which healthcare team will serve the patients who need to be treated at home in which order" or "the assignment of the patients located in a certain geographical region to the healthcare teams and providing the health services requested by the teams in the appropriate order" [10, 11].

In this study, clustering and genetic algorithm are proposed to solve a multipurpose HHCSRP which aims to provide a balanced workload distribution to HHC teams, to minimize the total daily travel distances of the teams and the total delays to patient appointments, taking into account the lunch breaks.

In the study, taking into account the home healthcare process of a state hospital, the solution was applied on a sample problem by utilizing the 2-day patient locations that the HHC unit would visit. Daily patient clusters are created by the hospital through an appointment system. Therefore, it is clear on which day the patients who will benefit from HHC will be visited. However, it is not definite which team will visit in which order.

2 Studies in the Literature

In the literature, many different heuristic and metaheuristic algorithms have been used for the HHCSRP problem, such as the branch-price-and-cut algorithm [12], heuristic algorithm based on the savings algorithm (Clarke-Wright algorithm) [13], and heuristic algorithm based on partitional clustering and large neighborhood search [14]. Table 1 summarizes some research from the literature, along with planning periods and solution methodologies.

It is seen in the literature that population-based [15, 16] and metaheuristic methods based on local search [17–19] are suggested for the short planning period. For exact solution methods, branch-price-and-cut algorithms are mainly used [20, 21].

According to Fikar and Hirsch [22], an HHCSRP considered for short-term planning includes an average of 20 personnel and 150 visits. The solution time varies from a few seconds to 75 min per sample, depending on the size of the problem, the programming language, and the method used.

3 Problem Definition

Scheduling and routing in home healthcare problems (HHCSRP) can be considered as a combination and an extension with some limitations that make it difficult to solve the problems of nurse scheduling and vehicle routing problems (VRP), which are very common in the literature, as it consists of designing a series of routes to be used by teams providing health services by assigning individuals living in the same geographical region and needing home treatment to the health teams [23, 24].

The aim of HHCSRP is to minimize or maximize a criterion determined according to some constraints for different planning periods. In the studies in the literature, classical scheduling and vehicle routing models were used in the modeling of the HHCSRP. However, they are modeled by using different constraint and objective functions from the classical scheduling and vehicle routing models in the resources due to the characteristics of the problem addressed, such as the presence of personnel with different competencies in the healthcare team and the precedence relationships between services. The HHCSRP literature is rich in model diversity depending on the different characteristics of the problems and the model selections related to them [23].

Table 1 Studies and solution methods in the literature on HHCSRP

Author	Planning period	Solution method
Fikar and Hirsch [22]	Short	Metaheuristic
Hiermann et al. [39]	Short	Hyper-heuristic
Cordeau et al. [40]	Short	Metaheuristic (Tabu search)
Bertels and Fahle [30]	Short	Hybrid approach: Two metaheuristic methods (simulated annealing and Tabu search)
Eveborn et al. [41]	Short	Heuristic: Iterative matching with nearest neighbor
Akjiratikarl et al. [15]	Short	Metaheuristic (particle swarm optimization)
Bennett and Erera [42]	Long	Heuristic (greedy search)
Rest and Hirsch [18]	Short	Tabu search
Rasmussen et al. [20]	Short	Branch-price-and-cut algorithm
Lanzarone et al. [29]	Short	Mathematical
Gayraud et al. [43]	Short	Integer linear programming and metaheuristic
Allaoua et al. [27]	Short	Mixed-integer programming and metaheuristic (simulated annealing)
Bowers et al. [44]	Short	Clarke-Wright
Mankowska et al. [17]	Short	Mixed-integer programming and heuristic
Nikzad et al. [45]	Short	Metaheuristic
Begur et al. [13]	Long	Heuristic (Clarke-Wright)
Cheng and Rich [4]	Short	Heuristic
Bard et al. [12]	Long	Two mixed-integer and three heuristic algorithms
Wirnitzer et al. [46]	Long	Mixed-integer programming
Breakers et al. [19]	Short	Metaheuristic algorithm that embeds neighborhood search heuristic in local search framework
Yalçındağ et al. [16]	Long	Two-phase decomposition method
Carello et al. [47]	Long	Integer linear programming
An et al. [26]	Long	Heuristic
Xiao et al. [48]	Short	Mixed-integer programming
Taş et al. [49]	Short	Integer linear programming
Cappanera and Scutellà [50]	Long	Metaheuristic
Erdem and Koç [51]	Long	Hybrid metaheuristic, combination of genetic and variable neighborhood algorithm
Decerle et al. [52]	Short	Simulated annealing and tabu search based approach
Şimşek [24]	Long	Heuristic
Grenouilleau et al. [14]	Long	Partitional clustering and large neighborhood search
Tanoumand et al. [21]	Short	Branch-price-and-cut algorithm

In the HHCSRP, there are three main stakeholders that are affected by medical home in different ways: the health institution, the patient, and the health personnel. The different constraints of each stakeholder bring different characteristics to the problem. Cissé et al. [23] classified the constraints of these stakeholders into three groups in their extensive literature review: time constraints, assignment constraints, and positional constraints. Classification according to the constraints in their study is given in Table 2.

In their study, Holm and Angelsen [25] found that 18–26% of the HHC process is travel time. This large percentage appears to be the main focus of travel in the sources. In the models developed in the literature, cost minimization is aimed by minimizing the total travel distance/time traveled [26]. Another common aim in the studies is to balance the workload among the teams, to minimize the working hours/overtime of the personnel, the number of personnel, or the personnel cost [27, 28]. In addition, some studies have assumed that delay time has a cost. Minimizing the delay time, waiting times, dissatisfaction, and the penalties arising from these is among the aims [17]. In addition, ensuring the continuity of care, which means assigning a minimum number of different personnel to a particular patient during a planning period, is one of the objectives of interest for a good quality of service [29]. In the literature, single-objective and multipurpose models have been proposed for the solution of HHCSRP [15]. There are two different alternative solution methods in multi-objective optimization problems that contain more than one objective function. The first and most widely used one is to make the objectives into a single-objective function by giving weight (coefficient) even if they are not in the same units [17, 30]. The other is to solve the objectives step by step. Each solved objective is added as a constraint to the other solutions, and the process continues. The most common method for solving multi-objective models in the HHCSRP literature is the weighting method, which is the first method. Only a few studies have adopted a multiple constraints approach [19].

HHCSRP consists of two NP-hard problems: scheduling and routing. In the literature, there are many studies in which HHCSRP is solved by a mathematical method. While mathematical methods scan the whole solution set to find the solution, heuristic methods aim to reach the best or near-best solution by approaching the solution set intuitively. For this reason, for problems with a large solution set,

Table 2 Classification scheme with constraints [23]

Stakeholders	Health institution	Patient	Health personnel
Time constraints:	Planning period Frequency of decision repetition	Frequency of visits Time windows Temporal relationship between services Discrete services	Contract type Capacity/working hours
Assignment constraints:	Continuity of the care	Preferences	Proficiency/skill Workload balancing
Positional constraints:	Zoning service type	Network type between home locations	Location of the personnel

the solution duration of mathematical methods takes a long time. In such problems, heuristic methods are more advantageous and more preferred [31]. Thus, while the mathematical models proposed for the solution of the HHCSRP are applied due to the acceptable solution time for small-scale problems, heuristic approaches are preferred for large-scale problems with high data density [10, 24].

3.1 Mathematical Model of the Problem

The mathematical model of the HHC problem discussed in this study was created by utilizing the mathematical model suggested by Şimşek [24] in his doctoral study. In the models developed in the studies, cost minimization was aimed by minimizing the total travel distance/time [26]. Tables 3, 4, and 5 show the parameters and indices of user in the mathematical model.

The objective function (Eq. 1) consists of three different objectives. The first goal is to distribute a balanced workload between teams, the second goal is to minimize the route length of the teams, and the third goal is to minimize the variable used to relax the strict time interval that occurs at the point of complying with the visiting hours.

$$\min \mathrm{z} = \sum_{e \in E} wl_e + \sum_{e \in E} \sum_{i e (H^{\circ} \cup P \cup TR)} \sum_{j \in (H' \cup P \cup TN)} \left(dij\ x\ eflow_{e,i,j} \right) + maxdepart \tag{1}$$

Here,

$$eflow_{e,\,i,\,j} = \begin{cases} 1, & \text{If team } e \text{ is going from node } j \text{ to node } i \\ & 0, \text{otherwise} \end{cases}$$

$$ep_{e,\,p} = \begin{cases} 1, & \text{If team } e \text{ is assigned } \textit{to} \text{ patient } p \\ & 0, \text{otherwise} \end{cases}$$

$edepart_{e,\,i}$= Departure time of team e from node i

$earrive_{e,\,j}$= Arrival time of team e to node j

$maxdepart$ = The largest possible difference between the end of the patient's appointment time and the time the team leaves the patient,

$workload_e$ = Team workload.

Table 3 Some notations and explanations in the model

Expression	Definition
H°	Hospital
H'	Virtual hospital
P	Patient cluster
E	Set of available teams
TN	Set of transfer points

Table 4 Indices and their equivalents

Indices	Definition
p	Patient index
e	Team index
i	Index of the cluster ($H^{\circ} \cup P \cup TN$) consisting of transfer points, patient, and hospital
j	Index of the cluster ($H' \cup P \cup TN$) consisting of patient, transfer points, and virtual hospital
h	Index of the cluster ($P \cup TN$) consisting of patient and transfer points
m	Index of the set of all nodes ($H^{\circ} \cup P \cup TN \cup H'$)

Table 5 Parameters and equivalents

Parameter	Definition
$Duration$	Visit duration
X_m	X coordinate of node m
Y_m	Y coordinate of node m
A_m	Opening time of node m
K_m	Closing time of node m
D_{mm}	Internode distance matrix
K	A large number

Constraints

Constraint to assign only one team per patient

$$\sum_{p \in P} \left(eP_{ep} \right) = 1 \tag{2}$$

Constraint so that the number of patients assigned to teams does not exceed the maximum workload

$$\sum_{\boldsymbol{P \in P}} \left(\boldsymbol{eP_{e,p}} \right) \leq workload_{e} \quad \forall \boldsymbol{e \in E} \tag{3}$$

Constraint for each team to start their routes from the hospital

$$\sum_{j \in \left(H' \cup P \cup TN \right)} eflow_{e,H^{\circ},j} \leq 1 \quad \forall e \in E \tag{4}$$

Constraint for each team leaving the hospital return to the hospital at the end of their route

$$\sum_{j \in \left(H^{1} \cup P \cup TN \right)} eflow_{e,H^{\circ},j=} \sum_{i \in \left(H^{\circ} \cup P \cup TN \right)} eflow_{e,i,H^{1}} \quad \forall e \in E,\ \forall h \in \left(P \cup TN \right) \tag{5}$$

Constraint for teams to leave the patients they reached

$$\sum_{i\in\left(H^{\circ}\cup P\cup TN\right)} eflow_{e,i,h=} \sum_{j\in\left(H^{1}\cup P\cup TN\right)} eflow_{e,h,j} \quad \forall e\in E, \forall h\in\left(P\cup TN\right) \tag{6}$$

Constraints for teams to start their routes after the shift starts and finish their routes by the end of the shift

$$edepart_{e,H^{\circ}} \geq A_{H^{\circ}} \left(1.6\right) \; e\alpha rrive_{eH^{1}} \leq K_{H^{1}} \tag{7}$$

Constraints to visit patients during visiting hours as much as possible: 1.8, 1.9, 1.10

$$edepart_{e,p} = e\alpha rrive_{e,p} + duration_{P} \tag{8}$$

$$e\alpha rrive_{ep} \geq A_{P} \tag{9}$$

$$edepart_{e,p} \leq K_{p} + maxdepart \forall P \in P \tag{10}$$

Constraint for time tracking

$$\begin{gathered} edepart_{e,i} + t_{i,j} - \left(Kx\left(1 - eflow_{e,i,j}\right)\right) \leq e\alpha rrive_{e,j} \\ \forall e\in E,\;\; \forall i\in\left(H^{\circ}\cup P\cup TN\right),\;\; \forall j\in\left(H'\cup P\cup TN\right),\;\; i\neq j \end{gathered} \tag{11}$$

4 Solution Approach

In this study, a clustering-based route optimization approach is presented for the home healthcare scheduling and routing problem. The patients are clustered according to their locations, and the resulting clusters are used to build up the medical home teams. Then, scheduling and routing problem for each team is solved separately.

4.1 Clustering Analysis

Clustering analysis is a technique of grouping similar samples into several clusters based on the values of some observed variables for each data sample [32]. Clustering tries to group similar data samples in the same cluster and different data samples in different clusters to minimize intercluster similarity while maximizing intra-cluster similarity [33].

K-Means Clustering Algorithm

The K-means algorithm is one of the most used and oldest clustering algorithms. The reason why it is so widely used and popular is that it works very fast in large volumes of data, has low time complexity, and is easy to implement [34].

In summary, K sample points are first selected as initial cluster centers in the algorithm. Then, the remaining samples are collected at the cluster centers according to the minimum distances. In this way, the first clustering is done, and the center point of the clusters is calculated repeatedly until a suitable clustering is obtained [35].

The K-means algorithm is mathematically expressed by the following Eq. (12).

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{ci} (\mathrm{xi} - vj)2 \tag{12}$$

"$\|x_i - v_j\|$" is the Euclidean distance between x and v.

"c" is the number of cluster centers and "c_i" is the number of data at the ith cluster [36].

Steps of the K-Means Algorithm

Let $X = \{x_1, x_2, x_3, \ldots, x_n\}$ be the dataset and $V = \{v_1, v_2, v_3, \ldots, v_c\}$ the set of cluster centers.

Step 1. K samples are randomly selected as the center of K clusters.
Step 2. The distance between each sample and the cluster centers is calculated.
Step 3. Samples are assigned to the closest cluster centers.
Step 4. Cluster centers are recalculated. The following equation is used for the calculation:

$$vi = \left(\frac{1}{ci}\right) \sum_{j=1}^{ci} \mathrm{xi} \tag{13}$$

With this equation, the cluster center is calculated by dividing the sum of the data in a cluster by the number of data in the cluster.

Step 5. The distance of each sample to the new cluster centers is recalculated.
Step 6. Instances are assigned to their nearest cluster centers according to their new cluster centers.
Step 7. If any instance is assigned to a different cluster than the previous one in step 6, it should stop. If assigned, it returns to step 3 [36].

4.2 *Genetic Algorithms*

Genetic algorithms are search and optimization techniques based on natural genetics and natural selection principles [37]. If these expressions are interpreted for Np class problems, in each iteration, solutions with high fit remain in the population,

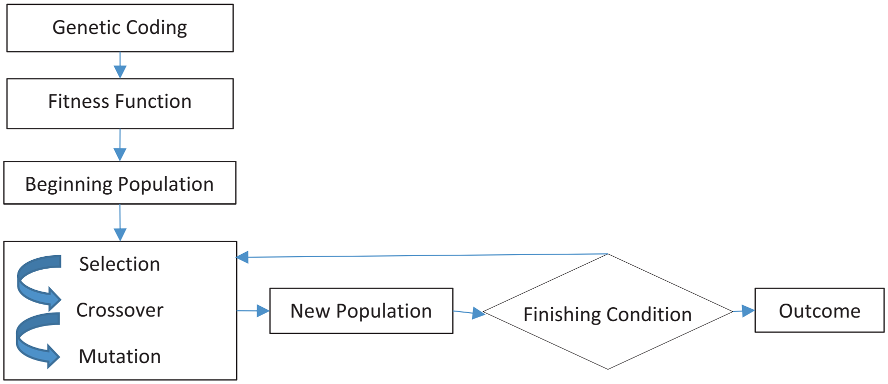


Fig. 1 Flowchart of GA

while solutions with low fit are eliminated from the population. In this way, the population, which gradually improves in each iteration as a result of natural selection and adaptation will eventually lead to the most suitable solution value. Since genetic algorithms adopt the continuation of the best solution, they use operators such as fitness function, crossover, mutation, and replication to generate new solutions [38]. The working mechanism of genetic algorithms is given in Fig. 1.

5 Experimental Study

Within the scope of the "Health Transformation" program in **Türkey**, home healthcare was put into practice in 2011 with the aim of mobilizing health services and providing health services to bedridden, elderly, or chronically ill patients at home in their own environment.

In this study, the determination of the patient clusters to be visited and the routes to be followed by the health teams, working in the home healthcare unit of a state hospital in Istanbul, is calculated with the proposed approach.

5.1 Data Collection

Four hundred eighty-nine patient locations where the state hospital provides home healthcare services were obtained. The state hospital provides HHC services to a total of five regions. In each region, visits are carried out by two dressing teams and two doctors, depending on the patient locations that need to be visited each day and the type of service that patients should receive. Dressing teams and medical teams are separate from each other, and their routes do not cross throughout the visits. During the visits, each team visits the patients of only one region and follows its own route throughout the process. Teams visit patient locations as a team, and there is no separation.

In summary, patients who benefit from home healthcare services provided by the state hospital are first clustered according to their region of residence during the appointment. Then, according to the type of health service that the patient will receive, that is, depending on whether the patient is visited by the team with doctors or for the purpose of dressing, patient clusters are formed that dressing teams and medical teams should visit. After the clustering, which is shaped according to the information received from the patient during the appointment, the patients in the patient cluster that must be visited by the dressing teams are assigned to two dressing teams, and those in the cluster that must be visited by the medical teams are assigned to a team of two physicians. These assignments are not made by any known method in the hospital.

In this study, the location data of 57 patients residing in Bahcelievler, which is 1 of the 5 regions served by the state hospital, who, according to the information given to the hospital, were known to be visited by the team of doctors and the day to be visited, were used.

5.2 Problem Definition

Patients who will benefit from HHC services and will be visited on a certain day are known in the beginning of the planning phase. However, the sequence in which they will arrive and the specific team that will be visiting are both unknown at this time. The hospital data indicates that on day #1, teams of physicians will visit 28 patients, and on day #2, teams will see 29 patients. On each day, two groups of physicians will pay a visit.

The HHC problem discussed in the study has some characteristic features according to the information obtained from the hospital. These are as follows:

- Each team starts their route at the hospital and finishes at the hospital.
- Teams leave the hospital at 08:55.
- Patient visits are carried out between 08.55 and 16.00.
- The lunch break is between 12.00 and 13.00, and teams do not visit during this interval. However, work starts at 13.00 and the teams – if any – must be in the location to be visited.
- A team can visit up to 18 patients in a day.

The problem is solved under certain assumptions:

- It is envisaged that patient visits would last 15 min.
- Urban speeds are up to 30 km/h.
- Distance between locations and speed were used to figure out how long it would take to get from one patient's home to another, assuming there was no traffic.

Aim of this problem

- Assigning the patients to be visited to the existing two teams of doctors, ensuring a balanced workload distribution as much as possible.
- The creation of patient clusters that each team of doctors will visit.
- Routing the patients in the patient clusters to be visited by the teams of doctors in such a way that the total distance traveled is minimized.
- Visiting patients as much as possible during their scheduled times and avoiding delays as much as possible.

The problem solved in line with these purposes has several benefits for the organization.

Expected benefits of the solution.

At a strategic level

- To make the workplace fairer by spreading out the work evenly and to reduce the number of mistakes that can be caused by too much work.
- Taking the necessary precautions in advance by testing the capacity for possible patient demand increases.

At the operational level

- The opportunity to visit all patients in less time by traveling a shorter distance by planning team routes and to use the existing capacity more efficiently by making use of the free time that may arise.

5.3 Clustering Analysis

The first step in the solution to the problem is to perform clustering analysis to create the patient clusters that the teams will visit by assigning the patients to two existing teams in a way that ensures a balanced workload distribution as much as possible. The number of teams is known beforehand, thus choosing the K-means algorithm for clustering is suitable, since it is a nonhierarchical clustering technique, and it takes the input parameter number of clusters from the user.

As a result of the clustering, the 28 patients who needed to be seen on day #1 were split into two teams as shown in Table 6.

The distribution of the patient locations that the teams will visit on the first day is shown in the Fig. 2.

The simple K-means algorithm grouped the 29 patients who needed to be seen on day #2 into two teams as in Table 7. And the clustering result can be seen in Fig. 3.

Table 6 Day 1 patient clusters

Team 1	Team 2
Patient no.	Patient no.
13	1
14	2
15	3
16	4
17	5
18	6
19	7
20	8
22	9
23	10
24	11
25	12
27	21
28	26

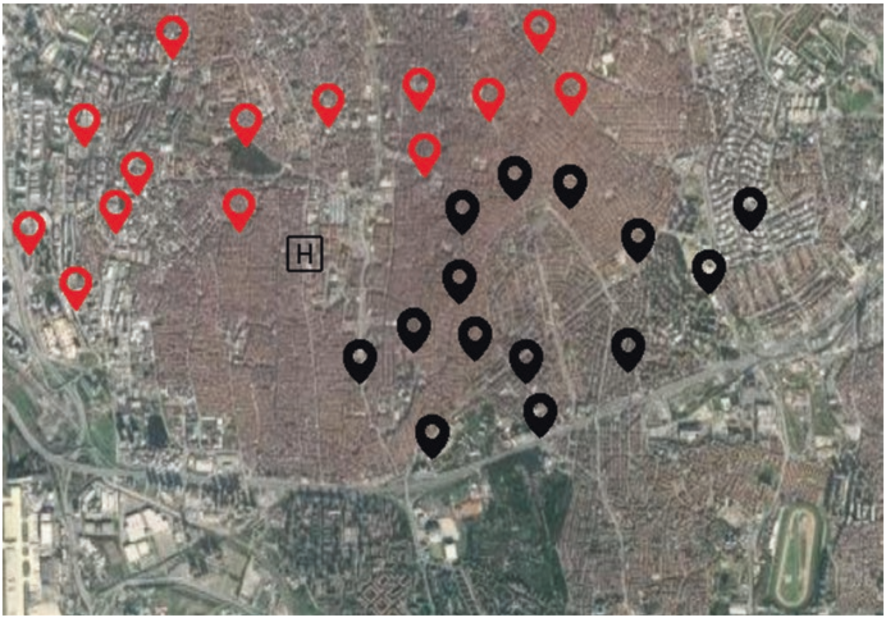


Fig. 2 Patient clusters for the first day

5.4 Routing with Genetic Algorithm

The second step in the solution of the problem is to solve the routing problem in each cluster to minimize the total distance traveled by the teams and the delays to

Table 7 Day 2 patient clusters

Team 1	Team 2
Patient no.	Patient no.
2	3
7	4
26	5
16	6
19	9
24	10
8	11
29	12
27	13
23	14
17	15
25	18
21	22
20	1
	28

Fig. 3 Patient clusters for the second day

the visiting hours. The aim of routing is to find the order in which the teams will visit the patients in the patient clusters in a way that will minimize the total distance and delays. For routing, it is necessary to know the distances between the locations of the patients in each patient cluster. In this study, the distances between the patient

Fig. 4 Time representation

locations assigned to each team were found using Google Maps, one of the geographic information systems. A from-to distance matrix was created for each team, which consists of the distances between the patient locations they will visit. Distances are in kilometers. The round-trip distances between the locations are assumed to be equal.

Visiting hours are included in the model with the logic of positioning them in a decimal number format on a number line, and the model has been clarified. According to the problem, the teams leave the hospital at 08.55 and visit the patients until 16.00, for example, 8 in the integer part of the decimal number at the start time and 55 min (55/60 min) from 0.9166 to approximately 0.90 h in the decimal part of the decimal number. Thus, the time is 08.55, expressed as a decimal 8.9 on the model, as shown in the Fig. 4. All time parameters in the model are in hours.

Since routes will be made for the 2-day planning period, the routes of the patients who will be visited on the first day will be created in the following section and then the routes of the patients who will be visited on the second day.

5.5 Routes of Patients to Be Visited

In the patient cluster that Team #1 will visit, there are patients numbered 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 27, and 28 who need to be visited on Day #1. Final routes and visiting times for each patient on day #1 can be seen in Table 8. The difference between the proposed solution's performance and the current situation is also reported in Table 9.

The routing process in the problem has two separate purposes:

1. Minimizing the total distance traveled along the route.
2. Minimizing the total delay to appointment times.

First, the model is solved separately for these two different purposes. The best value obtained for each separate optimization scenario is noted. And then the model is run once again to minimize the maximum deviation from the previously found best values in terms of both travel distance and total delay. The final optimization phase resulted in the following route in Fig. 5, for Team #1 on the first day.

According to this route, Team 1 departs from the hospital, visits Patient 19 first and Patient 17 last, and ends its route at the hospital.

The solutions of the routing model of Team #2, which will visit the rest of the patients on the first day, and Teams #1 and #2, which will visit the patients on the second day, are also solved.

Table 8 Route of Team #1 for the first day

From	To	Distance	Duration	Visit time	Appointment time	Delay
0	19	1.6	0.0333	89333.00	9	0.046667
19	13	0.06	0.02	9.205333	9.2	0.005333
13	28	0.06	0.02	9.482	9.4	0.082
28	27	1	0.0333	9.765333	9.6	0.165333
27	14	0.75	0.025	10.04367	10	0.043667
14	24	0.85	0.02833	10.31867	10.25	0.068667
24	20	1.3	0.04333	10.597	10.5	0.097
20	25	1.1	0.03667	10.867	10.8	0.067
25	18	1.7	0.05667	11.137	11	0.137
18	22	0.8	0.02667	11.44033	11.5	0.059667
22	23	0.06	0.002	11.727	11.75	0.023
23	16	1.4	0.04667	13	13	0
16	15	1.5	0.05	13.3	13.3	0
15	17	0.95	0.03167	13.58167	13.5	0.081667
17	0	2.5	0.08333	13.915	14	0.085
Total distance		16.11	0.537		Total delay	0.962

Table 9 Routing results for the first day for Team #1

	Original	Optimum	Deviation		
Route length	16.11	14.61	0.10267	Max dev.	0.10267
Delay	0.99967	0.92333	0.08267	Avg. dev.	0.0967

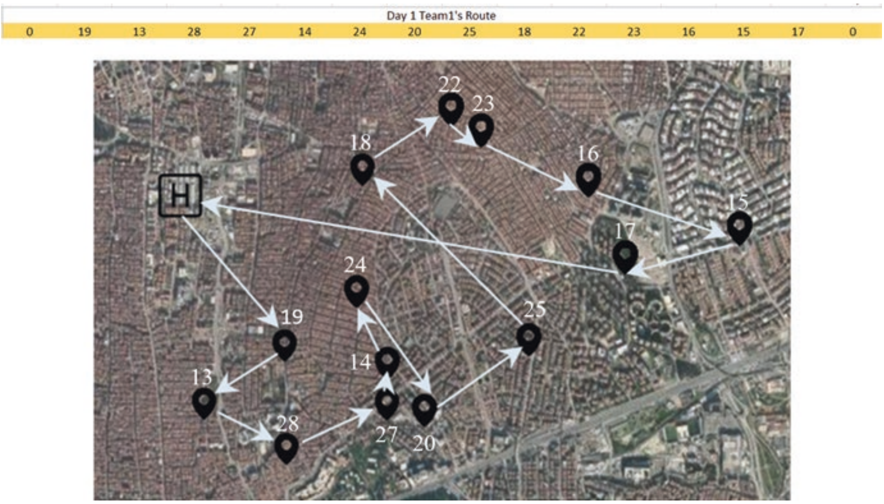


Fig. 5 Optimal routing found after multi-objective optimization for the first day of Team #1

5.6 Comparison of the Results

After the patient clusters that the teams will visit were created using the K-means algorithm, the patients in each cluster were routed with the genetic algorithm. In Tables 10 and 11, there are comparisons of the routes suggested in the study with the existing routes used by the hospital.

According to the existing routes in the hospital, the route length of Team #1 on the first day is 23.45 km and Team #2 is 25.62 km. Patients who needed service on the first day were visited by two teams covering a distance of 49.07 km. According to the suggested routes in the study, the route length of Team #1 on day #1 is 16.11 km and Team #2 is 14.81 km. Patients who needed service on the first day were visited by two teams covering a distance of 30.92 km. As a result, the total distance travelled for the day was decreased by 36.99% with the routes suggested in the study.

According to the existing routes in the hospital, the route length of Team #1 on the second day is 22.65 km and Team #2 is 31.2 km. Patients who needed service on the second day were visited by two teams covering a distance of 53.85 km. According to the suggested routes in the study, the route length of Team #1 on the second day is 13.71 km and Team #2 is 15.33 km. Patients who needed health services on the second day were visited by two teams covering a distance of 29.04 km. As a result, the total distance travelled on the second day decreased by 46.07% compared with the routes suggested in the study.

Table 10 Comparison of the current situation of the proposed approach on day #1

	Existing routing	Proposed routing
Team #1	23.45	16.11
Team #2	25.62	14.81
Total route	*49.07*	*30.92*
Total improvement	36.99%	

Table 11 Comparison of the current situation of the proposed approach on day #2

	Existing routing	Proposed routing
Team #1	22.65	13.71
Team #2	31.2	15.33
Total route	53.85	29.04
Total improvement	46.07%	

6 Discussion and Conclusion

Within the context of the "Transformation in Health" program in Turkey, home healthcare services were put into practice in 2011 with the aim of mobilizing and providing health services to bedridden, elderly, or chronically ill patients in their own home environment. In addition to the advantages that home healthcare services provide to their stakeholders, the continuous increase in the demand for health services with the aging of the world population and especially the COVID-19 pandemic we have been through have shown the importance of home healthcare services for high-risk groups.

In the study, a multi-objective HHC problem which aims to provide a balanced workload distribution between the teams while minimizing the total daily travel distances and the total delays to the patient appointment hours by considering the lunch breaks is discussed.

The solution of the multi-objective HHC problem consists of two steps: Clustering analysis was performed with the K-means algorithm to create clusters of patients to be visited by teams of doctors and route the patients in the clusters by using the genetic algorithm.

In the routes created by clustering analysis and genetic algorithm, the total distance travelled by the teams was improved by 36.99% on the first day and by 46.07% on the second day compared to the existing routes from the hospital. The results show that proposed approach has huge potential benefits in terms of service quality and costs. Employing a clustering and optimization-based home healthcare service will increase the efficiency of the resources and therefore increase the capacity of the current teams for the hospitals. Also, with the proposed approach, an evenly distributed workload will be provided for the separate healthcare teams.

The problem was solved according to the HHC operation conditions of the hospital. A problem can become more complex by making changes in the assumptions made in the solution to the problem. For example, the problem can include local traffic between locations based on the time of day, or delays can be calculated by fitting a probability distribution instead of using the relationship between travel time (distance/speed) and the locations.

Conflict of Interest and Ethical Compliance The authors declare no conflicts of interest. Ethics approval was not required for this study.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Çayır, Y. (2013). Birinci basamakta evde sağlık hizmetleri: Neler yapabiliriz? *Dicle Tıp Dergisi, 40*(2), 340–344.
2. Shepperd, S., & Iliffe, S. (2005). Hospital at home versus in-patient hospital care. *Cochrane Database of Systematic Reviews*, 3.

3. Hsiao, C. J., & Boult, C. (2008). Effects of quality on outcomes in primary care: A review of the literature. *American Journal of Medical Quality, 23*(4), 302–310.
4. Cheng, S. H., Hou, Y. F., & Chen, C. C. (2011). Does continuity of care matter in a health care system that lacks referral arrangements? *Health Policy and Planning, 26*(2), 157–162.
5. Çayır, Y. (2020). Home health care. *Turkish Journal of Family Medicine and Primary Care, 14*(1), 147–152.
6. Meriç, Ç. S., & Ayhan, N. Y. (2017). Yaşlılarda Evde Sağlık Hizmetleri ve Malnütrisyonun Önemi. *Beslenme ve Diyet Dergisi, 45*(3), 287–293.
7. Yilmaz, M., Sametoğlu, F., Akmeşe, G., Tak, A., Yağbasan, B., Gökçay, S. M., Doğanyılmaz, D., & Erdem, S. (2011). Sağlık hizmetinin alternatif bir sunum şekli olarak evde hasta bakımı. *İstanbul Tıp Dergisi, 11*(3), 125–132.
8. World Health Organization (WHO). (2004). *A glossary of terms for community health care and services for older persons* WHO Centre for health development ageing and health technical report (Vol. 5, pp. 43–45).
9. Doğusan, A. R. (2019). Türkiye'de Evde Sağlık Hizmetleri ile İlgili Mevzuat ve Gelişimi. *Ankara Medical Journal, 19*(3), 684–693.
10. Dengiz, A. Ö. (2021). *Evde Sağlık Hizmetleri Çizelgeleme ve Rotalama Problemi: Matematiksel Modeller Ve Meta-Sezgisel Algoritmalar*. Doktora, Başkent Üniversitesi Fen Bilimleri Enstitüsü.
11. Yurdakul, K., Alakaş, H. M., Tamer, E., & Gür, Ş. (2020). Yaşlılara Evde Bakım Hizmetinde Bulunan Ekiplerin Rotalanması: Büyükşehir Belediyesinde Bir Uygulama. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 9*(1), 206–223.
12. Bard, J. F., Shao, Y., & Wang, H. (2013). Weekly scheduling models for traveling therapists. *Socio-Economic Planning Sciences, 47*(3), 191–204.
13. Begur, S. V., Miller, D. M., & Weaver, J. R. (1997). An integrated spatial DSS for scheduling and routing home-health-care nurses. Interfaces, 27(4), 35–48.
14. Grenouilleau, F., Legrain, A., Lahrichi, N., & Rousseau, L. M. (2019). A set partitioning heuristic for the home health care routing and scheduling problem. *European Journal of Operational Research, 275*(1), 295–303.
15. Akjiratikarl, C., Yenradee, P., & Drake, P. R. (2007). PSO-based algorithm for home care worker scheduling in the UK. *Computers & Industrial Engineering, 53*(4), 559–583.
16. Yalçındağ, S., Cappanera, P., Scutellà, M. G., Şahin, E., & Matta, A. (2016). Pattern-based decompositions for human resource planning in home health care services. *Computers & Operations Research, 73*, 12–26.
17. Mankowska, D. S., Meisel, F., & Bierwirth, C. (2014). The home health care routing and scheduling problem with interdependent services. *Health Care Management Science, 17*(1), 15–30.
18. Rest, K. D., & Hirsch, P. (2016). Daily scheduling of home health care services using time-dependent public transport. *Flexible Services and Manufacturing Journal, 28*(3), 495–525.
19. Braekers, K., Hartl, R. F., Parragh, S. N., & Tricoire, F. (2016). A bi-objective home care scheduling problem: Analyzing the trade-off between costs and client inconvenience. *European Journal of Operational Research, 248*(2), 428–443.
20. Rasmussen, M. S., Justesen, T., Dohn, A., & Larsen, J. (2012). The home care crew scheduling problem: Preference-based visit clustering and temporal dependencies. *European Journal of Operational Research, 219*(3), 598–610.
21. Tanoumand, N., & Ünlüyurt, T. (2021). An exact algorithm for the resource constrained home health care vehicle routing problem. *Annals of Operations Research, 304*(1–2), 397–425.
22. Fikar, C., & Hirsch, P. (2017). Home health care routing and scheduling: A review. *Computers & Operations Research, 77*, 86–95.
23. Cissé, M., Yalçındağ, S., Kergosien, Y., Şahin, E., Lenté, C., & Matta, A. (2017). OR problems related to home health care: A review of relevant routing and scheduling problems. *Operations Research for Health Care, 13*, 1–22.
24. Şimşek, A. B. (2019). *Evde Sağlık Hizmetleri Atama ve Rotalama Problemi Üzerine*. Doktora, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.

25. Holm, S., & Angelsen, R. (2014). A descriptive retrospective study of time consumption in homecare services: How do employees use their working time? *BMC Health Services Research, 14*(439), 1–10.
26. An, Y. J., Kim, Y. D., Jeong, B. J., & Kim, S. D. (2012). Scheduling healthcare services in a home healthcare system. *The Journal of the Operational Research Society, 63*(11), 1589–1599.
27. Allaoua, H., Borne, S., L'etocart, L., & Wolfler Calvo, R. (2013). A matheuristic approach for solving a home health care problem. *Electronic Notes in Discrete Mathematics, 41*, 471–478.
28. Cheng, E., & Rich, J. L. (1998). *A home health care routing and scheduling problem*. Rice University.
29. Lanzarone, E., Matta, A., & Sahin, E. (2012). Operations management applied to home care services: The problem of assigning human resources to patients. *IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans, 42*(6), 1346–1363.
30. Bertels, S., & Fahle, T. (2006). A hybrid setup for a hybrid scenario: Combining heuristics for the home health care problem. *Computers & Operations Research, 33*(10), 2866–2890.
31. Çelik, Y., Yıldız, İ., & Karadeniz, A. T. (2019). Son Üç Yılda Geliştirilen Metasezgisel Algoritmalar Hakkında Kısa Bir İnceleme. *Avrupa Bilim ve Teknoloji Dergisi*, 463–477.
32. Sinharay, S. (2010). An overview of statistics in education. In *International encyclopedia of education* (pp. 1–11).
33. Rokach, L., & Maimon, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook* (pp. 321–352). Springer.
34. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR), 31*(3), 264–323.
35. Li, Y., & Wu, H. (2012). A clustering method based on K-means algorithm. *Physics Procedia, 25*, 1104–1109.
36. Takaoğlu, M. (2016). *Birkaç Veri Kümesi ile WEKA ve MATLAB Üzerinde Kümeleme Algoritmalarının Karşılaştırılarak İncelenmesi*. Doktora, İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü.
37. Deb, K. (1999). An introduction to genetic algorithms. *Sadhana, 24*(4), 293–315.
38. Çeyrekoğlu, S. (2017). *Araç Rotalama Problemine Genetik Algoritma Yaklaşımı Ve Örnek Bir Uygulama*. İstanbul Üniversitesi Fen Bilimleri Enstitüsü.
39. Hiermann, G., Prandtstetter, M., Rendl, A., Puchinger, J., & Raidl, G. R. (2015). Metaheuristics for solving a multimodal home-healthcare scheduling problem. *Central European Journal of Operations Research, 23*(1), 89–113.
40. Cordeau, J. F., Laporte, G., & Mercier, A. (2001). A unified Tabu search heuristic for vehicle routing problems with time windows. *Journal of the Operational Research Society, 52*(8), 928–936.
41. Eveborn, P., Flisberg, P., & Rönnqvist, M. (2006). Laps care – An operational system for staff planning of home care. *European Journal of Operational Research, 171*(3), 962–976.
42. Bennett, A. R., & Erera, A. L. (2011). Dynamic periodic fixed appointment scheduling for home health. *IIE Transactions on Healthcare Systems Engineering, 1*(1), 6–19.
43. Gayraud, F., Deroussi, L., Grangeon, N., & Norre, S. (2013). A new mathematical formulation for the home health care problem. *Procedia Technology, 9*, 1041–1047.
44. Bowers, J., Cheyne, H., Mould, G., & Page, M. (2015). Continuity of care in community midwifery. *Health Care Management Science, 18*(2), 195–204.
45. Nikzad, E., Bashiri, M., & Abbasi, B. (2021). A matheuristic algorithm for stochastic home health care planning. *European Journal of Operational Research, 288*(3), 753–774.
46. Wirnitzer, J., Heckmann, I., Meyer, A., & Nickel, S. (2016). Patient-based nurse rostering in home care. *Operations Research for Health Care, 8*, 91–102.
47. Carello, G., Lanzarone, E., & Mattia, S. (2018). Trade-off between stakeholders' goals in the home care nurse-to-patient assignment problem. *Operations Research for Health Care, 16*, 29–40.
48. Xiao, L., Dridi, M., & El Hassani, A. H. (2018). Mathematical model for the home health care scheduling and routing problem with flexible lunch break requirements. *IFAC-PapersOnLine, 51*(11), 334–339.

49. Taş, C., Bedir, N., Tamer, E., Alakaş, H. M., & Çetin, S. (2018). Evde Sağlık Hizmetlerinde Araç Rotalama İle Güzergahların Belirlenmesi: Devlet Hastanesinde Bir Uygulama. *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi, 4*(3), 264–283.
50. Cappanera, P., Scutellà, M. G., Nervi, F., & Galli, L. (2018). Demand uncertainty in robust home care optimization. *Omega, 80*, 95–110.
51. Erdem, M., & Koç, Ç. (2019). Analysis of electric vehicles in home health care routing problem. *Journal of Cleaner Production, 234*, 1471–1483.
52. Decerle, J., Grunder, O., El Hassani, A. H., & Barakat, O. (2016). A two-phases matheuristic for the home care routing and scheduling problem. *IFAC-PapersOnLine, 49*(12), 1484–1489.

Obesity Level Prediction Using Multinomial Logistic Regression

Shruti Srivatsan

Abstract With the increase in remote jobs and also due to the present ongoing pandemic for the past 1 year, most of the IT companies have taken a decision and advised their employees to work from home. This may lead to physical inactivity and result in obesity. Obesity occurs due to lack of proper diet and physical exercise. In this research work, a machine learning model has been used to predict the obesity level, which in turn helps the person to take suitable measures. The dataset used in this work is chosen from the well-known UCI Machine Learning Repository. The dirty data is cleaned and made suitable for further processing by the model. A new feature is added, and its performance is tested with and without that feature using the logistic regression model. The regression model built with all the input features yielded 82% accuracy, whereas with the addition of a new feature the accuracy improved to 96%. Further, the overall performance metrics of the proposed models are quite encouraging.

1 Introduction

Lifestyle diseases like obesity and hypertension are on the rise due to different changes made to our lifestyle. When an illness develops due to a person's routine habits and junk eating patterns, it is called a lifestyle disease. They are said to be chronic, which could be cured on early diagnosis only.

In the twenty-first century, there has been a massive increase in the number of people dying due to lifestyle diseases [1]. Some studies show a surge in Parkinson's disease among diabetics with a higher body mass index [2]. The increased inflammatory cytokines associated with obesity lead to a higher level of morbidity in COVID-19 patients who are obese [3].

S. Srivatsan (✉)
Master of Data Science and Artificial Intelligence, University of Waterloo,
Waterloo, ON, Canada

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_2

This study mainly focuses on implementing data mining (DM) to denote the obesity category of an individual. This work consists of five sections. After the introduction, the literature review is presented followed by a brief summary of the dataset. The proposed machine learning models suggested in this work are described in Sect. 4, and the final section shows the experimental results and conclusions.

1.1 Obesity

Obesity can be defined to be a state of imbalance in a person's body when the calories ingested are much higher than the calories expended. This leads to fat deposition in different parts of the body. An obese person is susceptible of having respiratory diseases, metabolic disorders, and cardiovascular diseases [4].

The body mass index (BMI) refers to the body fat accumulation based on an individual's height and weight. It is represented as follows:

$$\mathrm{BMI} = \mathrm{weight}\left(\mathrm{kg}\right) / \left(\mathrm{height}\left(\mathrm{m}\right)\right)^2 \tag{1}$$

There are six categories of BMI, highlighted in Table 1:

A person with a BMI greater than 25 can take necessary actions like modifying their dietary requirements and increasing physical activity to stay away from lifestyle diseases.

Some of the health concerns due to obesity are liver problems, sleep apnea, and coronary heart disease [5]. With early detection of obesity, these health complications can be prevented to an extent. This can be diagnosed earlier only by examining the symptoms which are less prevalent.

1.2 Data Mining (DM)

Data mining is the process of collecting data, analyzing it with different objectives, and giving an overview of the data in a useful form to identify patterns in the data, which is beneficial for effective decision-making.

Table 1 BMI categories

BMI category	BMI range
Underweight	<18.5
Normal	18.5–24.9
Overweight	25–29.9
Obesity I	30–34.9
Obesity II	35–39.9
Obesity III	> = 40

Knowledge discovery in databases (KDD) is usually seen as an exploratory analysis and extensive modeling of vast data repositories. The model used is for extracting the knowledge and analyzing and predicting the necessary data [6]. The process is highlighted in Fig. 1:

1.3 Applications of Data Mining

Data mining as a technique can be applied to varied forms of data that exist as *pictures*, *maps*, or *texts*.

- Text Mining
- Information extraction and text categorization using natural language processing can make text documents structured. Tokenization is performed for better understanding of the document [7].
- Web Mining
- Some websites spread dangerous content which could lead to a terrorist attack. Online social networks (OSNs) have helped in analyzing them [8].
- Process Mining
- Trust rate in customer's behavior has been computed using fuzzy systems. Analysis of customer bank transaction is performed, and a variety of steps have been followed for increasing the security of banking [9].
- Health Data Mining
- A patient's vital signs are analyzed using medical data mining to understand their illness and predict the outcome of the illness. A medical dataset mostly comprises

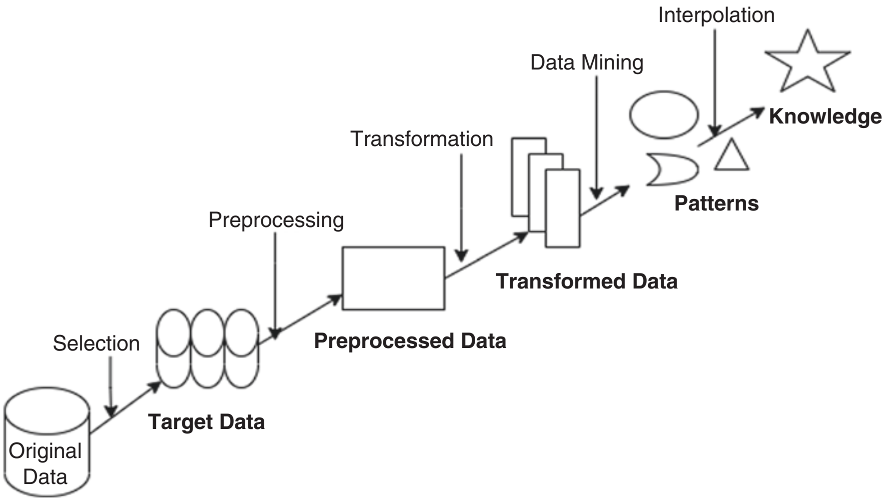


Fig. 1 KDD process [6]

of images either as scans or reports. Several techniques involving image processing have been performed to detect outliers in a scan [10]. Nowadays, smart watches are used to regularly check a person's heartbeat rate on a daily basis [11].

The remaining sections in this study elaborate on the literature review, followed by the dataset description and proposed methodology used in this work. The results are discussed and summarized in the conclusion section.

2 Literature Review

Analysis of medical data has been helpful in the diagnosis and prognosis of diseases including coronavirus, epilepsy, and heart attacks.

Algorithms like support vector machine (SVM), K-nearest neighbors (KNN), decision trees, and logistic regression have been used to predict the likelihood of a person getting heart disease. Decision tree yielded 93% accuracy, giving better results than the other models [12]. An optimal solution has been proposed by Ahmed et al. by using random forest classifier in the system to classify the medical data streams using Apache Spark. An accuracy of 95% has been obtained in the same [13]. Boosting algorithms have been used to study the coronavirus situation in different countries. The inference obtained after the analysis has been of great help in knowing the increase rate of the disease [14]. About 99% accuracy was obtained on a dataset from South Korea, on applying decision tree algorithm [15].

Epileptic seizure has been detected using electroencephalogram (EEG) signals, and the focal signals are detected using a fuzzy classifier called adaptive neuro fuzzy inference system (ANFIS) classifier [16]. Natural language processing (NLP) and text analytics have been used to infer that youngsters engage in more suicidal conversations about epilepsy than adults [17]. Models like random forest, support vector machine, and neural networks have been used for predicting the occurrence of glaucoma using Irish images. Neural networks resulted in 75% accuracy and 78% precision [18]. Backpropagation has been used with linear cyclic learning rate-based parabola function to classify eye diseases achieving an accuracy of 89.8% [19]. S. Srivatsan et al. have used random forest classifier to obtain an accuracy of 98% in predicting diabetes [20]. Photoplethysmography data has been used to train decision trees and K-nearest neighbors with the former resulting in 89.9% accuracy for predicting diabetes [21]. Shruti Srivatsan and Santhanam T in their work "Application of CART Decision Tree Model for Caesarean Section Prediction" used CART decision tree to obtain accuracy of 92.8% [22].

Childhood BMI values have been analyzed for predicting the risk of becoming overweight or obese during adolescence. A recall value of 96% was obtained by multilayer perceptron (MLP) after balancing the dataset using synthetic minority oversampling technique (SMOTE) [23]. Random forest was used to predict country-level obesity, which was higher with greater intake of cheese, baked goods, and carbonated drinks [24]. Electronic health record data was used to predict childhood obesity. Extreme gradient boosting (XG Boost) yielded 81% AUC score and significant performance metrics [25]. In [26], Weka tool was used to analyze the obesity dataset based on eating habits and physical conditions.

3 Dataset Summary

The UC Irvine Machine Learning Repository is widely recognized among researchers from various fields such as life sciences, engineering, and business. It currently contains a total of 588 datasets, out of which 138 pertain to the life sciences domain. The dataset utilized in this particular study belongs to the life sciences category, specifically dataset number [27]. It has a dimension of 2087, 17, consisting of 16 input features and 1 output feature.

Below is a concise overview of the attributes or features included in the dataset:

1. Gender: It denotes the gender of the individual {female, male}.
2. Age: It denotes the age of the person suspected to have obesity. The values range from 14 to 61.
3. Height: It represents the height of a person, ranging from 1.45 m to 1.98 m.
4. Weight: It indicates the weight of a person, ranging from 39 kg to 179 kg.
5. family_history_with_overweight: It determines whether there is a family history of obesity {yes, no}.
6. FAVC: It denotes the frequent consumption of high caloric food {yes, no}.
7. FCVC: It denotes the frequency of consumption of vegetables, ranging from 1 to 3.
8. NCP: It indicates the number of main meals consumed per day, ranging from 1 to 4.
9. CAEC: It tells the consumption of food between meals {no, sometimes, frequently, always}.
10. SMOKE: It gives information on whether the person smokes {yes, no}.
11. CH2O: It indicates the daily consumption of water, ranging from 1 to 3.
12. SCC: It finds whether the calorie consumption is monitored {yes, no}.
13. FAF: It represents the physical activity frequency, having values in the range from 0 to 3.
14. TUE: It denotes the time spent using technology devices, ranging from 0 to 2.
15. CALC: It refers to the consumption of alcohol {no, sometimes, frequently, always}.
16. MTRANS: It indicates the transportation used {automobile, motorbike, bike, Public_Transportation, walking}.
17. NObeyesdad: It is the output feature which indicates the obesity category of the person having values like Insufficient_Weight, Normal_Weight, Overweight_Level_I, Overweight_Level_II, Obesity_Type_I, Obesity_Type_II, and Obesity_Type_III.

4 Proposed Methodology

The proposed objectives of the work are as follows:

- The *primary* objective is to perform feature engineering and to check the significance of newly added feature(s) in the performance of the model.
- The *secondary target* is to decrease the memory and time complexity.
- The last *objective* would be to get a better recall value.

The head of the dataset is provided in Tables 2 and 3.

The dataset has 2087 observations. It has been split into the ratio of 70:30 where 70% of the data is utilized for training the model and the remaining 30% which is not seen by the model is made to validate the model.

Some of the preprocessing techniques carried out in this work are listed below:

When missing values are not dealt properly, it results in poor model learning and undesirable outcomes. Hence, it is vital to handle them properly using appropriate imputation techniques. The dataset used in this work has *no missing values and outliers*.

Out of 2087 records, from Fig. 2, it is observed that 351 records are of *Obesity_Type_I*, 297 are of *Obesity_Type_II*, 324 are of *Obesity_Type_III*, 276 are of *Overweight_Level_I*, 290 are of *Overweight_Level_II*, 282 are of *Normal_Weight*, and 267 are of *Insufficient_Weight* category. From this, one can infer that the dataset is not highly *imbalanced*. The last preprocessing step is to encode the variables which are categorical in the dataset.

One-hot encoding technique is used on the attributes gender, family_history_with_overweight, FAVC, SMOKE, SCC, and MTRANS, and label encoding is applied to encode the target variable into several categories. Next feature engineering is carried out by adding a new feature named BMI, which is computed using height and weight. This new feature is added and is interested to find its impact on the model with and without its presence.

The three models proposed to predict the various levels of obesity are as follows:

1. Logistic regression fitted on all input features excluding BMI – Proposed Model 1 (PM1).
2. Logistic regression fitted on all input features including BMI – Proposed Model 2 (PM2).
3. Logistic regression fitted on only BMI – Proposed Model 3 (PM3).

Table 2 Columns 1–8 of obesity dataset

Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP
Female	21	1.62	64.0	Yes	No	2.0	3.0
Female	21	1.52	56.0	Yes	No	3.0	3.0
Male	23	1.80	77.0	Yes	No	2.0	3.0
Male	27	1.80	87.0	No	No	3.0	3.0
Male	22	1.78	89.8	No	No	2.0	1.0

Table 3 Columns 9–17 of obesity dataset

CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
Sometimes	No	2.0	No	0.0	1.0	No	Public_Transportation	Normal_Weight
Sometimes	Yes	3.0	Yes	3.0	0.0	Sometimes	Public_Transportation	Normal_Weight
Sometimes	No	2.0	No	2.0	1.0	Frequently	Public_Transportation	Normal_Weight
Sometimes	No	2.0	No	2.0	0.0	Frequently	Walking	Overweight_Level_I
Sometimes	No	2.0	No	0.0	0.0	Sometimes	Public_Transportation	Overweight_Level_II

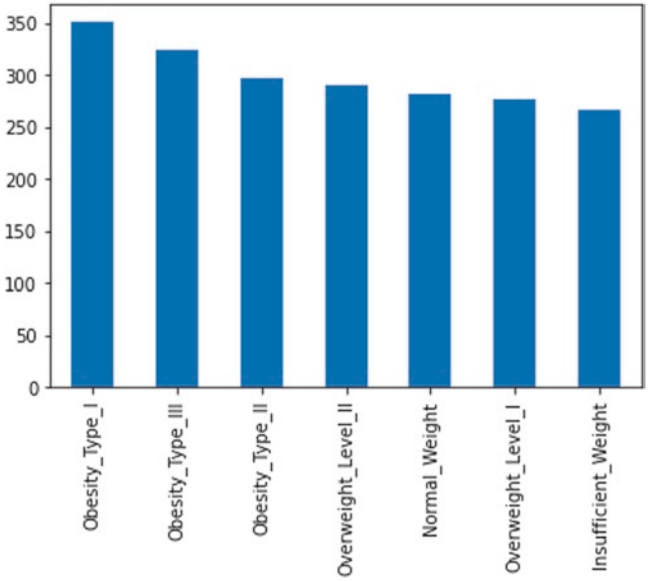


Fig. 2 Target variable count

From Fig. 3, one can observe that with only BMI, memory used is 32.6kB, whereas with all input features, it is 166.5 and with the new feature added to the existing features, it becomes 182.8 kB. In general, it is necessary to optimize the use of memory.

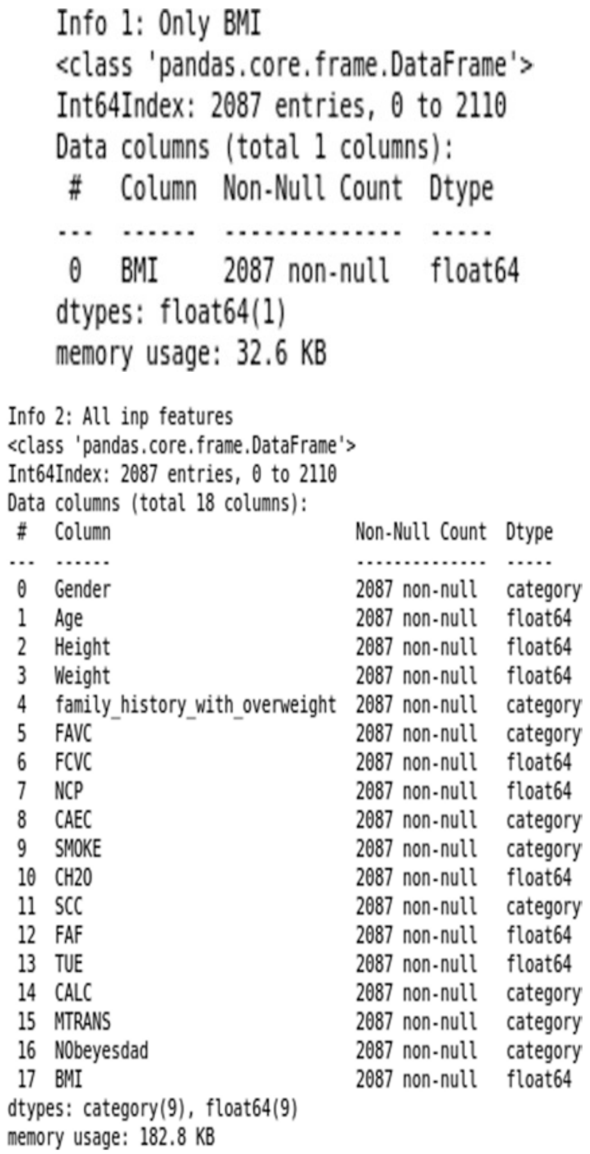
4.1 Steps in the Proposed Model

The following flowchart in Fig. 4 shows the proposed methodology in detail:

Logistic Regression (LR)

Logistic regression was initially developed as a model of population growth, and the term "logistic" was coined by Verlhust in the 1830s. Later, it was used as an alternative to the probit model by Edwin Wilson in 1943. David Cox introduced several refinements to the model in 1958 [28]. This model has been used for spam filtering using artificial bee colony algorithm to obtain significant results [29]. Bayesian logistic regression has been modeled for forecasting the amount of ice on the Arctic Sea cover [30].

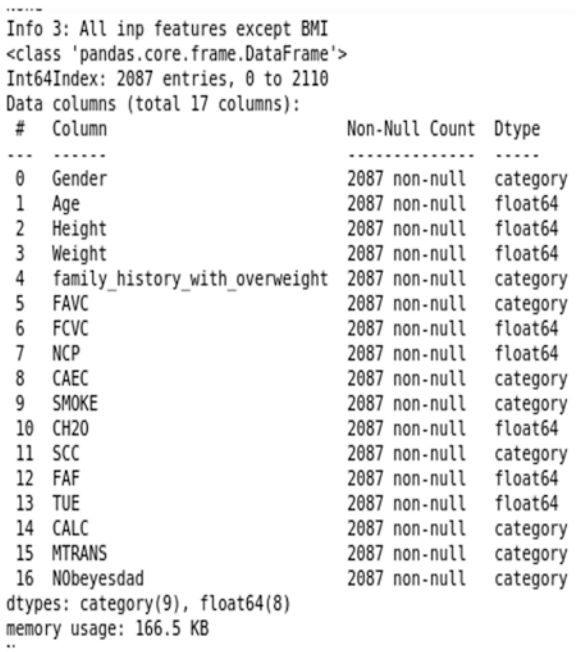
Though it is a regression technique implemented when the target variable is binary, it can be used for multiclass classification using multiclass = "multinomial" available in sklearn library for logistic regression model. The stratification is used to



```
Info 1: Only BMI
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2087 entries, 0 to 2110
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   BMI     2087 non-null   float64
dtypes: float64(1)
memory usage: 32.6 KB
```

```
Info 2: All inp features
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2087 entries, 0 to 2110
Data columns (total 18 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Gender                          2087 non-null   category
 1   Age                             2087 non-null   float64
 2   Height                          2087 non-null   float64
 3   Weight                          2087 non-null   float64
 4   family_history_with_overweight  2087 non-null   category
 5   FAVC                            2087 non-null   category
 6   FCVC                            2087 non-null   float64
 7   NCP                             2087 non-null   float64
 8   CAEC                            2087 non-null   category
 9   SMOKE                           2087 non-null   category
 10  CH2O                            2087 non-null   float64
 11  SCC                             2087 non-null   category
 12  FAF                             2087 non-null   float64
 13  TUE                             2087 non-null   float64
 14  CALC                            2087 non-null   category
 15  MTRANS                          2087 non-null   category
 16  NObeyesdad                      2087 non-null   category
 17  BMI                             2087 non-null   float64
dtypes: category(9), float64(9)
memory usage: 182.8 KB
```

Fig. 3 (**a**) Info of proposed model 1. (**b**) Info of proposed model 2. (**c**) Info of proposed model 3



```
Info 3: All inp features except BMI
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2087 entries, 0 to 2110
Data columns (total 17 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Gender                          2087 non-null   category
 1   Age                             2087 non-null   float64
 2   Height                          2087 non-null   float64
 3   Weight                          2087 non-null   float64
 4   family_history_with_overweight  2087 non-null   category
 5   FAVC                            2087 non-null   category
 6   FCVC                            2087 non-null   float64
 7   NCP                             2087 non-null   float64
 8   CAEC                            2087 non-null   category
 9   SMOKE                           2087 non-null   category
 10  CH2O                            2087 non-null   float64
 11  SCC                             2087 non-null   category
 12  FAF                             2087 non-null   float64
 13  TUE                             2087 non-null   float64
 14  CALC                            2087 non-null   category
 15  MTRANS                          2087 non-null   category
 16  NObeyesdad                      2087 non-null   category
dtypes: category(9), float64(8)
memory usage: 166.5 KB
```

Fig. 3 (continued)

ensure each cross-validation fold approximately possesses the same distribution of data for each class like the whole training dataset. Here, the number of repeats as 3 with k as 10 has been applied to evaluate the model performance [31].The model outputs well-calibrated probabilities on comparing with other classifiers and is less prone to over fitting in a low dimensional dataset. This algorithm allows models to reflect changes on data easily.

In this algorithm, the selection of a solver has played a vital role. A solver is an algorithm used in optimization problem. Other solver techniques applicable to logistic regression were tried, but did not get significant results. *Newton-cg*, an algorithm which calculates Hessian values explicitly, is used in this study. It may be computationally expensive at times when the data is of higher dimension in general, and in this work, it has yielded desirable results [32].

Algorithm

1. Apply the softmax function to a function (X,Y) having X which indicates the variables or features and Y which represents the label.
2. Set a decision boundary to predict the corresponding class.
3. Return the predicted class [33].

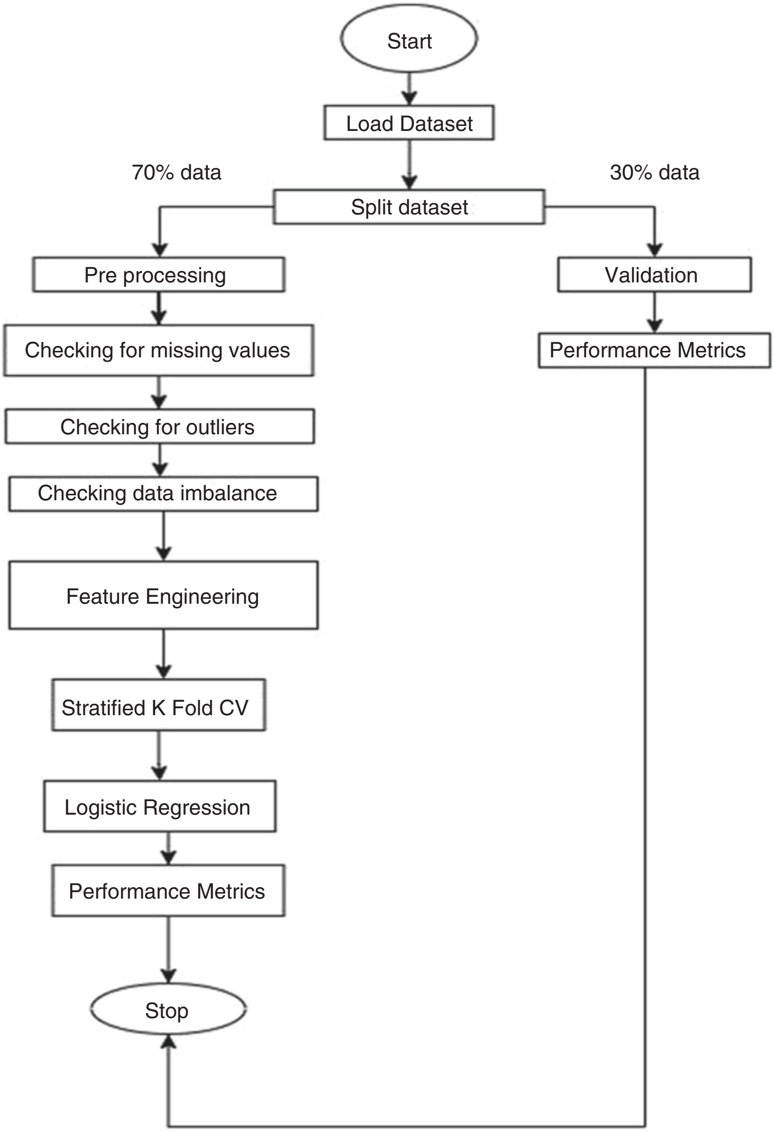


Fig. 4 Flow diagram of the proposed methodology

4.2 Performance Metrics

The following classification performance metrics are considered in this study:

1. *Confusion Matrix*

   It is a commonly used performance measurement evaluation metric used for finding the correctness of an algorithm. It consists of a table-like format with four values computed using actual and predicted values, shown in Table 4.

   The description of a confusion matrix is given below:

00 represents true negative (TN), 01 denotes false negative (FN), 10 indicates false positive (FP), and 11 shows true positive (TP).

- True negative: When the actual and predicted labels of a sample are negative.
- True positive: When the actual and predicted labels of a sample are positive.
- False negative: When the value of a sample is actually positive while its predicted label is negative.
- False positive: When the value of a sample is actually negative while its predicted label for the sample is positive [34].

2. *Precision*

   Precision refers to the measure of rightly classified classes of obesity among all positive classes:

$$\text{Precision} = \text{TP} / \left(\text{TP} + \text{FP}\right) \tag{2}$$

3. *Recall*

   Recall indicates the measure of correctly classified classes with all correctly predicted classes:

$$\text{Recall} = \text{TP} / \left(\text{TP} + \text{FN}\right) \tag{3}$$

4. *Accuracy*

   Accuracy, a crucial performance measure, is calculated as the ratio of correctly classified instances to the total number of instances in all classes [35]:

$$\text{Accuracy} = \left(\text{TP} + \text{TN}\right) / \left(\text{TP} + \text{TN} + \text{FP} + \text{FN}\right) \tag{4}$$

5. *f-Beta-Measure*

Table 4 Confusion matrix for binary classifier

Type	Actual positive	Actual negative
Predicted positive	TP (11)	FP (10)
Predicted negative	FN (01)	TN (00)

The f-beta-measure is an important performance metric utilized for comparing two models that exhibit varying levels of recall and precision by considering extreme values [36]:

$$\text{F - measure} = \left(\left(1+\beta^{2}\right) * \text{Recall}^{*}\text{Precision}\right) / \left(\left(\beta^{2*}\text{Precision}\right) + \text{Recall}\right) \tag{5}$$

A suitable beta-value is selected depending on the specific application domain. When equal weightage is assigned to both precision and recall, the beta-value is set to 1. If more importance is placed on precision, a value less than 1 is chosen, while a value greater than 1 is preferred for emphasizing recall [37].

6. *Area Under the Curve*

   This measure is employed to assess the ability of a model to accurately distinguish between different classes. If the value of the area under the curve is greater than 0.8, the model is considered to be well-fitted [38].

7. *Kappa*

   Kappa value denotes the inter-rater reliability among features. If it has a value >0.90, there is said to be almost perfect reliability [39].

5 Experimental Results

In this study, the performance evaluation of the proposed models was conducted using two different scenarios: utilizing all input features and utilizing reduced input features obtained through feature engineering.

Baseline accuracy serves as a simple prediction measure for assessing the model's performance without any rules, acting as a benchmark for achieving improved accuracy. In this study, the ZeroR classifier, which predicts the majority class solely based on the target variable while disregarding all predictors, was used as the baseline classifier. The baseline accuracy obtained from the ZeroR classifier (approximately 52%) can be used as a reference point for evaluating the performance of other classification techniques [40].

Subsequently, the stratified K-fold cross-validation technique was employed to train the model. This technique generally leads to better model performance compared to a single train-test split approach. The K-fold cross-validation technique with K set to 10 was utilized, where 70% of the data was used as the training set. The mean accuracy obtained after cross-validation was considered the training accuracy, while the accuracy predicted by the model for real-world data served as the final accuracy. This approach helps prevent overfitting and high bias in the model. Furthermore, the memory and training time requirements of the models were calculated, and Python with suitable libraries was used for model implementation.

The following sections provide detailed descriptions of the results obtained for the four proposed models:

5.1 Performance of Logistic Regression with All Input Features

The performance metrics of using logistic regression model using the all input features excluding BMI are shown in Tables 5 and 6.

The precision and recall values achieved in this study are deemed satisfactory, and the overall accuracy is considered reasonable. The F-measure, with a value of 0.82, indicates a decent performance in real-time data classification. Additionally, the AUC score of 0.970 indicates that the model can successfully classify between the two classes in 97% of cases.

As for the training and testing times, they were measured to be 45.114 s and 0.006 s, respectively. It is generally preferred to have a model that can learn quickly and perform well with a smaller amount of data.

5.2 Performance of Logistic Regression with All Input Features Including BMI

The performance metrics of using logistic regression is shown in Tables 7 and 8:

From Table 7, one can infer that the addition of a new feature, namely, BMI, has improved the model's performance to a significant level of about 14% in the performance metrics. The AUC score of 0.998 suggests that the model can distinguish well among the multiple classes. The training takes 64.100 s and testing takes 0.010 s. Among the two proposed models 1 and 2, it can be easily concluded that PM2 performs better.

The feature engineering done in this study (addition of BMI) has an impact in the performance of the model. This gave us an interest to exclusively analyze its performance in Proposed Model 3.

5.3 Performance of Logistic Regression Only with BMI

The metrics of modeling logistic regression is shown in Tables 9 and 10:

Based on the information presented in Table 9, it is evident that the constructed model demonstrates neither overfitting nor underfitting issues, as indicated by the

Table 5 Results of Proposed Model 1

Train acc (%)	Test acc (%)	Precision (%)	Recall (%)	f2-score	AUC score	Train time (seconds)	Test time (seconds)	Kappa
82.28	82.29	82	82	0.82	0.970	45.114 s	0.006 s	0.793

Table 6 Confusion matrix

Type	Actual Normal_Weight	Actual Overweight_Level_I	Actual Overweight_Level_II	Actual Obesity_Type_I	Actual Insufficient_Weight	Actual Obesity_Type_II	Actual Obesity_Type_III
Predicted Normal_Weight	**67**	**12**	**0**	**0**	**0**	**0**	**0**
Predicted Overweight_Level_I	**21**	**62**	**0**	**0**	**0**	**3**	**4**
Predicted Overweight_Level_II	**0**	**0**	**87**	**6**	**0**	**1**	**4**
Predicted Obesity_Type_I	**0**	**0**	**1**	**86**	**0**	**0**	**0**
Predicted Insufficient_Weight	**0**	**0**	**0**	**1**	**96**	**0**	**0**
Predicted Obesity_Type_II	**0**	**6**	**3**	**0**	**0**	**63**	**20**
Predicted Obesity_Type_III	**0**	**2**	**17**	**2**	**0**	**8**	**55**

Table 7 Results of proposed model 2

Train acc (%)	Test acc (%)	Precision (%)	Recall (%)	f2-score	AUC score	Train time (seconds)	Test time (seconds)	Kappa
95.89	96.49	96	96	0.96	0.998	64.100 s	0.010 s	0.959

Table 8 Confusion matrix

Type	Actual Normal_Weight	Actual Overweight_Level_I	Actual Overweight_Level_II	Actual Obesity_Type_I	Actual Insufficient_Weight	Actual Obesity_Type_II	Actual Obesity_Type_III
Predicted Normal_Weight	**77**	**2**	**0**	**0**	**0**	**0**	**0**
Predicted Overweight_Level_I	**5**	**84**	**0**	**0**	**0**	**1**	**0**
Predicted Overweight_Level_II	**0**	**0**	**96**	**1**	**0**	**0**	**1**
Predicted Obesity_Type_I	**0**	**0**	**0**	**87**	**0**	**0**	**0**
Predicted Insufficient_Weight	**0**	**0**	**0**	**0**	**97**	**0**	**0**
Predicted Obesity_Type_II	**0**	**5**	**0**	**0**	**0**	**83**	**4**
Predicted Obesity_Type_III	**0**	**0**	**0**	**0**	**0**	**3**	**81**

training and testing accuracies. This suggests that the model is able to generalize well.

The accuracy, precision, and recall values achieved by the model are highly satisfactory. Additionally, the F-measure, with a value of 0.95 (close to 1), showcases a superior performance in classifying real-time data. The kappa value of 0.944 indicates a near-perfect agreement.

Regarding the training and testing times, they were measured at 12.28 s and 0.008 s, respectively.

By examining both Tables 9 and 10, it can be concluded that the overall performance of the model is excellent. Furthermore, Proposed Model 3 outperforms Proposed Model 2.

Table 11 provides information on the memory and time requirements for all the proposed models. It is evident from the table that PM3 has significantly optimized both memory and time requirements.

Furthermore, in terms of accuracy, all the models proposed in this study outperform the baseline accuracy achieved using the ZeroR classifier, which is approximately 52%.

From all the three proposed models, Proposed Model 3 has yielded better performance metrics. Also, the stated objectives of this study have been realized using Proposed Model 3.

6 Conclusion and Future Work

In this research study, multinomial logistic regression has been applied to know the various levels of obesity. A model built with the newly added feature BMI has drastically improved the overall performance of the model. This suggests that while doing data analysis, it is essential to look for new features which may enable the model to learn better and provide better outcomes. In this study, the Proposed Model 3 outperformed the other two models with its encouraging performance metrics.

This work is focused exclusively on the dataset selected for this study, with no consideration for other datasets or broader scope. All the three objectives have been realized.

Table 9 Results of Proposed Model 3

Train acc (%)	Test acc (%)	Precision (%)	Recall (%)	f2-score	AUC score	Train time (seconds)	Test time (seconds)	Kappa
94.13	95.21	95	95	0.95	0.997	12.28 s	0.008 s	0.944

Table 10 Confusion matrix

Type	Actual Normal_ Weight	Actual Overweight_ Level_I	Actual Overweight_Level_ II	Actual Obesity_ Type_I	Actual Insufficient_ Weight	Actual Obesity_Type_ II	Actual Obesity_Type_ III
Predicted Normal_Weight	**78**	**1**	**0**	**0**	**0**	**0**	**0**
Predicted Overweight_Level_I	**5**	**84**	**0**	**0**	**0**	**1**	**0**
Predicted Overweight_Level_II	**0**	**0**	**96**	**2**	**0**	**0**	**0**
Predicted Obesity_Type_I	**0**	**0**	**3**	**80**	**4**	**0**	**0**
Predicted Insufficient_Weight	**0**	**0**	**0**	**6**	**91**	**0**	**0**
Predicted Obesity_Type_II	**0**	**6**	**0**	**0**	**0**	**85**	**1**
Predicted Obesity_Type_III	**0**	**0**	**0**	**0**	**0**	**1**	**83**

Table 11 Comparison of memory and time requirements of all proposed models

Model	Memory (kB)	Train time (seconds)
Proposed Model 1	*166.5*	*45.114*
Proposed Model 2	182.8	64.100
Proposed Model 3	32.6	12.286

The future work can explore the possibility of the following:

1. Using different feature selection algorithms.
2. Applying other machine learning classifiers suitable for this dataset.
3. Optimizing the performance using hyper parameter tuning methods and fine-tuning other parameters which are part of the model.

References

1. Rozpędowska-Matraszek, D. J. (2020). Efficiency frontiers in treating lifestyle diseases. *Folia Oeconomica Stetinensia, 20*(1), 341–359.
2. Jeong, S. M., Han, K., Kim, D., Rhee, S. Y., Jang, W., & Shin, D. W. (2020). Body mass index, diabetes, and the risk of Parkinson's disease. *Movement Disorders, 35*(2), 236–244.
3. Dietz, W., & Santos-Burgoa, C. (2020). Obesity and its implications for COVID-19 mortality. *Obesity, 28*(6), 1005.
4. Nammi, S., Koka, S., Chinnala, K. M., & Boini, K. M. (2004). Obesity: An overview on its current perspectives and treatment options. *Nutrition Journal, 3*(1), 1–8.
5. Ogunbode, A. M., Fatiregun, A. A., & Ogunbode, O. O. (2009). Health risks of obesity. *Annals of Ibadan Postgraduate Medicine, 7*(2), 22–25.
6. Feyyad, U. M. (1996). Data mining and knowledge discovery: Making sense out of data. *IEEE Expert, 11*(5), 20–25.
7. Shaalan, K., Hassanien, A. E., Tolba, F., & (Eds.). (2017). *Intelligent natural language processing: Trends and applications* (Vol. 740). Springer.
8. Goradia, R., Mohite, S., Jhakhariya, A., & Pinjarkar, V. (2020). Web mining to detect online spread of terrorism. *International Journal of Engineering Research & Technology (IJERT), 9*(7), 645–648.
9. Bekamiri, H., Ghasempour Ganji, S. F., Simonetti, B., & Seno, S. A. H. (2021). A new model to identify the reliability and trust of internet banking users using fuzzy theory and data-mining. *Mathematics, 9*(9), 916.
10. Antonie, M. L., Zaiane, O. R., & Coman, A. (2001, August). *Application of data mining techniques for medical image classification*. In Proceedings of the second international conference on multimedia data mining (pp. 94–101).
11. Muzammal, M., Talat, R., Sodhro, A. H., & Pirbhulal, S. (2020). A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. *Information Fusion, 53*, 155–164.
12. Abdar, M., Sharareh, R. N. K., Sutikno, T., Arji, G., & Subroto, I. M. I. (2015). Comparing performance of data mining algorithms in prediction heart diseases. *International Journal of Electrical and Computer Engineering (IJECE), 5*, 1569–1576. https://doi.org/10.11591/ijece.v5i6.pp1569-1576

13. Ahmed, H., Younis, E. M., Hendawi, A., & Ali, A. A. (2020). Heart disease identification from patients' social posts, machine learning solution on spark. *Future Generation Computer Systems, 111*, 714–722.
14. Albahri, A. S., Hamid, R. A., Alwan, J. K., Al-Qays, Z. T., Zaidan, A. A., Zaidan, B. B., et al. (2020). Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): A systematic review. *Journal of Medical Systems, 44*, 1–11.
15. Muhammad, L. J., Islam, M. M., Usman, S. S., & Ayon, S. I. (2020). Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. *SN Computer Science, 1*(4), 206.
16. Deivasigamani, S., Senthilpari, C., & Yong, W. H. (2021). Machine learning method based detection and diagnosis for epilepsy in EEG signal. *Journal of Ambient Intelligence and Humanized Computing, 12*, 4215–4221.
17. Falcone, T., Dagar, A., Castilla-Puentes, R. C., Anand, A., Brethenoux, C., Valleta, L. G., et al. (2020). Digital conversations about suicide among teenagers and adults with epilepsy: A big-data, machine learning analysis. *Epilepsia, 61*(5), 951–958.
18. Arunkumar, K. (2021). Machine learning models for the detection of human eye disease. *ICTACT Journal on Data Science and Machine Learning, 02*(03), 196.
19. Hameed, S. R., & Ahmed, H. M. (2021). Eye diseases classification using back propagation with parabola learning rate. *Al-Qadisiyah Journal of Pure Science, 26*(1), 1–9.
20. Srivatsan, S., & Santhanam, T. (2021, February). Application of ensemble classifiers for early diabetes detection. In *2021 grace hopper Celebration India (GHCI)* (pp. 1–6). IEEE.
21. Bavkar, V. C., & Shinde, A. A. (2021). Machine learning algorithms for diabetes prediction and neural network method for blood glucose measurement. *Indian Journal of Science and Technology, 14*(10), 869–880.
22. Srivatsan, S., & Santhanam, T. (2021). Application of CART decision tree model for caesarean section prediction. *Wesleyan Journal of Research*, 189–194.
23. Singh, B., & Tawfik, H. (2020). *Machine learning approach for the early prediction of the risk of overweight and obesity in young people*. In Computational science–ICCS 2020: 20th international conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part IV 20 (pp. 523–535). Springer.
24. Dunstan, J., Aguirre, M., Bastías, M., Nau, C., Glass, T. A., & Tobar, F. (2020). Predicting nationwide obesity from food sales using machine learning. *Health Informatics Journal, 26*(1), 652–663.
25. Pang, X., Forrest, C. B., Lê-Scherban, F., & Masino, A. J. (2021). Prediction of early childhood obesity with machine learning and electronic health record data. *International Journal of Medical Informatics, 150*, 104454.
26. Palechor, F. M., & de la Hoz Manotas, A. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. *Data in Brief, 25*, 104344.
27. https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition
28. Walker, S. H., & Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika, 54*(1–2), 167–179.
29. Dedeturk, B. K., & Akay, B. (2020). Spam filtering using a logistic regression model trained by an artificial bee colony algorithm. *Applied Soft Computing, 91*, 106229.
30. Horvath, S., Stroeve, J., Rajagopalan, B., & Kleiber, W. (2020). A Bayesian logistic regression for probabilistic forecasts of the minimum September Arctic Sea ice cover. *Earth and Space Science, 7*(10), e2020EA001176.
31. Brownlee, J. (2021). *Multinomial logistic regression with Python*. Retrieved from https://machinelearningmastery.com/multinomial-logistic-regression-with-python/
32. Logistic regression optimization parameters explained. Retrieved from https://holypython.com/log-reg/logistic-regression-optimization-parameters/

33. Maalouf, M. (2011). Logistic regression in data analysis: An overview. *International Journal of Data Analysis Techniques and Strategies, 3*(3), 281–299.
34. Hay, A. M. (1988). The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing, 9*(8), 1395–1398.
35. Jiao, Y., & Du, P. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quantitative Biology, 4*, 320–330.
36. Erika, D. *Accuracy, recall & precision*. Available [Online]: https://medium.com/@erika.dauria/accuracy-recall-precision-80a5b6cbd28d
37. Brownlee, J. *A gentle introduction to the Fbeta-measure for machine learning*. Available [Online]: https://machinelearningmastery.com/fbeta-measure-for-machine-learning/#:~:text=A%20default%20beta%20value%20is,the%20calculation%20of%20the%20score
38. Pandey, P. *Simplifying the ROC and AUC metrics*. Available [Online]: https://towardsdatascience.com/understanding-the-roc-and-auc-curves-a05b68550b69
39. McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.
40. Chan, P. K., & Stolfo, S. J. (1997). On the accuracy of meta-learning for scalable data mining. *Journal of Intelligent Information Systems, 8*, 5–28.

Importance of Feature Selection Methods in Machine Learning-Based Obesity Prediction

Usha Rani Gogoi

Abstract Obesity is an epidemic disease as being overweight or obese increases one's risk of developing serious health conditions such as diabetes, heart disease, hypertension, and certain cancers leading to premature death. However, early identification of causing factors makes obesity highly preventable. Owing to the objective of early detection and with the advancement of machine learning (ML) algorithms, author has evaluated the efficiency of the most widely used seven different ML algorithms in obesity detection. While evaluating, it has been observed that, depending on the set of selected features, the efficiency of the same ML algorithm may vary for the same experimental dataset. Hence, rather than comparing the performance of various ML techniques for obesity prediction, the feature selection techniques are also taken into account to make an unbiased decision. In this study, four different feature selection techniques: (a) univariate analysis, (b) feature importance, (c) mutual information, and (d) correlation were used. Experimental results show that among all the ML algorithms, the gradient boosting algorithm gives the highest classification accuracy of approximately 97% with only five features selected by the feature importance method. Thus, the proposed work demonstrates the importance of feature selection methods in ML-based obesity prediction.

1 Introduction

Since 1975, worldwide obesity has nearly tripled. More than 2.1 billion people are overweight or obese accounting for almost 30% of the global population for which obesity is considered a modern epidemic [1]. If this growth rate prevails, the proportion of people in the overweight and obese category is likely to rise up to 41% by the year 2030 [1]. Study shows that obesity is strongly linked to developing serious

---

U. R. Gogoi (✉)
Department of Computer Science & Engineering, School of Science and Technology, The Neotia University, Kolkata, West Bengal, India

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_3

illness like type 2 diabetes, respiratory problem, heart disease, cancers, and even death [1–5]. Considering these life risks, it is crucial to identify individuals at risk of developing obesity as early as possible. With the efficiency of machine learning (ML) algorithms in healthcare, researchers have made great efforts to develop a system for early obesity prediction. They analyzed the features by using various ML techniques. In [6], Dugan et al. used six ML models including random tree (RT), random Forest (RF), Iterative Dichotomiser-3 (ID3), J48, Naïve Bayes (NB), and Bayes Network on the CHICA dataset [7]. They reported that ID3 gave the highest accuracy of 85% with a sensitivity of 89%. The authors in [8] used a generalized linear model – RF – and partial least square for the prediction of obesity. They reported an accuracy of 89.68%. Dunstan et al. [9] used three types of ML algorithms – support vector machine (SVM), RF, and extreme gradient boosting (XGBoost) – for predicting the country-level obesity from food sales. Singh and Tawfik in [1] used seven ML algorithms: K-nearest neighborhood (KNN), J48 pruned tree, RF, bagging, SVM, multilayer perceptron (MLP), and voting. Among all, the highest precision value of 84% was obtained with MLP. Adnan et al. [10] combined Naïve Bayes (NB) and genetic algorithms to predict obesity. They reported a classification accuracy of 92% on a very small experimental dataset. Authors in Ref. [11] utilized three ML techniques – decision tree (DT), SVM, and k-means – for the estimation of obesity levels. They used a dataset of 178 samples and reported a precision value of 98.5%. They did not report the accuracy. In [12], the authors used nine ML algorithms: KNN, RF, logistic regression (LR), MLP, SVM, NB, adaptive boosting (AdaBoost), DT, and gradient boosting (GBoost) classifier. They reported the highest classification accuracy of 97.07% with LR using 28 features.

Although some works reported very high accuracy or other performance measures, it has been seen that the number of features they considered is very high. But to obtain high accuracy in ML models, training the model with relevant features is very crucial. To the best of my knowledge, no such work is there that analyzes the importance of identifying the significant features in ML techniques for obesity prediction. So, considering all these studies, the objective of the present work is to evaluate the importance of selecting discriminative features in the ML technique for obesity prediction. The contribution of this work is as follows:

(a) A comparative study of seven popular ML techniques has been made without applying the feature selection techniques.
(b) The efficiency of four different feature selection techniques has been evaluated to select the most efficient features for obesity prediction.
(c) The feature selection techniques are combined with different ML techniques to report the most efficient pair of feature selection techniques and ML algorithms for obesity prediction.

This chapter is divided into five sections. Section 2 describes the research methodology which includes the description of seven different ML techniques and the description of four different feature selection techniques. Section 3 describes the experimental results of our work. Section 4 provides a brief comparison of our work with existing works, and Section 5 concludes the chapter including future work.

2 Research Methodology

With the availability of various ML techniques, it is sometimes difficult for a new researcher to choose an ML algorithm that will solve the problem in the most efficient way. Moreover, depending on the set of selected features, the efficiency of the same ML algorithm may vary for the same dataset. Hence, rather than comparing the performance of various ML techniques for obesity prediction, the feature selection techniques are also taken into account to make an unbiased decision. The flow of the work is depicted in Fig. 1. Here, I have used seven supervised ML algorithms such as KNN, LR, DT, RF, XGBoost, SVM (linear and radial basis function kernel), and GBoost with four different feature selection techniques including (a) univariate analysis, (b) feature importance, (c) mutual information, and (d) correlation to come up with the best pair of feature set and ML algorithm for obesity prediction. A brief description of each phase of the methodology is provided below.

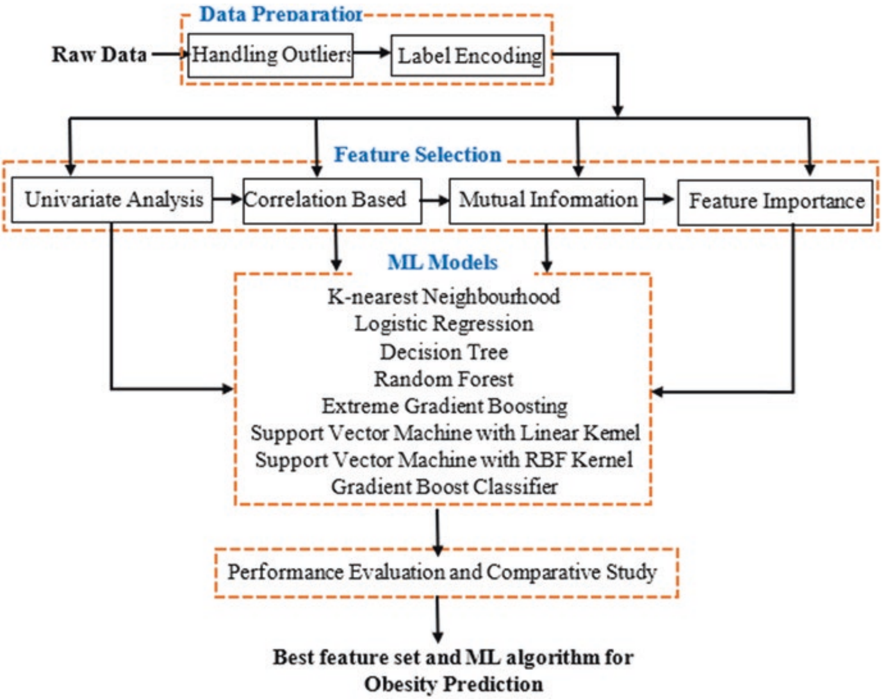


Fig. 1 Flow of the work

2.1 Data Collection and Preparation

In this study, an obesity dataset of 2111 samples aged between 14 and 61 years has been considered which is available in Kaggle [13, 14]. In this dataset, each sample is represented by using 17 attributes which were broadly categorized into three categories, namely, (1) attributes related to eating habits, (2) attributes related to the physical condition, and (3) others. Attributes related to eating habits include six attributes, namely, (a) frequent consumption of high caloric food (FAVC), (b) frequency of consumption of vegetables (FCVC), (c) the number of main meals (NCP), (d) consumption of food between meals (CAEC), (e) consumption of water daily (CH20), and f) consumption of alcohol (CALC). Attributes related to the physical condition include four attributes, namely, (a) calorie consumption monitoring (SCC), (b) physical activity frequency (FAF), (c) time using technology devices (TUE), and (d) transportation used (MTRANS). Other attributes included were gender, age, height, weight, family history of overweight, and smoking. The samples of the dataset are labeled in seven categories, namely, “Insufficient Weight,” “Normal Weight,” “Overweight Level I,” “Overweight Level II,” “Obesity Type I,” “Obesity Type II,” and “Obesity Type III,” based on the body mass index formula and information from the WHO and Mexican [14] normativity. The dataset is free from the unbalanced class problem, as 77% of the total data is generated synthetically by using the Weka tool and the filter SMOTE [14]. Moreover, there is no missing value in the dataset. However, to use this dataset for the experimental purpose, we need to prepare the data by using some preprocessing techniques as follows:

(a) *Handling Outliers*: In the experimental dataset, it is found that the distribution of some of the attributes is skewed, for which we have used the interquartile range approach to find the outliers. After removing the outliers using the quantile method, our experimental dataset comprised of 1856 sample data.
(b) *Conversion of Categorical Data to Numerical Data*: In the dataset, some of the attributes like gender, family history with overweight, FAVC, CAEC, SMOKE, SCC, CALC, and MTRANS are categorical in nature for which we need to convert them to numerical values before feeding them to any ML algorithm. So, one of the popular conversion methods, namely, “label encoding,” has been used to convert the categorical data to numerical values.

2.2 Machine Learning Algorithms

K-Nearest Neighborhood

KNN is one of the simplest supervised ML algorithms. The classification accuracy with different k-values like $K = 2$ to $K = 10$ has been evaluated, and the highest classification accuracy is obtained with $k = 5$. Thus, KNN classifier with $K = 5$ is used as the optimal model for performing the comparative study.

Logistic Regression

LR is one of the most commonly used and easy-to-implement ML algorithms for two-class classification problems. It can also be used for solving multiclass problems. LR uses a complex cost function defined by the sigmoid function. In LR, the estimation is done through maximum likelihood.

Decision Tree

DT is one of the popular and easiest classification techniques. Among various attribute selection measures, "entropy" is giving better results than "Gini index" or "Gini ratio." Likewise, the maximum depth of the tree was set as 3 as we obtain better accuracy with the maximum depth = 3. Except for these two parameters, the other parameters are set as default to obtain the optimal classifier to best fit the data.

Random Forests

RF is an ML technique that is used to solve regression and classification problems. RF uses ensemble learning, which is a technique that combines multiple decision trees to get a more accurate and stable prediction. A forest is supposed to be stronger when it has more trees. For obtaining a better prediction performance, a different number of decision trees are used to model the RF classifier, and the maximum classification accuracy was obtained with 20 numbers of decision trees. Besides this, all other parameters were set as default.

Gradient Boosting

GBoost algorithm is highly customizable to a specific problem. GBoost stands out for its prediction accuracy with large and complex datasets. The main idea behind the GBoost algorithm is to build a model that sequentially reduces errors in the subsequent models. So, the GBoost tries to build a base learner that is highly correlated with the negative gradient of the loss function, associated with the complete ensemble. No parameter tuning is done for GBoost as the highest classification accuracy in GBoost is obtained with the default parameters.

Extreme Gradient Boosting

XGBoost is an ensemble learning algorithm to build a strong classifier from several weak classifiers. It is the extension to GBoost. XGBoost uses advanced regularization which improves the model generalization capabilities. The training of XGBoost is very fast. No parameter tuning is done for XGBoost as its default parameters reported the highest accuracy.

Support Vector Machine

SVM is a very widely used classification algorithm. Unlike the other classification techniques, SVM aims in maximizing the margin between the decision boundary and data by minimizing the upper bound of generalization. SVM is a linear classifier; however, it may be easily extended to handle nonlinear data by using the kernel approach, which maps the low-dimensional input data to a high-dimensional space. For experimental purposes, we have used two different kernels: linear and radial basis function kernel. For both the kernels, no parameter tuning is done as the highest classification accuracies were obtained with the default values of these kernels.

2.3 Feature Selection Techniques

For designing a best-fitted model by reducing the computational cost and by improving the model performance, it is very crucial to identify the best discriminative features from the set of all features. By removing irrelevant and redundant features, feature selection reduces overfitting and improves the generalization of any model. It makes the model simpler. In identifying the best discriminative and important features, feature selection techniques play an imperative role. However, the efficiency of the feature selection technique varies depending on the dataset and the ML algorithms. Hence, in this work, our objective is to evaluate the efficiency of some of the most widely used feature selection techniques in obesity prediction. The four most popular feature selection techniques are used, and their efficiency in obesity prediction is studied in Section 4. The detail of each of these feature selection techniques is provided below.

Correlation-Based Method

Correlation describes the relationships between the features or the target variable. High correlation features are more linearly dependent and hence have same effect on the target variable. Correlation between two features can be positive, i.e., an increase in one feature increases the value of the other feature, or it can be negative which means one feature's value decreases, as the other feature increases. Whether positive or negative, our objective is to drop one of two features when they are highly correlated.

Mutual Information-Based Method

Mutual information is used to evaluate the arbitrary dependency between random variables. Mutual information is a measure between two features X and Y that quantify the amount of information obtained about one feature through the other feature. Mutual information is defined as follows.

Mutual information between two features X and Y is the amount of knowledge on feature Y supplied by X or the amount of knowledge of feature X supplied by Y. If feature X and Y are independent, i.e., X contains no information about Y, then their mutual information will be zero [15].

Feature Importance-Based Method

The feature importance-based feature selection method assigns a score to each feature based on the usefulness and relevancy of the feature to the output variable. The higher score indicates the relative importance of the feature in making a prediction. There are various methods like linear regression based, logistic regression based, ExtraTreesClassifier, etc. For our work, we have used the ExtraTreesClassifier method for obtaining scores against each feature. The ExtraTreesClassifier fits a number of randomized decision trees to the data. The random splits of all observations are carried out to ensure that the model does not overfit the data.

Univariate Analysis

The univariate feature selection method allows us to have a better understanding of the features and their relationship to the response variables. It examines each of the features to determine the strength of the relationship of the feature with the response variable [16]. It is called univariate because while analyzing the relationship between a feature and the target variable, we ignore the other features [16].

3 Experimental Results

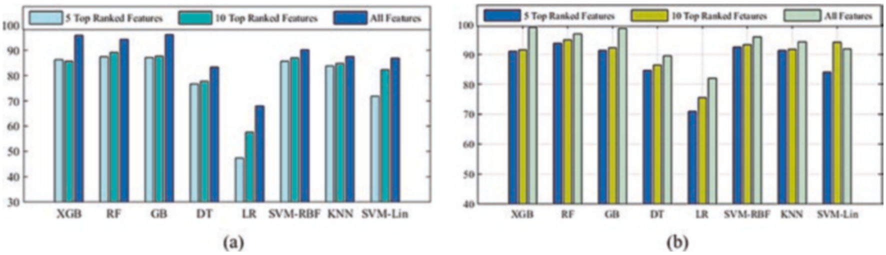
To perform an efficient comparative study, we evaluate the classification performance of each ML algorithm with and without the feature selection techniques. For evaluating the classification performance of the ML algorithms, five most widely used metrics – accuracy (ACC), area under the receiver operating characteristic curve (AUC), precision (PR), recall (RC), and F1 score – are used. The experimental dataset is split into two sets: a training set and a testing set with a ratio of 8:2. Moreover, to train the model in the most efficient way and to tackle the overfitting and underfitting, we have validated each model with k-fold cross-validation. We tried with the different numbers of folds and obtained the optimum result with k = 10 folds. The results of different experiments are discussed in the following subsections.

3.1 Classification without Feature Extraction Techniques

Here, the potential of all the features of the dataset is evaluated without applying any feature selection technique, and we fed the normalized features to each of the ML algorithms. We consider this setup as the base and compare the other experimental results with this only to evaluate the importance of identifying the discriminative features in ML-based obesity prediction. As illustrated in Fig. 2a, b, it has been seen that among all the ML algorithms, the GBoost model gives the highest classification ACC of 96.23% and AUC value of 98.74%. Likewise, the XGB and RF also provide classification ACC of 95.96% and 94.36%, respectively. As listed in Table 1, like the ACC and AUC, the PR, RC, and F1 score of GBoost, XGB, and RF are also found to be greater than equal to 94% each.

3.2 Classification with Univariate Feature Selection

To investigate the importance of identifying the significant features in obesity prediction and also to remove the redundant features, the univariate feature selection technique is used. The ranking of the features is illustrated in Table 2. The higher the score, the higher the rank. Out of all the features, two sets – one is of five high-ranked features (weight, age, gender, SCC, FAF) and the other is of 10 high-ranked features (weight, age, gender, SCC, FAF, family_history_with_obesity, MTRANS, FCVC, CAEC, NCP) – were considered to evaluate the performance of the ML algorithms. As illustrated in Fig. 2a, b, the highest classification accuracy (AUC) obtained with the set of five high-ranked features and with the set of ten high-ranked features is 87.40% (91.33%) and 89.15% (94.87%), respectively. The values of the



(a) (b)

Fig. 2 (**a**) Classification ACC and (**b**) AUC values of various ML algorithms without feature selection (all features) and with univariate feature selection (with five top-ranked features and ten top-ranked features) technique

Table 1 The PR, RC, and F1 score value with univariate feature selection

		Number of features								
		5	10	All	5	10	All	5	10	All
	Metrics	Precision			Recall			FI Score		
ML algorithms	XGB	86	86	98	84	85	98	85	85	98
	RF	89	91	94	89	91	94	89	91	94
	GB	87	87	98	85	86	98	85	86	98
	DT	74	78	85	72	76	85	71	76	80
	LR	46	56	68	47	58	68	44	56	67
	SVM-RBF	87	88	93	87	88	93	87	88	92
	KNN	85	85	90	84	85	90	84	85	89
	SVM – Lin	74	83	86	71	85	86	71	85	86

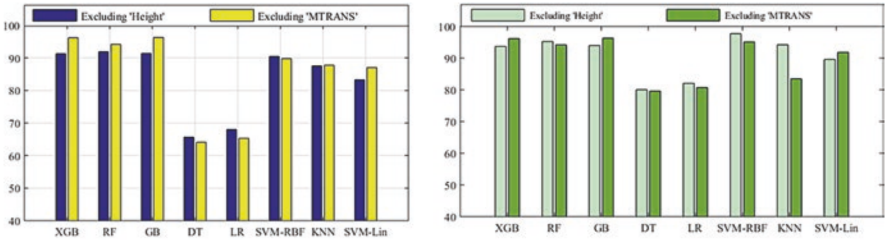
Table 2 Univariate feature selection based ranking of features

Feature number	Feature	Score	Feature number	Feature	Score
3	Weight	11136.19	8	CAEC	41.01
1	Age	418.39	7	NCP	33.53
0	Gender	281.15	5	FAVC	24.75
11	SCC	88.57	9	SMOKE	22.57
12	FAF	87.56	14	CALC	22.13
4	Family_history_with_overweight	85.07	13	TUE	20.57
15	MTRANS	80.03	10	CH2O	12.49
6	FCVC	49.65	2	Height	0.85

metrics PR, RC, and F1 are listed in Table 1. So, as illustrated in Fig. 2 and listed in Table 1, the selection of features using the univariate feature selection method reduces the classification performance of ML algorithms in obesity prediction. So, we can conclude that the way in which univariate feature selection ranked the features is not improving the classification accuracy of the base model.

3.3 Classification with Correlation-Based Feature Selection

In this section, we will discuss the importance and efficiency of the correlation-based feature selection method in obesity prediction. By observing the correlation coefficients, we found that none of the features are highly correlated. However, two features “Height” and “MTrans” are found to have a correlation coefficient of 0.64. Hence, the classification performance of the ML algorithms has been evaluated with only any of these two features at a time to evaluate which factor is more efficient in obesity prediction. Figure 3 clearly demonstrates that if *MTRANS* is considered over *Height*, then we obtain the highest classification accuracy (AUC) of 91.38% (93.99%). In contrary, if *Height* is considered over *MTRANS* with all other features,



(a) (b)

Fig. 3 (**a**) Classification accuracies and (**b**) AUC values of various ML algorithms with correlation-based feature selection technique

Table 3 The PR, RC, F1 score value with correlation-based feature selection

		Number of features					
		No ‘Height’	No ‘MTRANS’	No ‘Height’	No ‘MTRANS’	No ‘Height’	No ‘MTRANS’
	Metrics	Precision		Recall		FI Score	
ML algorithms	XGB	90	98	88	98	88	98
	RF	92	95	91	95	91	95
	GB	90	97	89	97	89	97
	DT	70	71	63	63	62	63
	LR	68	65	68	65	68	65
	SVM-RBF	96	92	96	90	96	90
	KNN	90	89	89	87	89	87
	SVM - Lin	83	89	81	85	82	85

then we obtain a classification accuracy (AUC) value of 96.29% (98.53%) which is much better than the previous.

Likewise, from the values of PR, RC, and F1 scores as listed in Table 3, we can conclude that although MTRANS is correlated with Height, Height is more discriminative and efficient in obesity prediction.

3.4 *Classification with Mutual Information-Based Feature Selection*

In this section, we have used the mutual information-based feature selection technique to reduce the feature set. But as listed in Table 4**,** none of the features are found to have mutual information of zero, for which we cannot discard any of the features. So, we can conclude that for obesity prediction, the mutual information-based feature selection method is not effective.

Table 4 Mutual information value of each feature

Feature number	Mutual information	Feature	Mutual information
Weight	1.303	Gender	0.226
Age	0.583	CAEC	0.183
Height	0.434	family_history_with_overweight	0.151
FCVC	0.389	CALC	0.098
FAF	0.334	MTRANS	0.062
CH2O	0.305	SMOKE	0.033
TUE	0.273	FAVC	0.031
NCP	0.253	SCC	0.002

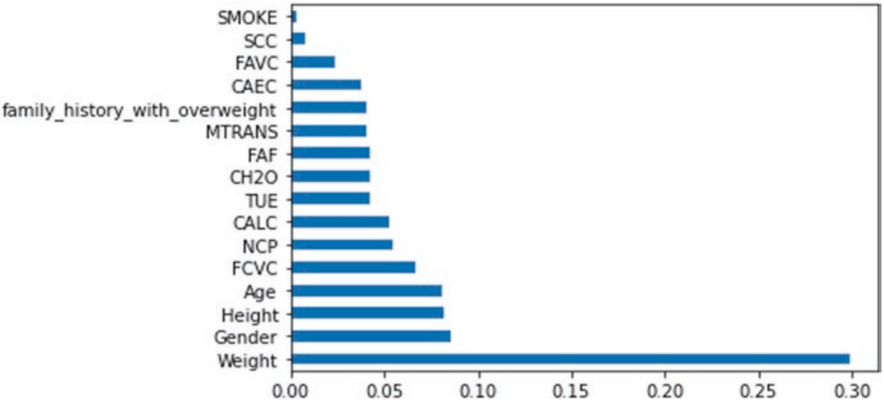


Fig. 4 Feature importance of each feature in the dataset

3.5 Classification with Feature Importance-Based Feature Selection

In this section, we have applied the feature importance-based feature selection method to remove all the unnecessary or nondiscriminative features. As depicted in Fig. 4, among all the features, weight has the highest feature importance and Smoke has the lowest feature importance.

Now, based on the feature importance, we have created two sets of features. In one set, we consider five higher-importance features, and in the other set, we consider those ten higher-importance features. We evaluate the performance of each of these two feature sets in obesity prediction. The accuracy and AUC values of each of the ML algorithm are depicted in Fig. 5a, b, respectively. As illustrated in Fig. 5a, among all the ML algorithms, the highest classification accuracy (AUC) reported by the GBoost algorithm with the set of five features is 96.87% (98.83%) and with the set of ten features is 95.82% (98.48%). Like GB, XGB and RF also provide higher classification accuracies (AUCs) compared to other remaining algorithms. To compare the classification performance of all the ML algorithms, the classification

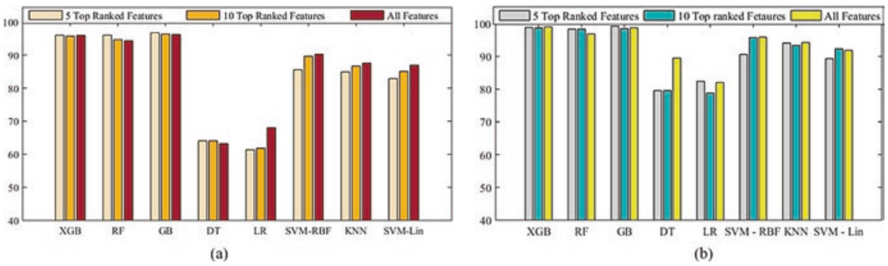


Fig. 5 (**a**) Classification accuracies and (**b**) AUC values of various ML algorithms with feature importance-based feature selection technique

Table 5 Classification performance of different ML algorithms with top 5 and top 10 important features, selected based on the feature importance method

	5					10					All Features				
Classifiers	Acc	AUC	Pr	Rc	F1	Acc	AUC	Pr	Rc	F1	Acc	AUC	Pr	Rc	F1
XGB	96.16	98.68	98	98	98	96.09	98.17	97	97	97	96.02	99	98	98	98
RF	95.69	97.16	95	95	95	94.21	98.26	97	97	97	94.88	97.12	95	95	95
DT	64.08	79.65	71	65	65	64.08	79.59	71	63	63	64.08	79.59	71	63	63
GBoost	**96.87**	98.83	98	98	98	**95.82**	98.48	98	98	99	**96.23**	98.74	98	98	98
LR	57.26	77.33	62	57	53	61.83	78.23	62	62	60	66.40	80.43	65	66	65
SVM-RBF	61.19	76.02	48	56	50	60.82	91.95	86	86	87	57.55	74.88	57	55	50
KNN	91.24	93.77	89	88	88	91.94	89	81	81	81	77.42	86.44	75	75	75
SVM-linear	71.23	82.27	69	67	66	74.60	84.88	75	73	72	77.36	87.76	79	78	78

performance with all 16 features is also plotted in the same figure. While comparing, we found that instead of considering all 16 features, it is sufficient to consider only 5 features with higher importance feature, because, as Fig. 5a depicts, the set of 5 high importance features provides higher classification accuracy (96.87%) than the set of 16 features (96.23%).

Similarly, the AUC value also proves the effectiveness of using the proper feature selection method to identify the discriminative feature and to improve the accuracy by reducing unnecessary, redundant features. The PR, RC, and F1 score values are listed in Table 5. So, we can conclude that instead of using 16 raw features, we can use only 5 features with higher feature importance values to obtain a higher obesity prediction performance.

4 Discussion

This section discusses the observations of this study. Although our intention is to compare our work with the other existing works, due to the absence of the experimental datasets of the existing works, it is not possible to directly compare our work with others. However, we tried to summarize the findings of the existing as well as our study in Table 6. As illustrated in Table 6, the highest obesity prediction was reported by F. Ferdowsy et al. [12] by using 28 features. In contrast, our approach of classification also reported almost 97% accuracy with only five important features selected based on the "feature importance"-based feature selection method. So, these classification results signify the importance of the feature selection techniques for comparing the performance of ML algorithms in obesity prediction.

5 Conclusion and Future Work

The main focus of this chapter is to evaluate the importance of feature selection methods in identifying the most efficient features as well as the ML algorithms for obesity prediction. In the literature, it is observed that although authors reported the highest classification accuracy of 97.07%, it requires a large set of 28 features. So, to make a fair comparison of the ML algorithms, we have applied four different feature selection methods, out of which, the experimental result reveals that the feature importance-based feature selection method is the most efficient for obesity prediction. With the set of five selected features, the gradient boosting algorithm

Table 6 Comparison of different obesity prediction approaches

Authors/year	ML algorithm used	Dataset used	Number of features	Classification accuracy
Dugan et al. [6], 2015	RT, RF, ID3, J48, NB, Bayes network	CHICA dataset	167 features	85%
K. Jindal et al. [8], 2018	Linear model, RF and partial least square	200 for each model	5 features	89.68%.
Dunstan et al. [9], 2020	SVM, RF, and XGBoost	–	–	–
Singh and Tawfik [1], 2020	KNN, J48 pruned tree, RF, bagging, SVM, MLP, and voting	(8160 N, 2126 overweight, and 824 obese) (imbalanced dataset)	1 feature (body mass index)	52% for Imbalanced data 92% for balanced data
Adnan et al. [10], 2012	Combined NB and GA	Very small experimental dataset	19 features	92%
R. C. Cervantes [11], 2020	Decision tree (DT), SVM, and k-means	178 samples	18 features	A precision value of 98.5%
F. Ferdowsy et al. [12], 2021	Nine ML algorithms: k-NN, RF, LR, MLP, SVM, NB, AdaBoosting, DT, and GBoost	1100 samples	28 features	97.07% with LR
Ours	Seven ML algorithms: KNN, LR, DT, RF, GBoost, XGBoost, SVM	1856 samples	5 features	96.87% with GBoost

gives us the highest accuracy of 96.86% (approximately 97%). So, based on the experimental result, we can conclude that the consideration of feature selection is very essential while evaluating the efficiency of various ML algorithms in obesity prediction.

Acknowledgments The work presented herein was conducted at the School of Science and Technology of the Neotia University, Diamond Harbour Rd., Sarisha, West Bengal 743368, West Bengal.

Conflict of Interest The author declares no potential conflict of interest with respect to the authorship and/or publication of this article.

References

1. Singh, B., & Tawfik, H. (2020). Machine learning approach for the early prediction of the risk of overweight and obesity in young people. In *International conference on computational science (ICCS)* (pp. 523–535). Springer.
2. Chatterjee, A., Gerdes, M. W., & Martinez, S. G. (2020). Identification of risk factors associated with obesity and overweight – A machine learning overview. *Sensors, 20*(9), 2734.
3. Engeland, A., Bjørge, T., Søgaard, A. J., & Tverdal, A. (2003). Body mass index in adolescence in relation to total mortality: 32-year follow-up of 227,000 Norwegian boys and girls. *American Journal of Epidemiology, 157*(6), 517–523.
4. Colmenarejo, G. (2020). Machine learning models to predict childhood and adolescent obesity: A review. *Nutrients, 12*(8), 2466.
5. Freedman, D. S., Mei, Z., Srinivasan, S. R., Berenson, G. S., & Dietz, W. H. (2007). Cardiovascular risk factors and excess adiposity among overweight children and adolescents: The Bogalusa heart study. *The Journal of Pediatrics, 150*(1), 12–17.
6. Dugan, T. M., Mukhopadhyay, S., Carroll, A., & Downs, S. (2015). Machine learning techniques for prediction of early childhood obesity. *Applied Clinical Informatics, 6*(3), 506–520.
7. Anand, V., Biondich, P. G., Liu, G. C., Rosenman, M. B., & Downs, S. M. (2004). Child health improvement through computer automation: The CHICA system. *Medinfo*, 187–191.
8. Jindal, K., Baliyan, N., & Rana, P. S. (2018*). Obesity prediction using ensemble machine learning approaches*. In Proceedings of the 5th ICACNI conference (Vol. 2, pp. 355–362). Springer.
9. Dunstan, J., Aguirre, M., Bastías, M., Nau, C., Glass, T. A., & Tobar, F. (2020). Predicting nationwide obesity from food sales using machine learning. *Health Informatics Journal, 26*(1), 652–663.
10. Adnan, M. H. B. M., & Husain, W. & Abdul Rashid, N. (2012, June). *A hybrid approach using Naïve Bayes and Genetic Algorithm for childhood obesity prediction*. In Proceedings of 2012 ICCIS (Vol. 1, pp. 281–285). IEEE.
11. Cervantes, R. C., & Palacio, U. M. (2020). Estimation of obesity levels based on computational intelligence. *Informatics in Medicine Unlocked, 21*, 100472.
12. Ferdowsy, F., Rahi, K. S. A., Jabiullah, M. I., & Habib, M. T. (2021). A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences, 2*, 100053.
13. Obesity Levels & Life Style | Kaggle. Available at: https://www.kaggle.com/code/mpwolke/obesity-levels-life-style/data
14. Palechor, F. M., & de la Hoz Manotas, A. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. *Data in Brief, 25*, 104344.

15. Amiri, F., Yousefi, M. R., Lucas, C., Shakery, A., & Yazdani, N. (2011). Mutual information-based feature selection for intrusion detection systems. *Journal of Network and Computer Applications, 34*(4), 1184–1199.
16. Feature selection – Part I: Univariate selection. Available at: https://blog.datadive.net/selecting-good-features-part-i-univariate-selection/

A Clinical Decision Support System Using Machine Learning to Forecast the Risk of Chronic Pulmonary Disease and Anthracosis

Ritu Aggarwal

Abstract Several diseases arise influencing the life span of people. There are many diseases such as lung disease, diabetes, chronic obstructive pulmonary disease (COPD), lung cancer, breast cancer, and anthracosis. Medical professionals have a greater challenge of effectively forecasting and identifying COPD early on. One significant cause of worldwide death is COPD. It is a respiratory disease due to the accumulation of carbon in the lungs. İn this proposed work, datasets have 430 instances and 20 attributes. The total sample used is 4000 in which 2000 samples were analyzed and were found to have no disease, while the other samples have a common disease problem. In this paper, the system with supporting clinical decisions for affected patients that are suffering from COPD and anthracosis is studied. To detect and diagnose COPD at the early stage, different machine learning classifiers are used such as support vector machine, random forest, decision tree, Naïve Bayes, and artificial neural network with best results in terms of accuracy. The accuracy achieved by the support vector and random forest is 96.866% and 92.678%.

1 Introduction

Chronic obstructive pulmonary disease is a meticulous respiratory condition that is due to chronic airflow limitation. COPD is considered as the third cause of death in the world in terms of pervasiveness and mortality. In the year 2010, more than three million people were affected by this disease and they all died [12]. In the USA alone, the operating cost of this disease is 50 billion dollar annually. Nowadays, at its pervasiveness, the costs related to COPD increased more [1, 11]. Due to increase cost for operating and treatment, the rate of severity for admitted hospitalization patients is less due to growth in treatment rate. With the help of CDSS, home

---

R. Aggarwal (✉)
Seth Jai Prakash Mukand Lal Institute & Technology, Radaur, Haryana, India

Maharishi Markandeshwar Engineering College, Ambala, Haryana, India

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_4

telemonitoring symptoms and managements of COPD patients become easy. The main ideas behind CDSS are to easily monitor patient health and their severity irrespective of COPD on a regular basis [10]. At in early stages detection and prevention of the exacerbation with the CDSS monitoring system the costs are reducing in the hospital admissions and with antibiotics and corticosteroids improvement of patient's quality of life are improved. Exacerbations are the main cause of COPD for severe hospitalization [30].The main reasons of increasing rate of COPD are smoking, environmental changes, increased age, and population rate [16]. With the help of spirometry, the detection and diagnosis of COPD become easy. Respiratory problems are the early sign of COPD [3] exacerbation. The previous algorithms that have been implemented are based on the international definitions which need immediate clinical alerts. In 2007, according to the WHO, the estimated number of deaths due to anthracosis is approx. 250,000. Environmental pollution, dust, and carbon are the most important factors for spreading out of this disease [5–7]. For the last 20 years, COPD is main leading cause of death. This study focuses on the important factors of COPD and reasons that portray the classify of Anthracosis. Machine learning algorithms help in the early prediction and diagnosis of COPD and anthracosis [4, 13] . The following are the major symptoms of anthracosis and chronic obstructive pulmonary disease: breathing difficulty and sputum color and quantity with minor symptoms like cold, nausea, stinging throat, cough, and fever. Symptom measures according to number of key attribute of a patient present on that is nKey, same as minor symptoms and by the combination of two evaluate the total symptom measures [2]. Section 1 discusses the introduction of COPD; Sect. 2, literature review; Sect. 3, methods and tools; Sect. 4, predictive classification models; Section 5, performance evaluation matrices; and Sect. 6, conclusion and references.

1.1 Research Gap and Objective

The research gap and objective for this current study is improved accuracy and better prediction of disease at early stages. To identify and classify these diseases, medical or health sciences are more effective to use. Its the various diseases the Anthracosis and COPD (is also known as chronic bronchitis). Anthracosis is due to massive fibrosis of the lungs resulting in excessive amount of carbon and is an unceasing seditious of the airways [8, 9]. According to previous studies, 100 million people were affected by anthracosis disease.

2 Literature Review

Iadanza, E et al. [1] develop a CDSS based classification model that tests some respiratory functions; analyzer tool is used to get accurate results in testing COPD. Siddhi Vora1 et al. [2] in this work proposed different machine learning algorithms such as support vector machine and KNN to obtain better predicted results. It obtained an accuracy of 96.97% for SVM and 92.30% for KNN. Swaminathan S et al. [3] in this study for detecting the COPD disease self identification machine is design and developed by using various machine learning algorithms and obtained results with higher accuracy. Woldeamanuel, G et al. [4] proposed a model to identify the disease risk factors at the early stages of COPD. If patients have high tobacco intake, the chances of disease occurring are more which will affect the lungs.

X Ma et al. [5] in this study develop a new prediction model using the SNP model and other machine learning algorithms. The author identifies the disease by the clinical attributes of affected patients. Peng, J et al. [6] used the technique for predicting the disease and obtained accuracy by identifying 28 clinical attributes of patient that show the disease risk factors.

3 Methods and Tools

The proposed methodology is applied to achieve the following goals:

1. Evaluation of the capability of the dataset presented for the classification of the respiratory function test results.
2. The results evaluated based on to identify the predictive model that outperforms the various performance metrics.
3. Implementation and comprehensive study are based on the well-known machine learning algorithms.
4. Public and real dataset is used for anthracosis and chronic obstructive pulmonary disease as shown in Fig. 1.

3.1 Dataset

In this study, three datasets are to be taken for anthracosis disease groups, COPD groups, and healthy groups. Each dataset has 20 separate descriptive variables and the total sample is 4000. The dataset has 430 attributes. The number of persons who have no disease is 2000, and other samples are considered for common disease. With the help of spirometer, the medical measurements are obtained. It is a device that is used to provide the inhalation volume. The predicted percentage according to age group and male and female number depends upon the absolute value. The absolute value in pulmonary function is measured in ML. The anthracosis disease group and COPD group of patients were evaluated based on the basal pulmonary function test values. For a person suffering from anthracosis and COPD disease, the properties, attributes, and their details have been considered as shown in Tables 1 and 2 (ErkutBolat et al.).

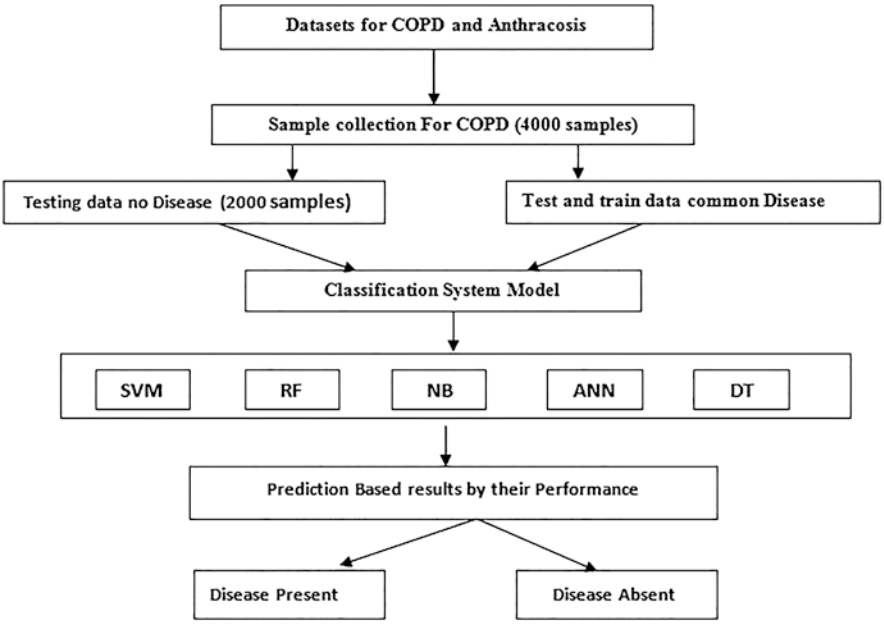


Fig. 1 Proposed methodology

Table 1 Dataset for COPD and anthracosis [1]

Sr. no.	Attribute name	Value	Description
1.	Age (AG)	Definite	1, 0–14; 2, 15–29; 3, 30–44; 4, 45–59; 5, 60–74; 6, 75+
2.	Sex (SX)	Definite	Male or female
3.	Emergency (EG)	Arithmetic	How many emergency admission
4.	Outpatient (OP)	Arithmetic	How many outdoor patient admission
5.	FEV1	Arithmetic	Forced expiratory for 1 s
6.	FVC1	Arithmetic	Forced vital capacity
7.	PEF1	Arithmetic	Peak expiratory flow
8.	MEF (75)	Arithmetic	75% of maximal expiratory flow of the FVC curve
9.	MEF (50)	Arithmetic	Maximal expiratory flow = 50% of the FVC curve
10.	MEF25	Arithmetic	25% of maximal expiratory flow of the FVC curve
11.	MMEF1 max	Arithmetic	Maximal-mid expiratory flow
12.	FEV3	Arithmetic	For 3 s, forced expiratory volume
13.	PIF	Arithmetic	Peak in respiratory flow
14.	TIFF	Arithmetic	Tiffeneau-Pinelli index (FEV1/FVC)
15.	Disease (anthracosis)	Definite	Yes/no
16.	Disease (COPD)	Definite	Yes/no

Table 2 Different diseases and their description

Name of disease and description		(Effected with low risk)	(Effected with high risk)
Number of cases		220 (51.16%)	210 (48.83%)
Sex	Male	170 (79.4%)	180 (90.1%)
	Female	54 (29.9%)	46 (26.9%)
Smoking history		149 (87.5%)	187 (94.7%)
Age		89 ± 9	92 ± 9
Hospitalization cases		6.8 ± 4.8	7.5 ± 8.5
Temperature (TEMP)		36.8 ± 0.5	36.7 ± 0.6
Pulse rate (PULSE)		82.5 ± 14.6	88.1 ± 16.6
Respiratory rate (RES)		22.7 ± 4.1	34.5 ± 7
Systolic pressure (SP)		123.1 ± 18.3	142.1 ± 34.7
Diastolic pressure (DP)		56.1 ± 10.2	74.6 ± 11.9
Pulmonary heart disease (PHD)	With	26 (16.3%)	75 (42.2%)
	Without	162 (92.7%)	147 (69.9%)
Bronchiectasis (BRCH)	With	15 (8%)	9 (5%)
	Without	174 (83%)	184 (89%)
Hypertension (HTN)	With	76 (41.4%)	80 (40.6%)
	Without	120 (58.8%)	120 (60.2%)
Diabetes mellitus (DM)	With	18 (9.2%)	42 (21.3%)
	Without	181 (81.9%)	151 (89.6%)
Coronary heart disease (CHD)	With	31 (11.1%)	46 (16.9%)
	Without	177 (79.7%)	156 (72.3%)
Chronic kidney disease (CKD)	With	6 (4%)	56 (4%)
	Without	104 (88%)	156 (89%)
Malignant tumor (MT)	With	17 (9%)	30 (20%)
	Without	181 (82%)	172 (80%)
Cerebrovascular disease (CD)	With	7 (3%)	9 (4%)
	Without	176 (76%)	184 (86%)
Viral hepatitis (HBV)	With	4 (2%)	5 (3%)
	Without	204 (94%)	187 (87%)
Cirrhosis (CRHS)	With	2 (1%)	3 (2%)
	Without	107 (89.4%)	305 (99.1%)

There is less number of patient's records because some of patient's data were taken from the follow-up records. According to the clinical records and the estimations as indicated by the physiological boundaries by the respiratory capacity test and diffuse limit of lung against the carbon monoxide (DLCO) Test. The distinctive physiological boundaries considered for every quiet were as: Forced Expired Volume in 1 s (FEV1), Forced Vital Capacity (FVC), Slow Vital Capacity (SVC), FEV1/FVC proportion, FEV1/SVC proportion, Forced Expired Flow at 25–75% (FEF 25–75), Peak Expiratory Flow (PEF), Vital Capacity (VC), Total Lung Capacity (TLC), Residual Volume (RV), Functional Residual Capacity (FRC), Expiratory Reserve Volume, DLCO, Alveolar Volume (VA) and DLCO/VA. These tests are estimated when given the nebulizer.

4 Predictive Classification Models

The above data are analyzed using the Python Jupyter Notebook. It is a very useful platform that enables the outcome of predictive models, without much more configuration. The main objective of creating the prediction models depends on the input parameters in respect to their output parameters: point of respiratory dearth, diffusing capacity of the lungs for carbon monoxide, and possibility of exacerbation. The models used to get the experiment results in the case of anthracosis and COPD are decision tree, support vector machine, Naive Bayes, random forest, artificial neural networks, etc. The performance evaluation metrics are used such as accuracy, precision, F1-score, recall, kappa, etc.

4.1 Artificial Neural Networks (ANN)

It is neural networks that works like a human brain and provides solutions in respect to problems given. It is very difficult but the flexible approach used for ANN is no limitations on the variable composition. ANN consists of three basic elements: input layer, hidden layer, and output layer. ANN have nodes and their connections could be combined in different ways. The hidden layer has two or more layers [6]. The directions of connections between nerves are within network. Basically ANN works on approach such as back propagation algorithm.

4.2 Support Vector Machine (SVM)

SVM is a machine learning algorithm for solving the classification and regression problems [11]. The hyperplane is used separating the marginal lines. The primary beginning stage of the help vector machines is to attempt to play out the characterization by picking the highest line whose edge is out of a limitless number of lines. The edge line framed ought to be corresponding to the individuals from the class individuals nearest to the chose line. The lines drawn are called hyperplanes.

4.3 Decision Tree (DT)

In the calculation utilizing decision tree to characterize the information as indicated by certain variable qualities, the sources of info and yields are the deciding factors of the information. Decision trees are quite possibly the most utilized calculations in order issues, a calculation created to partition an information base containing an enormous number of information utilizing a few procedures [15].

4.4 Random Forest

RF is a type of machine learning algorithm which is used to solve the classification and regression problems. Random forest used the approach of bagging when applying splitting approach in each tree. In bagging approach, the selected features are to be used. RF approach computed the relevant and consistent number of variables to be selected. RF is based on the vote classifier. It is a hierarchical structure of nodes.

4.5 Naïve Bayes

Naïve Bayes is a classification algorithm. NB works on Bayes theorem. The common principles used for NB are implementing the individual pair of features which is dependent on each other [14].

5 Performance Evaluation Metrics

During the implementation of machine learning algorithms at the training process, the minimum error and maximum accuracy are analyzed by the hyper parameters that are kept as final implementation parameters when the final model is tested on the testing data. The various performance metrics are to be taken: accuracy, recall, kappa, precision, F1-score, etc. These metrics could be obtained by their individual confusion matrix. In the confusion matrix, the predicted and actual class of observations can be considered. The confusion matrix is given in Table 3.

$$\text{Accuracy} = \frac{True_{\_Pos} + \text{True}\_Neg}{True_{\_Pos} + \text{false}_{\_Pos} + True_{Neg} + \text{False}\_neg} \tag{1}$$

$$\text{Precision} = \frac{\text{True}\_Pos}{True_{Pos} + \text{False}\_Pos} \tag{2}$$

$$\text{Recall} = \frac{\text{True}\_Pos}{True_{Pos} + \text{False}\_Neg} \tag{3}$$

Table 3 Confusion matrix prediction

		Corrected/actual predictions	
		Positive (Pos)	**Negative (Neg)**
Prediction	Positive (Pos)	True positive (True_pos)	False positive (False_Pos)
	Negative (Neg)	False negative (False_Neg)	True negative (True_Neg)

$$\mathrm{F1-score} = \frac{2*\mathrm{Recall}*\mathrm{precision}}{\mathrm{Recall}+\mathrm{Precision}} \tag{4}$$

The results of mentioned algorithms are used for both COPD and anthracosis datasets by which the training and testing were performed. The various performance metrics such as accuracy, precision, recall, and F1-score have been implemented, and their corresponding results for both diseases were presented, respectively. In the proposed methodology, firstly the dataset and their samples for both of diseases are taken, while in the second step, a new system was developed by using machine learning classifier and train and test the data accordingly [17]. The performance evaluation metrics provide the results in terms of their accuracy and metrics that they have as shown in Tables 3 and 4. The proposed methodology used is shown in Fig. 1 and results in Fig. 2.

According to the given table, for COPD and anthracosis, the best results in terms of accuracy rate are achieved by the SVM and RF classifier. The accuracy obtained for SVM is 96.866% and 92.678% for RF.

5.1 *Descriptive Information for Datasets Used for COPD and Anthracosis*

With the help of machine learning algorithms, the dataset and their related information could be understood easily. From the given table, the descriptive information of disease related patients is mentioned. The age is selected for 0–75+. No any kind of imbalance data and information is found there. The range and ratio is describing for COPD and anthracosis are be adjusted as shown in Table 5 and the descriptive results for different value for COPD and Antracosis is shown in Table 6.

The above table shows the COPD and anthracosis control data results using different machine learning classifier. It gives best performance results with the help of evaluation metrics such accuracy, recall, F1-score, precision, etc. as per results in Table 6. The accuracy obtained for SVM is 96.866% and 92.678% for RF. It gives best results as compared to other classifiers. When this model system tested and train with other classifiers give not best results in terms of accuracy. In order to evaluate and determine the significant factors, SVM and random forest are best for both COPD and anthracosis. The above table describes the importance of

Table 4 Description analysis for COPD and anthracosis

Algorithm	Accuracy	Precision	Recall	F1-score	Kappa
ANN	87.567	78.557	70.567	84.567	70.437
SVM	96.866	86.866	66.755	96.78	86.866
DT	89.789	79.689	79.759	83.789	89.789
NB	80.567	70.467	80.327	80.567	80.567
RF	92.678	82.5481	81.05	92.067	92.216

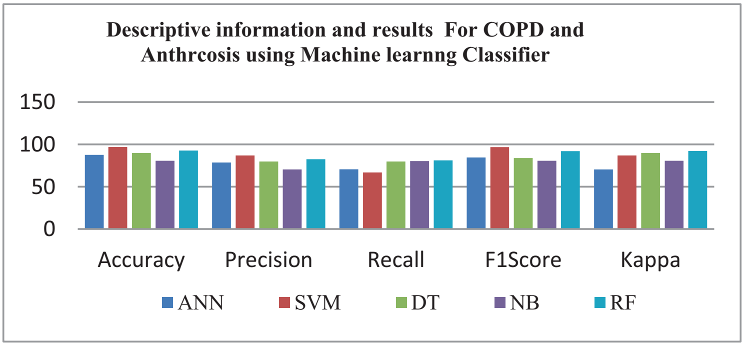


Fig. 2 Descriptive information and results for COPD and anthracosis using machine learning classifier

Table 5 Dataset range for COPD and anthracosis

	Disease group	Frequency (%)	Percentage (%)
Age group	Control	1	0.04
0–15	Anthracosis	0.5	0
	COPD	0	0
	Control	12	0.25
16–40	Anthracosis	10	0.24
	COPD	7	0.3
	Control	110	4.8
40–50	Anthracosis	89	3.73
	COPD	48	2.5
	Control	341	11.06
50–60	Anthracosis	339	11.1
	COPD	350	12
	Control	382	14.4
60–75	Anthracosis	563	25.88
	COPD	488	26.27
	Control	233	5.99
75+	Anthracosis	130	5.37
	COPD	174	7.27
	Control	658	8.93
Male	Anthracosis	569	29.63
	COPD	699	22.97
	Control	572	25.6
Female	Anthracosis	405	13.5
	COPD	318	32.34
	Control	2000	44.44
Disease	Anthracosis	2000	44.44
	COPD	2000	44.44

Table 6 Descriptive results for different values for COPD and anthracosis

		Mean value	SE	Min value	Max value	Range	Standard deviation	Variance
FEV1	Control	86.52	.46	26.3	157	129.5	16.52	410.61
	Anthracosis	83.81	.48	42.8	128.6	88.7	8.98	83.37
	COPD	71.46	.25	55.6	95.7	45.9	7.99	66.07
FVC1	Control	86.66	.72	19.8	162.6	143.8	20.59	483.6
	Anthracosis	77.89	.39	36.99	139.4	95.34	12.08	148.8
	COPD	69.08	.39	48.9	120.4	69.12	12.86	148.32
FEV1M	Control	99.82	.48	60	147.9	88.99	14.29	192.14
	Anthracosis	89.09	.8	55.32	144.98	89.9	19.99	386.65
	COPD	92.9	.46	55.6	145.6	83.9	17.17	250.08
PEF	Control	66.15	.64	18.6	135.6	121.9	23.07	404.54
	Anthracosis	62.29	.65	22.9	132.3	105.5	17.35	286.46
	COPD	52.7	.45	18.1	116.8	98.8	25.66	261.99
MEF75 (maximal expiratory flow)	Control	69.26	.69	6.86	149.2	143.38	28.38	420.38
	Anthracosis	56.25	.63	15.6	155.8	131.9	17.86	298.88
	COPD	48.06	.7	15.8	121.7	97.8	16.9	256.66
MEF50 (maximal expiratory flow)	Control	68.59	.77	14.08	174.7	157.29	25.39	550.8
	Anthracosis	54.5	.59	12	138	128	19.86	371
	COPD	38.87	.59	9.6	104.67	98.44	25.24	240.39
MEF25 (maximal expiratory flow)	Control	68.08	.25	10	245	230.02	46.44	1425.27
	Anthracosis	56.09	.97	9.02	230.64	211.53	28.96	887.44
	COPD	45.57	.79	5.91	240.8	235.81	25.68	620.5
MMEF (maximal-mid expiratory flow)	Control	65.64	.77	25.35	165.8	150.35	24.79	624.92
	Anthracosis	69.52	.88	12.9	148.9	170	28.69	585.46
	COPD	59.9	0.88	18	178.9	157.6	47.88	789.58
FEV3 (forced expiratory volume)	Control	98.97	.14	89.67	110.99	14.45	4.58	13.68
	Anthracosis	95.74	.19	66.55	110.99	46.48	5.29	27.85
	COPD	95.38	0.19	74.89	110.98	26.09	5.26	27.62
PIF (peak inspiratory flow)	Control	99.6	.07	24.78	225.67	192.88	35.79	1088.98
	Anthracosis	87.73	.07	22.88	225.62	195.74	35.37	1109.29
	COPD	88.27	.87	56.58	182.37	137.52	27.65	809.21
TIFF (Tiffeneau-Pinelli index)	Control	108	.47	63.33	138.19	78.96	14.04	134.6
	Anthracosis	98.46	.42	58.08	134.57	86.01	12.35	160.08
	COPD	95.9	.55	57.04	139.58	85.51	12.19	230.95

(continued)

Table 6 (continued)

		Mean value	SE	Min value	Max value	Range	Standard deviation	Variance
Emergency	Control	1	0	0	35	35	2	8
	Anthracosis	1	0	0	25	25	1	6
	COPD	2	0	0	300	300	18	180
Outpatient	Control	6	0	0	54	54	7	38
	Anthracosis	5	0	0	80	80	3	30
	COPD	6	0	0	40	40	2	15

variables for both the diseases. The highest value describes the higher importance. According to medical test analysis, the various coefficients and test analysis were implemented.

6 Conclusion and Future Works

A review of COPD and the clinical-choice emotionally supportive networks explicitly applied to COPD is depicted in the writing. In particular, an examination was made of the calculations most generally used to make predictive models on which current choice emotionally supportive networks are based. At that point, the physiological boundaries generally used to prepare these models to make forecasts on the objective boundaries were distinguished. When these features were distinguished, an endeavor was made to reproduce the outcomes, regarding accuracy and precision, of the two frameworks found: the random forest and the support vector machine. In this manner another predictive model dependent on random forest and decision tree on which machine learning calculation was created. The random forest gave preferred execution over the two past models. Early detection of COPD and anthracosis is a very big challenge. The machine learning approaches when applying during testing and training of complexity level of COPD and Anthracosis using the RF,DT,SVM,ANN,NB. To effectively calculate the results, more dataset value is required. In this work developed a system by which the early detection and prediction of COPD and anthracosis disease is effectively resulted. The outcome of this system used for both diseases is giving best results in terms of accuracy. The results obtained for SVM is 96.866% and 92.678% for RF. As compared to other work, that is best results. In the future, this work can be enhanced by using more dataset values and developing new system.

References

1. Iadanza, E., Mudura, V., & Melillo, P. (2020). An automatic system supporting clinical decision for chronic obstructive pulmonary disease. *Health Technology, 10*, 487–498. https://doi.org/10.1007/s12553-019-00312-9

2. Vora, S., & Shah, C. (2019). COPD classification using machine learning algorithms. *International Research Journal of Engineering and Technology, 6*, 608–611.
3. Swaminathan, S., Qirko, K., Smith, T., Corcoran, E., Wysham, N. G., Bazaz, G., et al. (2017). A machine learning approach to triaging patients with chronic obstructive pulmonary disease. *PLoS ONE, 12*(11), e0188532. https://doi.org/10.1371/journal.pone.0188532
4. Woldeamanuel, G. G., Mingude, A. B., & Geta, T. G. (2019). Prevalence of chronic obstructive pulmonary disease (COPD) and its associated factors among adults in Abeshge District, Ethiopia: A cross sectional study. *BMC Pulmonary Medicine, 19*, 181. https://doi.org/10.1186/s12890-019-0946-z
5. Ma, X., Wu, Y., Zhang, L., Yuan, W., Yan, L., & Fan, S. (2020). Comparison and development of machine learning tools for the prediction of chronic obstructive pulmonary disease in the Chinese population. *Journal of Translational Medicine*.
6. Peng, J., Chen, C., Zhou, M., et al. (2020). A machine-learning approach to forecast aggravation risk in patients with acute exacerbation of chronic obstructive pulmonary disease with clinical indicators. *Scientific Reports, 10*, 3118. https://doi.org/10.1038/s41598-020-60042-1
7. Global Initiative for Chronic Obstructive Lung Disease (GOLD). (2015). GOLD. http://www.goldcopd.org/
8. Pauwels, R. A., Buist, A. S., & Calverley, P. M. (2001). Global strategy for the diagnosis management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO global initiative for chronic obstructive lung disease (GOLD) workshop summary. *American Journal of Respiratory and Critical Care Medicine, 163*(5), 1256–1276.
9. Wilkinson, T. M. A., Donaldson, G. C., Hurst, J. R., & Seemungal, T. A. (2004). Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine, 169*(12), 1298–1303.
10. Chandra, D., Tsai, C. L., Camargo, C. A., & Acute Exacerbations of COPD. (2009). Delay in presentation and the risk of hospitalization. *Chronic Obstructive Pulmonary Disease, 6*(2), 95–103.
11. Aggarwal, R., & Kumar, S. (2022). An enhanced fusion approach for meticulous presaging of HD detection using deep learning. In *2022 IEEE international conference on distributed computing and electrical circuits and electronics* (ICDCECE), pp. 1–4. https://doi.org/10.1109/ICDCECE53908.2022.9793141
12. Nunavath, V., Goodwin, M., Fidje, J. T. & Moe, C. E. (2018). Deep neural networks for prediction of exacerbations of patients with chronic obstructive pulmonary disease. In *EANN*, pp. 217–228.
13. Aggarwal, R., Podder, P., & Khamparia, A. (2022). ECG classification and analysis for heart disease prediction using XAI-driven machine learning algorithms. In *Biomedical data analysis and processing using explainable (XAI) and responsive artificial intelligence (RAI)* (Intelligent systems reference library) (Vol. 222). Springer.
14. Aggarwal, R., & Thakral, P. (2022). Meticulous presaging arrhythmia fibrillation for heart disease classification using oversampling method for multiple classifiers based on machine learning. In *Advances in data computing, communication and security* (Lecture notes on data engineering and communications technologies) (Vol. 106). Springer. https://doi.org/10.1007/978-981-16-8403-6_9
15. Aggarwal, R., & Kumar, S. (2022, March). An automated perception and prediction of heart disease based on machine learning. In *AIP conference proceedings* (Vol. 2424, No. 1, p. 020001). AIP Publishing LLC.
16. Sugumaran, V., Muralidharan, V., & Ramachandran, K. (2007). Feature selection using Decision Tree and classifcation through Proximal Support Vector Machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing, 21*(2), 930–942.
17. Aggarwal, R., & Kumar, S. (2023). Classification model for meticulous presaging of heart disease through NCA using machine learning. *Evolutionary Intelligence*. https://doi.org/10.1007/s12065-023-00830-6

Smart Healthcare: A Breakthrough in the Growth of Technologies

Divya Singh, Deepti Singh, Manju, and Umesh Gupta

Abstract Artificial intelligence (AI) technologies have significantly grown in recent years and have become a reality in many aspects of our everyday lives. Various attempts are being made in the healthcare area to integrate AI for practical medical treatments. With the fast advancements in machine learning algorithms and hardware performance, AI technology is projected to play a significant role in properly evaluating and using large volumes of health and medical data. However, AI technology has some distinct properties that set it apart from contemporary healthcare technologies. As a result, several areas within the present healthcare system must be supplemented for AI to be used more successfully and often in healthcare. Furthermore, the number of medical practitioners and general persons who embrace AI in healthcare is still low; additionally, there are different difficulties about the safety and dependability of AI implementations. As a result, the purpose of this article is to describe the current research and application status of AI technology in healthcare, as well as to explore the challenges and issues that must be addressed.

D. Singh (✉)
School of Computer Science Engineering and Technology, Bennett University, Noida, India

D. Singh
Computer Engineering Department, NSIT, New Delhi, India

Manju
Department of Computer Science and Engineering & Information Technology, Jaypee Institute of Information Technology, Noida, India

U. Gupta
School of Computer Science and Artificial Intelligence, S R University, Warangal, Telangana, India

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_5

1 Introduction

The emphasis on medicine has changed in recent years from curing diseases to preventing them. Consistent with these shifts, hospitals have implemented AI, robots, and other Fourth Industrial Revolution-related technology, leading to optimistic hopes for better disease detection and treatment. There is a growing demand to lower national medical costs by establishing networks among hospitals, clinics, pharmacies, and rehabilitation centers; incorporating cutting-edge information and communication technologies; and offering patients access to a wide range of health information from which to make informed treatment decisions [1].

New opportunities in digitalization and data collection have opened the door to the possibility of informed digital transformations in traditional biotechnology and medicine. Smart healthcare integration has led to the development of novel data processing methods. Transforming into a "smart" healthcare system involves several steps. Smart healthcare systems analyze patient behavior by implementing changes such as refocusing on the individual rather than the disease, rethinking the role of the healthcare provider in disease prevention, and moving away from a focus on regional to one that is more clinically focused in terms of information creation [2–4]. With the use of smart healthcare technologies, such as real-time vital sign monitoring systems, improved diagnostic tools can deliver cutting-edge care for patients. The main aim of "smart healthcare" is to aid patients by giving them information about medical problems and their remedies. So, they are better prepared to respond during emergency situations, thanks to smart healthcare [1].

The key technological component of AI-based services for illness detection and prognosis is intelligence (the capacity for learning, reasoning, perception, and comprehension). Increased efficacy, efficiency, and safety of therapy may be the result of this aid to clinical decision-making by doctors during diagnosis and medication administration. Beneficial for both patients and doctors, AI-based healthcare systems enhance illness prognosis, treatment, and monitoring [7]. There are four ways in which the application of AI to the medical field can help doctors, patients, and healthcare providers: i. by estimating the likelihood of treatment success and analyzing disease onset prior to treatment initiation, ii. by managing complications, iii. by actively supporting patient care during diagnosis and/or treatment, and iv. by identifying the pathology of the disease for the best treatment [8]. The technologies used led to advancements in the transport of data, interoperability of data, exchange of information, and communication between machines. Better use of technology and communication protocols enhanced planning and allocations, leading to fewer unnecessary patient visits [4]. To sum up, smart healthcare enables professionals to have timely and accurate access to the information and solutions they need to decrease medical mistakes and save costs [5].

Table 1 shows the review of different researchers related to smart healthcare. Our contribution is to deliver a review of smart healthcare with a different perspective that includes taxonomy for smart healthcare (shown in Fig. 1), covering communication technologies, network types, requirements, characteristics, applications, services, open issues, and challenges.

Table 1 Survey on smart healthcare

References	Author contribution
Qi et al. [9]	In this paper, researchers have reviewed smart healthcare from the perspective of monitoring heartbeat, blood pressure, and oxygen saturation
Islam et al. [10]	In this paper, the focus was on architecture, technology, and platform that enable reception and transmission of medical data
Baker et al. [11]	This paper has designed a model for special condition monitoring and general system. They also presented an overview of wearable component and blood pressure sensor monitoring along with review on long- and short-range communication
Mahmoud et al. [13]	In this researcher, researchers have implemented cloud of things in smart healthcare and its efficiency
Dhanvijay et al. [13]	In this paper, the focus was on healthcare application for wireless body area network
Ahad et al. [14]	In this paper, researchers have done a survey on scheduling and routing to attain objective of smart healthcare
Pratap et al. [15]	This author used medical technologies to treat orthopedic patients
Al Turjman et al. [16]	In this paper, the main focus is on developing a generic IoT framework comprising of data collection, communication gateways, and server clouds
Jagathee saperumal et al. [17]	This paper has done a survey on various technologies that can provide security services in healthcare systems
Alabdulatif et al. [18]	This paper explored security and privacy solutions in smart healthcare systems
Wazid et al. [19]	The main focus in this paper was to develop a secure framework to detect ransomware in smart healthcare systems using machine learning
Zia et al. [20]	The researchers have proposed a lightweight authentication mechanism for wearable devices
Kong et al. [21]	The researchers have used oversampling with AI models and improved the security of Android wearable applications

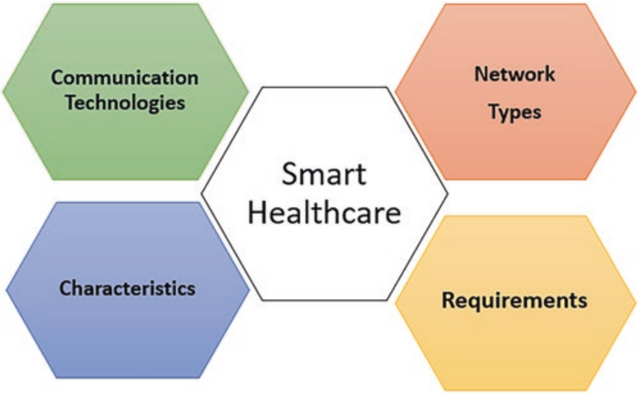


Fig. 1 Taxonomy of smart healthcare

2 Taxonomy of Healthcare

2.1 *Communication Technology*

The ability to send and receive data across both short and long distances is crucial to the operation of many different types of smart healthcare systems. Technologies like Bluetooth, ZigBee, and Wi-Fi are the most well-known examples of short-range wireless technology used in smart healthcare systems like the body area network (BAN). Long-distance technologies like LTE and WiMAX are employed in the smart healthcare system to transmit data from the local server to the BS. Moreover, LTE-M is seen as an improvement to help IoT. Nevertheless, 3GPP must make more enhancements to tackle coverage, battery life, and device complexity [6]. To guarantee compatibility across different service providers, LoRa has standardized the LoRa-WAN protocol for use in smart healthcare. Additionally, SIGFOX provides a highly flexible global network by considering low-power-usage smart healthcare applications.

2.2 *Network Types*

Topologies play significant role in efficient operation of IoT-based healthcare systems. Several short-range services may be accessed using IoT networks. WLANs, BANs, and WPANs are all employed in indoor healthcare applications, for instance. MANs, WANs, and mobile communication networks are used to provide e-health services to patients on the go. Several unique features of the systems under discussion include size, data, coverage, delay, and coverage requirements of structures.

2.3 *Requirements*

Requirements for smart healthcare may be broken down into functional and nonfunctional categories, as shown in Fig. 3. A technical specification is a document that details the design of healthcare. As an example, the temperature monitoring system is structured according to the purpose for which it is used, the temperature range over which the thermometer or thermistor operates, and the method by which temperatures are monitored.

It's possible that the data collecting cycle and the typical interval between operations will be different. Thus, functional needs are context-dependent, depending on the specific segment used in that smart healthcare system.

On the contrary, nonfunctional requirements tend to be vague. This criterion suggests capabilities that may be used to address problems in the healthcare system. A more broad perspective would classify the nonfunctional necessities of healthcare as falling under the headings of ethical and performance requirements. It is necessary to characterize performance requirements in terms of hardware and software requirements due to the vast number of obstacles involved in designing a comprehensive intelligent healthcare setup. The fundamental requirements for the intelligent healthcare technique are system dependability, higher efficiency, a small form factor, low power, quality of service, an improved client experience, a reputation for consistently helpful assistance, the adaptability of the method to move up to more recent forms and advances, and widespread availability.

As a result, the first and foremost consideration of intelligently designed healthcare is to ensure that medical services can be provided promptly. Besides these requirements, the building also needs ambient intelligence to improve the nature of the facility for cutting-edge applications. Researchers and businesses can have quite different perspectives on what constitutes "smart healthcare," depending on the desired outcome. Components of an intelligent healthcare system may be classified according to their role as sensors or actuators, processing devices, information storage components, or network components. The term "sensor" refers to a sensing device that works in tandem with a living thing to detect changes in the environment [22].

Depending on the kind of monitoring system in place, different sensors or actuators may be used. Common sensors used in a smart healthcare system include electromyography (EMG), electrocardiography (ECG), temperature sensors, sensors for oxygen saturation (SpO2), accelerometers, orientation sensors, motion sensors, and heart rate sensors. Many kinds of computers, from simple handhelds like PDAs and smartphones to very complicated machines like servers and supercomputers, are in use today. Being the most crucial part of the systems, data storage plays a pivotal role in intelligent healthcare. The implanted memory on the detecting devices and the enormous servers used to handle big data analyses are both examples of information repository components in the innovative healthcare network. Sensor nodes, switch nodes, and base stations are all examples of network nodes. The modules become more or less sophisticated depending on the importance of the smart healthcare system being implemented. An intelligent healthcare system mostly worked on the wireless network. As shown in Fig. 2, a variety of wireless technologies enable data interchange across the many different physical nodes that will make up the smart healthcare network.

2.4 Characteristics of Smart Healthcare

Figure 4 depicts the essential components needed to implement the smart healthcare system.

Major needs may be broken down into three categories: objects, apps, and semantics. Appointed architectures are solely responsible for establishing a personalized network in the sensors and the user's computing device, and for information security, to ensure the authenticity of data transfers between apps on smartphones

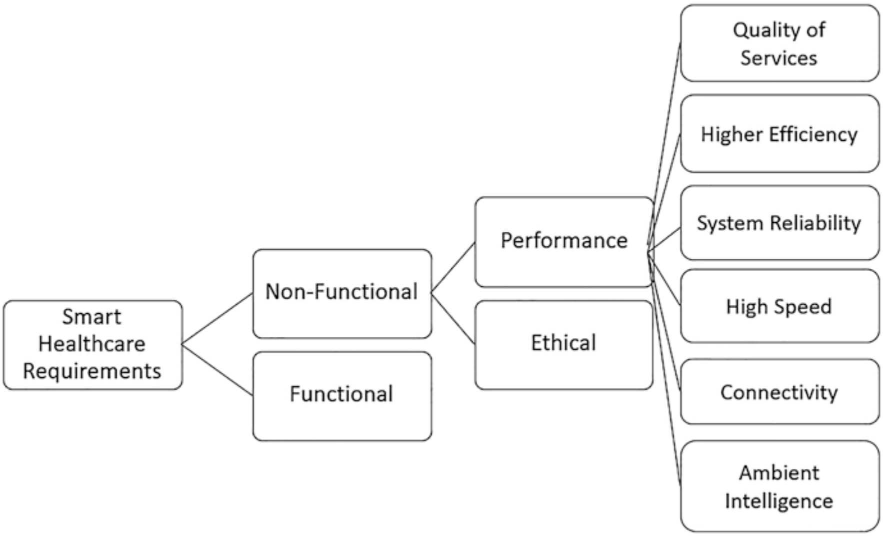


Fig. 2 Requirements of smart healthcare

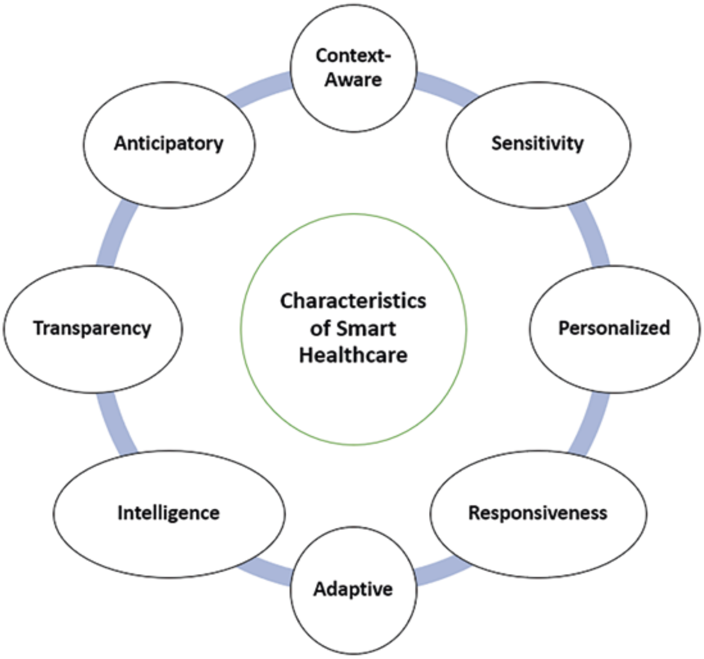


Fig. 3 Characteristics of smart healthcare

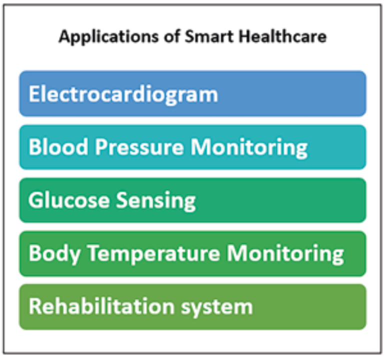


Fig. 4 Applications of the smart healthcare

and sensors. Things-oriented architectures must perform tasks such as real-time monitoring, adaptable implementation, high-level responsiveness, reduced power consumption, increased efficiency, and activation of the intelligent process. Semantic-oriented systems should be able to process natural language execution techniques for enhancing the user experience, developing detectable specimens based on the previously gathered data, and possessing remarkable computational skills [23] (Fig. 3).

3 Applications and Services

3.1 Applications

1. *Electrocardiogram (ECG) Monitoring*: This is a simple test to capture electrical activity of the heart. Checking a patient's cardiac activity on a regular basis using a portable ECG device is convincing and provides long-term cost savings. Since they don't need to make physical touch with the skin, noncontact capacitive coupled anodes are also often used for ECG signal securing, opening the door to their use as wearable sensors implanted in clothing [24]. A method to track ECG readings in real time is suggested in [25]. Edge servers and cloud servers can process the sent ECG signals. Remote access to electrocardiogram (ECG) data is now possible for both physicians and patients, thanks to an Android-based software. The suggested application's ability to offer a dependable and low-latency interface for the patients' ECG has been shown via simulation.
2. *Blood Pressure Monitoring*: It's no exaggeration to say that hypertension is a real risk factor for developing heart attacks or strokes. Although minute-to-minute variations in blood flow are possible, patients benefit from continuous monitoring of blood pressure by medical sensors. The method of measuring and standardizing blood pressure is proposed in [26]. The sensor end and the server end make up the proposed system. The sensing end includes a blood pressure

monitor and an internal measurement unit (IMU) to offer a physical movement correlation to the data being sensed.

Processing and analyzing the raw sensed data is done on the server end, which also handles network function management. To calibrate the measured blood pressure and IMU data, it is necessary to first extract features from each dataset and then generate a matrix vector from it. The system predicts future blood pressure readings using a regressing algorithm like KNN, SVM, Bayesian, or DT. While doing regression, a root-mean-square error of the original data is submitted. The server system's web application provides users and clinicians with access to graphical representations of analyzed data and the system's operational condition.

3. *Body Temperature Monitoring*: (Point 3) Increased core body temperature is thought to be one of the first signs of many viral infections. When doctors get quick and precise notifications from real-time monitoring and body sensors whenever a patient's temperature rises or falls outside of a predetermined range, they are better able to treat their patients.
4. *Rehabilitation System*: Those who suffer from a physical deficiency might benefit from medicine and rehabilitation since it not only improves their physical condition but also restores their functional independence and sense of self-worth. When it comes to comprehensive rehabilitation, the IoT aids all potential remote meetings. As a result, a sophisticated rehabilitation system is an essential component of modern pharmaceuticals. In this research [27], for instance, the author proposes a smart framework for the at-home rehabilitation of stroke patients. Using plastic bands, four biosensors are fastened to the patient's torso, and a pressure sensor is strapped to the patient's feet in this SmartPants. These sensors allow the framework to ensure that the patient is doing the exercises correctly and to provide feedback to the patient if they are not.
5. *Glucose Sensing*: Diabetes is a group of metabolic illnesses characterized by persistently elevated blood glucose levels (sugar). Keeping tabs on your blood sugar levels may alert you to any fluctuations and provide valuable information for planning your meals, workouts, and medication administration schedule. In [28], the authors offer a strategy for the construction of m-IoT devices that can instantly detect average glucose levels. In this article, we present a method for linking sensors (connected with patients) across the IPv6 network to the industry's leading healthcare providers (Fig. 4).

3.2 Services

1. *Ambient Assisted Living*: For the elderly who need round-the-clock care at home, a smart healthcare service known as ambient assisted living (AAL) is available. The primary goal of AAL is to increase the independence of seniors while also ensuring their safety in their own homes. A patient's ability to live alone may be restored by AAL. It's true that the outcomes of today's AAL care systems for the

elderly at home are very adequate; yet, most of these systems contribute very little to the provision of mobility for their users. Since older individuals may forget to charge their phones, inactive phones may be a contributing factor to the aforementioned problem. Another reason is that a short-range protocol cannot reestablish contact and send the medical data to an unidentified gateway if it is lost between an IoT device and its designated gateway. A technique for detecting urgent situations in AAL is presented in [28]. An Internet of Things sensor using Bluetooth Low Energy may track vital signs and detect falls, sending that information to a remote mobile relay. Data collected from sensors can only be uploaded to a cloud server if a mobile device, known as a donor in this system, is nearby. The detected data is not processed by the mobile relays; instead, they simply relay it to a cloud server based on their RSSI ratings. With this much room, the data collected by each Internet of Things sensor may be safely kept. If the server determines that anything is wrong with the sensed data, it may notify the people responsible for the patients by text message or electronic mail.

2. *Internet of m-Health Things (m-IoT)*: The product's software application uses ways for a variety of wearable sensors to collect data on a user's physiological indications. Signals in the form of brief messages may be used to alert medical care professionals to the causes of a health crisis and guide them in taking the necessary actions, depending on the circumstances. More opportunities to increase human services quality, for greater improvements in persistent fulfillment, and for easier mediation are made possible by instant access to medical data [29].
3. *Adverse Drug Reaction*: This is a health problem brought on by medication use. The medical data is then arranged with the use of a pharmaceutically perceptive database [10]. A major obstacle to providing adequate medicine-based healthcare is the presence of counterfeit medications in the pharmaceutical supply chain. For a safe and manageable medication supply chain, see [30] for a discussion of blockchain-based IoT solutions for medication management. There are four distinct phases in which the medications' purity is compromised. Owing to possible counterfeiting throughout both the ingredient acquisition and medication manufacturing processes, the manufactured pharmaceuticals may not be as effective as advertised. The distribution phase of medicine delivery is notoriously murky, particularly at the frontiers of developing nations. Moreover, inadequate medication usage occurs at the pharmacy level due to the prevalence of the illicit market and online drug sales. All players in today's pharmaceutical sector benefit from a safe and transparent supply chain, thanks to blockchain and the traceability of IoT solutions.
4. *Community Healthcare*: Benefits for healthcare systems may be expected from the proliferation of IoT devices, such as the ability to remotely monitor patients and provide quick alert mechanisms. Those living in rural areas, where it might be more challenging to receive quality and efficient treatment, can also reap the advantages of smart healthcare systems, thanks to their increased accessibility. An electronic health monitoring system is suggested to lessen the need for patient visits and other inconvenient hospital procedures in [31]. For the benefit

of underserved rural areas, the suggested model captures and sends vital signs to a medical professional. By using medical data and analysis from a broadband viewpoint, community healthcare is achieved. There is a great need for a comprehensive study of persons to take targeted and effective action during crises like floods, earthquakes, cyclones, and the spread of infectious diseases. For instance, a smart city emergency response strategy is shown in [32]. In the suggested concept, a sensor and fog node form a monitoring unit that can track patients' vitals and interpret the collected data. The path designer unit determines the most efficient route for sending the collected data from its origin to its destination. Cloud-based path designer units connect patients with nearby medical facilities. The route designer module also has an alert system that may provide critical, warning, and safe notifications to the user interface.

4 Research Limitations and Future Directions

The 5G infrastructure is anticipated to increase the adoption rate of AI in the healthcare service industry. Medical imaging systems, such as MRIs, produce very big files; 5G networks may enhance the quality and availability of treatment by facilitating the rapid transmission of such data accurately. Patients may quickly obtain medical care or consult with experts, thanks to the 5G network's ability to be utilized for telemedicine. The best method to limit losses when the future is uncertain is to act promptly. Artificial intelligence (AI)- based technology, which can exceed the limits of expert knowledge, is shattering barriers in healthcare, well-being, and even life itself beyond the constraints of time and space. It is anticipated that the purview of AI-based medical application would grow from predisease through treatment and post-care to everyday living, even though they now concentrate on patient-centered management of illnesses. Consequently, it is important to use AI-enabled technology with a holistic view that considers the whole scope of people's lives, not only in the context of healthcare.

Current applications of AI-based technologies form the basis of the recommendations offered in this paper, which may prevent us from realizing the full potential of future technologies. This work has contributed to the successful use and management of AI by providing guidelines based on a survey of the literature and actual uses of AI systems in healthcare organizations. Our research will pave the way for future investigation into the effective use of AI in healthcare with aim of providing quality care to patients and reducing preventable illness and death. There are gaps in this study. First, there was a dearth of AI operations that were unavailable due to the fact that research articles seldom provide specifics on how AI worked due to the mostly secret nature of these features. Second, not all relevant research on AI in healthcare was found, especially gray literature and reports that were not published in the databases that were examined.

We anticipate that AI will play a significant part in future medical services. Its application to machine learning has resulted in precision medicine, which is generally acknowledged as a much needed improvement in patient care. Notwithstanding the difficulties encountered in the past, we anticipate that AI will eventually be able to provide accurate diagnoses and effective treatment suggestions. A growing number of healthcare facilities are using speech and text recognition technology for routine activities including patient communication and note-taking.

Not whether the technologies will be competent enough to be helpful is the biggest obstacle facing AI in many healthcare fields, but rather securing their acceptance in everyday clinical practice. Artificial intelligence (AI) systems need regulatory approval, integration with electronic health record (EHR) systems, sufficient standardization so that similar products function similarly, education of clinicians, funding from public or private payer organizations, and ongoing field updates to be widely adopted. These problems can be solved, but it will take far longer than the development time of the underlying technology. So, we anticipate some incorporation of AI into clinical practice over the next 5 years, with broader use following within the next 10. All these evidence suggest that AI will not replace human therapists but will instead complement their work. In the future, doctors and other medical professionals may focus more on activities and job designs that make use of human strengths like empathy, persuasion, and the ability to see the big picture. In the long run, it is possible that only healthcare workers who refuse to collaborate with AI will be let go.

References

1. Sundaravadivel, P., Kougianos, E., Mohanty, S. P., & Ganapathiraju, M. K. (2017). Everything you wanted to know about smart health care: Evaluating the different technologies and components of the internet of things for better health. *IEEE Consumer Electronics Magazine, 7*(1), 18–28.
2. Rathert, C., Williams, E. S., McCaughey, D., & Ishqaidef, G. (2015). Patient perceptions of patient-centred care: Empirical test of a theoretical model. *Health Expectations, 18*(2), 199–209.
3. Kadhim, K. T., Alsahlany, A. M., Wadi, S. M., & Kadhum, H. T. (2020). An overview of patient's health status monitoring system based on internet of things (IoT). *Wireless Personal Communications, 114*(3), 2235–2262.
4. Srivastava, J., & Routray, S. (2023, January). *AI enabled internet of medical things framework for smart healthcare*. In Innovations in Intelligent Computing and Communication: First International Conference, ICIICC 2022, Bhubaneswar, Odisha, India, December 16–17, 2022, Proceedings (pp. 30–46). Springer.
5. De Mattos, W. D., & Gondim, P. R. (2016). M-health solutions using 5G networks and M2M communications. *IT Professional, 18*(3), 24–29.
6. Mehmood, Y., Ahmad, F., Yaqoob, I., Adnane, A., Imran, M., & Guizani, S. (2017). Internet-of-things-based smart cities: Recent advances and challenges. *IEEE Communications Magazine, 55*(9), 16–24.
7. Chen, X., Xie, H., Li, Z., Cheng, G., Leng, M., & Wang, F. L. (2023). Information fusion and artificial intelligence for smart healthcare: A bibliometric study. *Information Processing & Management, 60*(1), 103113.

8. Tariq, F., Riazul Islam, S. M., & Mostefaoui, G. K. (2023). Future challenges in artificial intelligence for smart healthcare. In *Artificial intelligence for disease diagnosis and prognosis in smart healthcare* (pp. 05–310). CRC Press/Taylor Francis.
9. Qi, J., Yang, P., Min, G., Amft, O., Dong, F., & Xu, L. (2017). Advanced internet of things for personalised healthcare systems: A survey. *Pervasive and Mobile Computing, 41*, 132–149.
10. Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The internet of things for health care: A comprehensive survey. *IEEE Access, 3*, 678–708.
11. Baker, S. B., Xiang, W., & Atkinson, I. (2017). Internet of things for smart healthcare: Technologies, challenges, and opportunities. *IEEE Access, 5*, 26521–26544.
12. Mahmoud, M. M., Rodrigues, J. J., Ahmed, S. H., Shah, S. C., Al-Muhtadi, J. F., Korotaev, V. V., & De Albuquerque, V. H. C. (2018). Enabling technologies on cloud of things for smart healthcare. *IEEE Access, 6*, 31950–31967.
13. Dhanvijay, M. M., & Patil, S. C. (2019). Internet of Things: A survey of enabling technologies in healthcare and its applications. *Computer Networks, 153*, 113–131.
14. Ahad, A., Tahir, M., & Yau, K. L. A. (2019). 5G-based smart healthcare network: Architecture, taxonomy, challenges and future research directions. *IEEE Access, 7*, 100747–100762.
15. Singh, R. P., Javaid, M., Haleem, A., Vaishya, R., & Ali, S. (2020). Internet of Medical Things (IoMT) for orthopaedic in COVID-19 pandemic: Roles, challenges, and applications. *Journal of Clinical Orthopaedics and Trauma, 11*(4), 713–717.
16. Al-Turjman, F., Nawaz, M. H., & Ulusar, U. D. (2020). Intelligence in the Internet of medical things era: A systematic review of current and future trends. *Computer Communications, 150*, 644–660.
17. Jagatheesaperumal, S. K., Mishra, P., Moustafa, N., & Chauhan, R. (2022). A holistic survey on the use of emerging technologies to provision secure healthcare solutions. *Computers and Electrical Engineering, 99*, 107691.
18. Alabdulatif, A., Khalil, I., & Saidur Rahman, M. (2022). Security of Blockchain and AI-empowered smart healthcare: Application-based analysis. *Applied Sciences, 12*(21), 11039.
19. Wazid, M., Das, A. K., & Shetty, S. (2022). BSFR-SH: Blockchain-enabled security framework against ransomware attacks for smart healthcare. *IEEE Transactions on Consumer Electronics*
20. Zia, M., Obaidat, M. S., Mahmood, K., Shamshad, S., Saleem, M. A., & Chaudhry, S. A. (2022). A provably secure lightweight key agreement protocol for wireless body area networks in healthcare system. *IEEE Transactions on Industrial Informatics, 19*(2), 1683–1690.
21. Kong, K., Zhang, Z., Guo, C., Han, J., & Long, G. (2022). PMMSA: Security analysis system for android wearable applications based on permission matching and malware similarity analysis. *Future Generation Computer Systems, 137*, 349–362.
22. Syed, R. A. (2019). *Next generation and advanced network reliability analysis* (pp. 1–27). Springer.
23. Bader, A., Ghazzai, H., Kadri, A., & Alouini, M. S. (2016). Front-end intelligence for large-scale application-oriented internet-of-things. *IEEE Access, 4*, 3257–3272.
24. Rachim, V. P., & Chung, W. Y. (2016). Wearable noncontact armband for mobile ECG monitoring system. *IEEE Transactions on Biomedical Circuits and Systems, 10*(6), 1112–1118.
25. Yew, H. T., Ng, M. F., Ping, S. Z., Chung, S. K., Chekima, A., & Dargham, J. A. (2020, February) *IoT based real-time remote patient monitoring system*. In 2020 16th IEEE international colloquium on signal processing & its applications (CSPA) (pp. 176–179). IEEE.
26. Zhong, D., Yian, Z., Lanqing, W., Junhua, D., & Jiaxuan, H. (2020). Continuous blood pressure measurement platform: A wearable system based on multidimensional perception data. *IEEE Access, 8*, 10147–10158.
27. Bisio, I., Garibotto, C., Lavagetto, F., & Sciarrone, A. (2019). When eHealth meets IoT: A smart wireless system for post-stroke home rehabilitation. *IEEE Wireless Communications, 26*(6), 24–29.
28. Kumar, P. M., & Gandhi, U. D. (2018). A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases. *Computers & Electrical Engineering, 65*, 222–235.

29. Ahmadi, H., Arji, G., Shahmoradi, L., Safdari, R., Nilashi, M., & Alizadeh, M. (2019). The application of internet of things in healthcare: A systematic literature review and classification. *Universal Access in the Information Society, 18*, 837–869.
30. Ahmadi, V., Benjelloun, S., El Kik, M., Sharma, T., Chi, H., & Zhou, W. (2020, February). *Drug governance: IoT-based blockchain implementation in the pharmaceutical supply chain*. In 2020 Sixth International conference on mobile and secure services (MobiSecServ) (pp. 1–8). IEEE.
31. Jabirullah, M., Ranjan, R., Baig, M. N. A., & Vishwakarma, A. K. (2020, February). *Development of e-health monitoring system for remote rural community of India*. In 2020 7th International conference on signal processing and integrated networks (SPIN) (pp. 767–771). IEEE.
32. Sundaravadivel, P., Lee, I., Mohanty, S., Kougianos, E., & Rachakonda, L. (2019, December). *RM-IoT: An IoT based rapid medical response plan for smart cities*. In 2019 IEEE International symposium on smart electronic systems (iSES) (Formerly iNiS) (pp. 241–246). IEEE.

A Multidisciplinary Explanation of Healthcare AI Uses, Trends, and Possibilities

Siddharth Vats and Anurag Dixit

Abstract AI could enhance patient outcomes, reduce healthcare expenditures, and increase efficiency. AI can analyze medical information, imaging scans, and test results to help clinicians make better decisions. On the basis of an individual's medical history, AI is also capable of developing new drugs and customizing treatments. Radiologists can employ AI algorithms to more accurately diagnose medical imaging problems. AI can also contribute in the development of cancer detection and diagnosis tools. In addition to diagnosis, AI can streamline administrative tasks such as appointment scheduling and the maintenance of electronic medical records, allowing healthcare professionals more time to focus on patient care. Concerns about data privacy and the need for universal rules and norms could impede the deployment of AI in healthcare. To use AI ethically and effectively, healthcare professionals require training.

1 Introduction

AI in healthcare boosts efficiency, budgets, and patient outcomes [1]. Any AI revolution must start with healthcare data. Artificial intelligence now has a vast range of data to work with, including electrocardiograms, MRI scans, and X-rays from hospitals; DNA sequences from labs; and heart rate, blood pressure, and sleep cycle patterns from smart devices [2]. Data issues do exist.

Strict rules govern data security, privacy, and confidentiality. Despite this, a lot of startup businesses have been successful in navigating regulatory processes and obtaining approvals and licenses to market their products to hospitals, pharmacies,

---

S. Vats
Department of Biotechnology, Noida International University, Greater Noida, India

A. Dixit (✉)
Department of Computer Science and Engineering, School of Engineering and Technology, Noida International University, Greater Noida, India

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_6

biotechnology companies, and private consumers despite the fact that these companies are notoriously difficult to regulate.

Since 2014, more than 50 companies have acquired FDA approval for machine learning and AI algorithms. Due to COVID-19, digital health services have increased. Rapid testing, respiratory monitoring, and telehealth diagnostics are growing in popularity, which is encouraging businesses to accelerate product development. Governments around the world have developed new coronavirus R&D programs to aid business owners and corporations in the fight against COVID-19 [3, 4]. Costs for healthcare have gone up, especially in developed countries. Large financial investments are required for drug development and clinical trials, and even a little increase in healthcare costs could contribute hundreds of millions of dollars. Companies in the biotech, pharmaceutical, and clinical research sectors might save a lot of money by increasing productivity. This market's current size is sufficient to sustain a competitive environment, which is supported by the low entry barriers that attract a significant number of new competitors [5]. As people's awareness of technology expands, they anticipate obtaining more personalized and interconnected medical care. Providers of artificially intelligent medical care are under intense pressure to innovate in order to meet the evolving needs of their patients. This involves a greater emphasis on patients as clients, the adoption of preventative care models, the formation of strategic alliances with disruptive new firms (which may include healthcare AI technologies), and the reduction of healthcare expenses [8]. However, this list of potential remedies is not exhaustive. Due to improvements in consumer technology, namely, the Internet of Things, individuals may now assume responsibility for their own health and make educated decisions.

A growing number of AI healthcare businesses are building consumer-facing apps that employ AI to assist users in tracking and improving their health, selecting the appropriate physician, and coordinating care [13–15]. These applications were created with the user in mind. As a result of the epidemic, certain nations are prioritizing the use of AI in the healthcare industry. AI has a positive effect on every element of healthcare, including diagnosis, data creation, interpretation, patient data management, therapy, and future treatment plans (Fig. 1).

2 AI Application in Healthcare Sector

AI in biomedical engineering, biotechnology, biochemical engineering, and healthcare appears promising [15]. AI can improve illness detection and treatment by creating new medical devices. AI systems can analyze massive medical data to identify trends and predict outcomes, enabling more accurate and timely diagnosis. AI can also design patient-specific medical products like prostheses. AI can also be utilized in computer-aided design and simulation to make medical equipment more effective and efficient. AI in biomedical engineering could improve patient care worldwide.

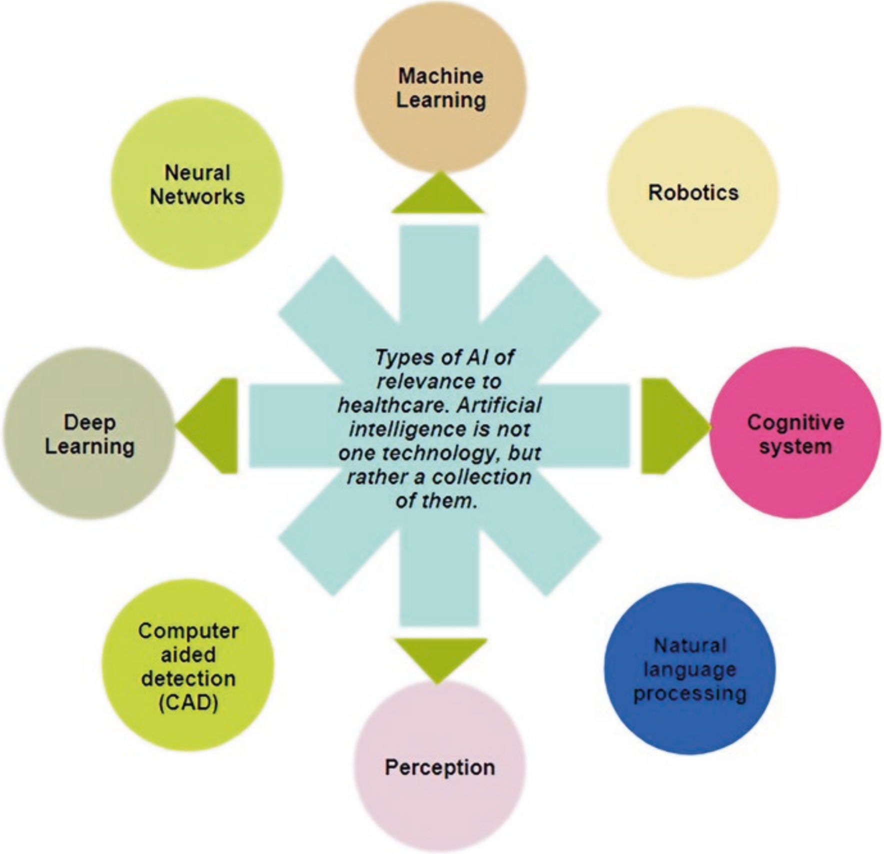


Fig. 1 Types of AI as per their relevance in the field of health sector

2.1 Medical Imaging

By analyzing X-rays, CT scans, and MRIs, AI systems can help radiologists diagnose diseases more quickly and accurately [16, 17]. AI could help medical imaging, which uses X-rays, CT scans, MRIs, and PET scans, to detect and monitor medical conditions. AI technologies can automate tasks and provide insights and recommendations to enhance the precision and efficacy of medical imaging. In medical imaging, AI is used to detect diseases and issues that professionals may overlook. To spot trends and predict patient health, AI systems can analyze vast quantities of medical image data, including patient demographics, lifestyle factors, and medical history. This data can aid clinicians in making more accurate diagnoses. AI is also employed in the development of computer-aided detection (CAD) systems that can automatically identify structures or anomalies in medical images that may be indicative of diseases or disorders. These techniques can help clinicians identify these structures more rapidly and precisely [18]. AI also contributes to the development of new medical imaging technologies, such as virtual reality and augmented reality systems, which can enhance the visualization of images. These technologies improve patient outcomes by providing more precise and detailed images, hence reducing misdiagnoses and increasing the diagnostic accuracy. AI in medical imaging is still relatively novel, but it potentially improves precision and efficacy. As AI technology progresses, it is anticipated that medical imaging would benefit [20].

2.2 Diagnosis

AI can assist physicians in making diagnoses by analyzing massive volumes of medical data such as electronic health records and test results to identify patterns and connections that humans may overlook. AI is increasingly employed for medical diagnostics [21, 22]. Medical images, patient information, and test results can be analyzed with the assistance of AI. Thus, AI can facilitate early diagnosis, accelerate the process, and improve diagnostic precision. In medical diagnostics, CAD systems are a prominent use of artificial intelligence. These technologies provide a second opinion on X-rays, MRIs, and CT scans to radiologists. Artificial intelligence algorithms trained on a big dataset of images can detect diseases that human doctors may overlook. Additionally, AI is analyzing EHRs to identify disease-prone individuals. In order to anticipate disease risk, AI systems can identify trends in demographic, lifestyle, and medical data. Anxiety and despair are also diagnosed by AI systems. Voice and facial expressions can alert AI systems about mental health risks. AI helps and enhances the decision-making of medical professionals; it does not replace them. AI should assist physicians, not replace them [22].

2.3 Drug/Pharmacological Discovery

Artificial intelligence can analyze vast datasets to identify pharmacological targets. The use of AI in drug discovery is expanding. AI systems can analyze vast quantities of data to identify potential therapeutic targets and accelerate the drug discovery process [23, 24]. AI analyzes genomic data for drug discovery. Large genomic datasets can be mined by AI to identify medicinal targets. The identification of promising targets can accelerate drug discovery and increase success rates. Molecular data are analyzed using AI in drug research. Artificial intelligence programmers can evaluate complex chemical compounds and predict protein interactions. This information can aid in the discovery of new therapeutic targets and the design of more effective medicines. To accelerate drug discovery, AI algorithms reduce the number of compounds evaluated in the laboratory. AI systems can analyze the structures of compounds and predict their efficacy, allowing researchers to concentrate on the most promising chemicals. In drug research, AI is still evolving; therefore, its full potential has not yet been realized [23, 24]. However, AI in drug research is growing rapidly and is anticipated to play a greater role in drug discovery in the future.

2.4 Personalized Medicine

AI is capable of creating personalized medications for patients based on their medical histories and characteristics. AI could help personalized medicine, which tailors medication to a patient's genetics, lifestyle, and medical history [25, 26]. By analyzing vast quantities of medical data, AI can help personalize medication. AI algorithms are able to recognize patterns and develop predictions from this data in order to personalize treatment regimens. AI can detect genetic alterations that may influence a patient's response to treatment. This information helps determine the optimal treatment for each patient. In personalized medicine, AI also analyzes medical imaging [25, 26]. AI systems can analyze medical images such as MRIs and CT scans to detect diseases that clinicians may overlook. This information can be used to tailor therapies and improve diagnoses. AI also facilitates medical research and development. Artificial intelligence algorithms can review vast amounts of data to identify the most promising chemicals for development, removing the need for laboratory research. This can expedite and enhance the drug development process. AI in personalized medicine is still in its infancy, but it has the potential to improve diagnostic accuracy, accelerate the discovery of novel medications, and personalize treatment regimens. As AI technology progresses, personalized medicine will gain advantages.

2.5 Clinical Decision Support

AI can provide clinicians with real-time recommendations and data. AI could boost clinical decision support, which uses information technology to help physicians make better judgments [19, 27]. Algorithms powered by artificial intelligence are capable of analyzing vast quantities of medical data and providing recommendations to improve clinical decision- making. Artificial intelligence analyzes digitized medical records for clinical decision support (EHRs). Massive amounts of EHR data, including patient demographics, lifestyle characteristics, and medical history, can be analyzed by AI algorithms to find patterns and generate health forecasts. This data can aid clinicians in making more accurate diagnoses. In clinical decision support, medical image analysis employs AI. AI algorithms can analyze X-rays, MRIs, and CT scans to discover diseases and abnormalities that clinicians may overlook

[26]. This data can aid clinicians in making more accurate diagnoses. AI facilitates healthcare decision-making in real time. Using vital signs and other data, AI systems can detect health problems. With this knowledge, healthcare practitioners may make more informed decisions and improve patient outcomes. AI in clinical decision support is still novel, but it has the potential to improve the precision and efficacy of clinical decision-making. As AI technology advances, clinical decision support will reap the rewards.

2.6 Administrative Tasks

By automating appointment scheduling, electronic medical record keeping, and invoicing with artificial intelligence, healthcare practitioners may focus on patient care. Healthcare administrative tasks that are performed outside of direct patient care but are essential to the efficient and effective operation of healthcare organizations may benefit from the use of artificial intelligence [21, 24]. These processes can be automated by AI algorithms, thereby lowering the workload of healthcare personnel and enhancing organization efficiency. In healthcare administrative duties, AI is used for claims processing. AI systems can detect errors and inconsistencies in insurance claims, saving time and effort. This can improve the accuracy and efficiency of claims processing, reducing denied claims and the administrative burden on healthcare providers. Appointment scheduling in healthcare also utilizes AI. AI algorithms examine patient information and appointment schedules to determine the most efficient appointment times, thereby saving time and effort. This can improve appointment scheduling and wait times for patients. Manual medical coding and billing are time-consuming and prone to error; AI is automating these processes. Large volumes of data can be analyzed by AI algorithms to determine the most appropriate codes and bills for each patient, reducing administrative burden on healthcare practitioners and increasing billing precision. AI in administrative tasks in healthcare is still novel, but it has the potential to increase efficiency and accuracy. As AI develops, it will automate administrative tasks in the healthcare industry. AI could enhance the quality, efficiency, and patient care of healthcare. AI should support healthcare practitioners, not replace them. The use of AI in healthcare creates moral and regulatory concerns.

3 Healthcare Difficulties and AI

Costs are increasing, healthcare professionals are few, and an aging population requires increased medical care. AI could enhance healthcare, efficiency, and costs. However, there are disadvantages to implementing AI in healthcare [10–12].

3.1 Data Privacy

AI in healthcare collects and analyzes huge quantities of sensitive patient data. This data is tough to secure. Patient data confidentiality is a big obstacle for healthcare AI [9]. Vital is the protection of health data against unauthorized access and use. AI systems can protect patient data in a variety of ways:

Encryption

Encrypting data prior to storage or transmission prevents unauthorized access.

Access Control

Authentication and authorization can restrict access to patient data via access control.

Data Minimization

Collecting and storing only the minimum amount of data required by the AI system reduces data breaches and unauthorized access.

Anonymization

Data removing or hiding patient data could avoid breaches of privacy while facilitating data analysis.

Privacy-Enhancing Technologies

Differential privacy and holomorphic encryption are privacy-enhancing technologies that help protect patient data. In addition to these technical safeguards, AI systems should have clearly established policies and procedures for collecting, keeping, and utilizing patient information. This may require obtaining patient consent, regularly reviewing the system for privacy breaches, and developing a reaction strategy for data breaches. When applying AI in healthcare, technical safeguards and explicit policies and procedures can help preserve patient data.

3.2 Regulation

Standards and regulations are required for the ethical and successful application of AI in healthcare.

3.3 Bias in AI Algorithms

If trained on biased data, AI systems can exacerbate prejudices in healthcare.

3.4 Interoperability

The healthcare sector employs complex electronic health records and other information technologies. Integration of AI into these systems is challenging.

3.5 Workforce Training

Healthcare personnel must use AI technology ethically and successfully, and they must be trained to do so. Despite these obstacles, AI has tremendous potential in healthcare. AI can improve patient outcomes, efficiency, and costs and assist healthcare providers in providing the most effective care. By overcoming AI adoption obstacles, the healthcare industry may be able to enhance patient outcomes and efficiency. AI research and applications in healthcare are vital. AI-based firms boost consumer fitness, clinical trials, and hospital diagnoses.

4 Government Initiatives for AI in Healthcare

AI can enhance illness diagnosis, patient outcomes, and delivery of care. To maximize benefits and limit risks, governments worldwide are supporting the development and application of AI in healthcare.

4.1 Research and Development Funding

Government funding is provided to universities, research institutions, and corporations for AI healthcare research. These investments contribute to the development of breakthrough AI algorithms; enormous, diversified datasets for training AI systems; and clinical trials to verify AI systems.

4.2 National AI Strategies

Numerous countries have formed AI development and deployment policies, particularly in the healthcare sector. These policies influence government AI funding, ethical and legal requirements, and the development of AI in healthcare.

4.3 Regulation and Standardization

Governments are creating legislation and standards to ensure safe, effective, and ethical healthcare AI systems. These regulations may include testing and validation of AI systems, data privacy and security, and healthcare decision-making criteria.

4.4 Data Sharing and Interoperability

Governments are pushing the sharing and interoperability of healthcare data in order to assist AI systems in analyzing vast quantities of patient data. This includes attempts to standardize data formats and protocols, build data networks, and stimulate data sharing among healthcare professionals, researchers, and patients.

4.5 Workforce Development

Governments encourage the expansion of AI in healthcare by training healthcare professionals and creating new AI-related positions. This includes AI training for clinicians and healthcare administrators, as well as support for new AI-related professions such as data scientists and machine learning engineers.

4.6 Public-Private Partnerships

Governments are using public-private collaborations and co-funded research programs to advance AI in healthcare. Government entities, healthcare providers, researchers, and technology firms work together to develop and implement AI technologies in healthcare.

4.7 International Collaboration

International research and data sharing agreements are helping governments worldwide boost AI in healthcare. Examples include harmonizing norms and standards, sharing best practices, and working on massive data initiatives.

5 New Startups in the Field of Healthcare Based on AI

New healthcare technologies and discoveries emerge daily. AI could change healthcare. Many firms use AI to improve healthcare quality and accessibility. AI helps one startup diagnose. The business uses machine learning algorithms to quickly and correctly detect diseases and ailments by analyzing massive volumes of medical data. Medical professionals save time and patients get the right treatment faster. Another startup manages diabetes and heart disease with AI. The business monitors patients in real time and offers individualized health advice using wearable technology and AI algorithms. By limiting hospital visits, this improves patient outcomes and lowers healthcare costs. A startup uses AI to develop new pharmaceuticals and therapies. The startup finds new drug targets and improves disease treatments by analyzing massive genetic and medical data. This could save millions of lives and improve patient health worldwide.

6 Ethical Issues in the Application of AI to Healthcare

AI can enhance patient outcomes and increase the efficiency and effectiveness of healthcare. It also raises moral issues [6, 7]. Concerns include the following:

6.1 Bias and Discrimination

Artificial intelligence systems based on biased data can exacerbate healthcare biases. This may result in prejudice and disparate care. An AI algorithm trained on the data of predominantly white, middle-class persons may be incapable of identifying diseases or conditions that are more prevalent in other communities.

6.2 Data Privacy

Artificial intelligence in healthcare collects and analyzes massive amounts of sensitive patient data, such as personal and medical information. Data security and privacy is a major ethical concern. Data breaches can result in identity theft or the disclosure of sensitive information, making this an important consideration.

6.3 Responsibility for Errors

Errors resulting from AI systems can be difficult to identify. This raises questions regarding medical decision-making and responsibility. Who is accountable if an AI system misdiagnoses a patient: the algorithm, the healthcare professional who depended on it, or the developer of the program?

6.4 Impact on Employment

AI could automate numerous healthcare tasks, resulting in job losses and other workforce consequences. This raises ethical considerations regarding how AI may affect employment and how organizations should help workers as AI becomes more widespread in healthcare. How will healthcare staff be supported if AI algorithms automate their tasks?

6.5 Autonomy and Informed Consent

Patients have the right to make educated medical decisions. However, if AI algorithms choose treatment alternatives, people may not grasp the rationale, raising concerns about their autonomy and informed consent. For instance, if an AI algorithm proposes a treatment, the patient should know why and be able to make an informed choice.

6.6 Transparency and Interpretability

Complicated AI algorithms make it difficult to comprehend the decision-making process. Transparency, AI algorithm interpretability, accountability, and oversight are raised. If an AI algorithm is making patient care decisions, it’s crucial to understand how it’s making them and make sure they’re ethical. These ethical considerations emphasize the need for healthcare AI legislation and guidelines. In addition to considering the ethical implications of their AI initiatives, healthcare organizations should include key stakeholders, such as patients and healthcare professionals, in AI research and deployment. In conclusion, AI can improve healthcare, but ethical considerations must be addressed. The healthcare industry may use AI to improve patient outcomes and build a more efficient and effective system by considering these ethical issues and implementing suitable legislation and standards.

7 Conclusion

AI will transform healthcare. AI technology can enhance patient outcomes, efficiency, productivity, and healthcare costs as they evolve. AI aids medical imaging analysis, disease diagnosis, and individualized therapy planning. In population health management, drug discovery and development, and telemedicine, AI will become increasingly crucial. AI could enhance healthcare outcomes, lower costs, and boost efficiency. However, AI must be employed ethically and help all patients, regardless of socioeconomic class or geography.

References

1. Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism, 69*, S36–S40.
2. Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 14*(4), 337–339.
3. Kumar, A., Gupta, P. K., & Srivastava, A. (2020). A review of modern technologies for tackling COVID-19 pandemic. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 14*(4), 569–573.
4. Erion, G., Janizek, J. D., Hudelson, C., Utarnachitt, R. B., McCoy, A. M., Sayre, M. R., White, N. J., & Lee, S.-I. (2021). CoAI: Cost-aware artificial intelligence for health care. *medRxiv*, 2021-01.
5. Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine, 260*, 113172.
6. Karimian, G., Petelos, E., & Evers, S. M. (2022). The ethical issues of the application of artificial intelligence in healthcare: A systematic scoping review. *AI and Ethics, 2*(4), 539–551.
7. Kamel Boulos, M. N., & Zhang, P. (2021). Digital twins: From personalised medicine to precision public health. *Journal of Personalized Medicine, 11*(8), 745.
8. Panch, T., Mattie, H., & Celi, L. A. (2019). The "inconvenient truth" about AI in healthcare. *NPJ Digital Medicine, 2*(1), 77.
9. Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare* (pp. 295–336). Academic.
10. Wang, F., & Preininger, A. (2019). AI in health: State of the art, challenges, and future directions. *Yearbook of Medical Informatics, 28*(1), 16–26.
11. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine, 28*(1), 31–38.

12. Kaviani, S., Han, K. J., & Sohn, I. (2022). Adversarial attacks and defenses on AI in medical imaging informatics: A survey. *Expert Systems with Applications, 116815*.
13. England, J. R., & Cheng, P. M. (2019). Artificial intelligence for medical image analysis: A guide for authors and reviewers. *American Journal of Roentgenology, 212*(3), 513–519.
14. Subasi, A. (2022). *Applications of artificial intelligence in medical imaging*. Elsevier Science Publishing Company Incorporated.
15. Briganti, G., & Le Moine, O. (2020). Artificial intelligence in medicine: Today and tomorrow. *Frontiers in Medicine, 7*, 27.
16. Ström, P., Kartasalo, K., Olsson, H., Solorzano, L., Delahunt, B., Berney, D. M., et al. (2020). Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: A population-based, diagnostic study. *The Lancet Oncology, 21*(2), 222–232.
17. Neri, E., Coppola, F., Miele, V., Bibbolino, C., & Grassi, R. (2020). Artificial intelligence: Who is responsible for the diagnosis? *La Radiologia Medica, 125*, 517–521.
18. Chan, H. S., Shan, H., Dahoun, T., Vogel, H., & Yuan, S. (2019). Advancing drug discovery via artificial intelligence. *Trends in Pharmacological Sciences, 40*(8), 592–604.
19. Montani, S., & Striani, M. (2019). Artificial intelligence in clinical decision support: A focused literature survey. *Yearbook of Medical Informatics, 28*(01), 120–127.
20. Zihni, E., Madai, V. I., Livne, M., Galinovic, I., Khalil, A. A., Fiebach, J. B., & Frey, D. (2020). Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome. *PLoS One, 15*(4), e0231166.
21. Silcox, C., Dentzer, S., & Bates, D. W. (2020). AI-enabled clinical decision support software: A “trust and value checklist” for clinicians. *NEJM Catalyst Innovations in Care Delivery, 1*(6).
22. Kindle, R. D., Badawi, O., Celi, L. A., & Sturland, S. (2019). Intensive care unit telemedicine in the era of big data, artificial intelligence, and computer clinical decision support systems. *Critical Care Clinics, 35*(3), 483–495.
23. Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in healthcare: A structured literature review. *BMC Medical Informatics and Decision Making, 21*, 1–23.
24. Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine, 112*(1), 22–28.
25. Susar, D., & Aquaro, V. (2019, April). Artificial intelligence: Opportunities and challenges for the public sector. In *Proceedings of the 12th international conference on theory and practice of electronic governance* (pp. 418–426).
26. Chatterjee, S. (2020). AI strategy of India: Policy framework, adoption challenges and actions for government. *Transforming Government: People, Process and Policy, 14*(5), 757–775.
27. Kankanhalli, A., Charalabidis, Y., & Mellouli, S. (2019). IoT and AI for smart government: A research agenda. *Government Information Quarterly, 36*(2), 304–309.

Optimum Utilization of Bed Resources in Hospitals: A Stochastic Approach

Tarun Kumar Vashishth, Sachin Chaudhary, and Vikas Sharma

Abstract Hospital management may encounter situations where patients are turned away because all beds are occupied, and medical management is shut down due to insufficient beds. Inadequate funding or poor performance of resources often contributes to this situation. This article is devoted to using queuing theory to reduce waiting time in hospital allocation. Over time, hospital management was faced with the problem of allocating beds for various in-house medical services, as patient admissions and demand for medical services increased. This article presents the complex analysis of resource allocation in hospital management through systems integration, mathematical models of cutoff values using select circle line, and variation by quality on the same basis. This research demonstrates how to allocate hospital beds with respect to services required by hospital managers. Hospital beds are the main resources managed in hospitals. However, the management strategies adopted in practice or discussed in the literature do not clearly show inconsistencies in methods and associated costs.

1 Introduction

When we need help, waiting can seem hopeless. Nationally, the average wait time in an emergency room is currently 222 min or 3 h and 42 min. Andhra Pradesh (274.2 min), Maharashtra (282 min), Uttar Pradesh (245 min), Tamil Nadu (214.8 min), and Nataka (214.2 min) are the least desirable of among cities, having a standby time of up to 4 h or more. But at *Spectrum Health Hospitals*, they are constantly working at 85% capacity without changing staff. Divide the patients into three groups as emergency, convalescent, and nonurgent, run quickly on a

---

T. K. Vashishth (✉) · S. Chaudhary · V. Sharma
School of Computer Science and Applications, IIMT University, Meerut, Uttar Pradesh, India
e-mail: tarun_socsa@iimtindia.net; sachin_socsa@iimtindia.net; vikas_socsa@iimtindia.net

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_7

well-designed computer, and reduce the door-to-door waiting time of more than 4 h for treatment to 1.5 h.

Existence of beds in hospital is a persistent issue in emerging nations. Healthcare issues in developing nations are complex and stem from a variety of factors, including cultural, political, and economic factors, poor planning, and/or the misuse of laws and programmers. Bed availability is one of the most important factors in deciding to use a hospital. In India, with an average bed occupancy rate of 6.8/10.0002, insufficiency of hospital beds is a major problem.

2 Related Literature Review

Our study contributes to the practice of waiting queues in order to reduce the waiting time in the hospital. The term "hospital use" refers to the way community makes use of hospital facilities. Fully exploring a hospital involves a comprehensive examination of all its facets, encompassing services catering to outpatient and home patients alike. However, many countries, the concept of "hospital identity" remains unfamiliar. Christina et al. [1] conducted an investigation into the adequacy of access and the duration of stay. A Derby hospital where patients trust the NHS Trust. Green et al. [2] presented the implications of activity-based cost accounting for healthcare investment decisions. Harry [3] developed a career in hospitals and public policy. Jeffrey et al. [4] build efficiency for US public hospitals. Milton et al. [5] developed a mathematical method for selecting the optimal mix of health plans. Richard et al. [6] establish a health planning and management program. Robinson [7] examined the drivers and constraints of growth of healthcare companies. Roland et al. [8] develop the concept of cost-effectiveness in clinical evaluation. Ross [9] developed a health screening system using the ABC. Samuel et al. [10] design health management.

3 Data of the Problem

To comprehensively assess the utilization of hospital beds, it is essential to discuss various terms that pertain to this context. These terms encompass patient arrivals denoted as (λ) the cost of service represented by the minimum time required for bed allocation noted as (μ), the inter-arrival time (IAT), bed occupancy rate (BOR), and the average duration of stay (ALS). These parameters are intrinsically linked to population dynamics and are governed by an exponent of –

1. Per-person hospitalization rate (HC), bed occupancy rate (BC)
2. Ped/index of population (B/P), etc.

Indicators are the only thing available in this study, though, when a bed is changed only, in the hospital (BTR).

The utilization of hospital beds within a community is influenced by a multitude of factors, encompassing treatment practices and individual attributes linked to social, economic, educational, and cultural dimensions. Given these diverse influences, it's important to approach the competition for hospital bed usage judiciously. Disparities in utilization across regions should not necessarily be interpreted as indicative of discrepancies in the quality or standard of care. However, due to the increase in treatment costs in hospitals, these studies are beneficial in directing the economy without affecting the appearance control well. Below are some factors that affect hospital use. Age groups, coverage and allocation of hospital beds, access to hospital services, hospital "lungs," treatment and communication quality, access to doctors, payment for hospital services, and patents are some of the factors that affect hospital beds. Hospital design, accommodation, and health conditions are further contributors to this intricate landscape.[1]

A descriptive study was conducted to examine bed use in a 492-bed hospital between January 2010 and March 2010. Enter data on inpatient admissions, lengths of stay, discharges, deaths, and hospital days for the full year 2009, from medical records and preplanned information. The gathering of information was limited for emergency (179 beds), surgical (41 beds), nonemergency (41 beds), combined with other hospitals (36 beds), and the total hospital bed utilization index for selected patients. The number of hospitalizations in 2009 was 12,550; 6405 of them are men (51.03%) and 6,145 are women (48.96%). The average hospital stay in the emergency department is 6–12 days. Maximum height on bed (50.13%) was below the best in June. It is also less in many months.

Table 1 Statistics for the Department of Medicine each month

Month	Admissions	Discharge	Hospital bed	Average	ALS H/ (D−d)	BOR (N/B)*100	TOI	BTR
	+ Deaths	Days	Occupancy					
Jan	207	231	2497	80.5	10.8	44.9	13.2	1.3
Feb	250	205	2510	84.9	11.6	47.4	12.8	1.1
Mar	148	269	2663	85.9	9.8	47.9	10.7	1.5
Apr	259	263	2420	80.6	9.2	45.0	11.2	1.5
May	252	279	2618	84.4	9.3	47.1	10.5	1.6
June	224	232	2342	78.0	10.0	43.6	13.0	1.3
July	277	334	2782	89.7	8.3	50.1	8.2	1.9
Aug	367	378	2302	74.2	6.0	41.1	8.5	1.1
Sep	338	347	2368	76.3	6.8	44.0	8.6	1.9
Oct	347	271	2453	79.1	9.0	45.6	11.4	1.5
Nov	271	286	2229	74.3	7.7	41.5	10.9	1.6
Dec	251	251	2384	76.9	9.4	42.9	12.6	1.4
Department of Medicine (covering pediatrics, oncology, geriatrics, and VD)								

H = Days spent in bed overall in a year

D (D1 + D2) = Deaths and discharges in the same year, respectively

BOR = 100 times the usual number of occupied beds

The arrival time is very long throughout the year and reaches a highest rate (13.2%) in January. It indicates the beds have been there for a future scope. The cost of the bed is very low every year (Table 1).

Throughout the year, a patient's typical operating room stay is 6–12 days, with the exception of July, which has less than 3 days [approx. 2.7]. The highest rate of beds occupied was in January (113.7%), with the highest quality, and lowest in July (4.1%). It was pretty much the sweet spot for last year. The January transition period is negative, showing that the bed occupancy is above 100%. November and December have the lowest bed turnover rate (1.9%) (Table 2).

Table 2 Statistics for the Department of Surgery each month

Month	Admissions	Discharge	Hospital bed	Average	ALS H/ (D−d)	BOR(N/B)*100	TOI	BTR
	+ Deaths	Days	Occupancy					
Jan	131	131	1446	46.6	11.0	113.7	−1.33	3.19
Feb	97	113	961	34.3	8.5	83.7	1.6	2.7
Mar	101	107	1079	34.8	10.0	84.8	1.7	2.6
Apr	103	108	970	32.3	8.9	78.8	2.4	2.6
May	101	113	1071	34.5	9.4	84.2	1.7	2.7
June	98	130	841	28.0	6.4	68.3	2.9	3.1
July	117	110	307	9.7	2.7	24.1	8.7	2.6
Aug	81	93	1040	33.5	11.1	81.8	2.4	2.2
Sep	102	96	1165	30.6	12.1	94.7	1.1	2.3
Oct	89	104	1088	35.0	10.4	85.6	1.7	2.5
Nov	77	81	875	29.1	10.8	71.1	4.3	1.9
Dec	87	81	1015	32.7	12.5	79.8	3.1	1.9
Department of Medicine (covering pediatrics, oncology, geriatrics, and skin ward VD)								

Table 3 Orthopedics Department monthly statistics

Month	Admissions	Discharge	Hospital bed	Average	ALS H/ (D−d)	BOR (N/B)*100	TOI	BTR
	+ Deaths	Days	Occupancy					
Jan	92	77	1044	33.6	13.5	82.1	2.9	1.8
Feb	97	77	1055	37.6	13.7	91.8	1.2	1.8
Mar	114	114	1485	47.9	13.0	116.8	−1.87	2.7
Apr	99	114	1139	37.1	9.9	92.6	0.7	2.7
May	88	86	851	27.4	9.8	66.9	4.8	2.0
June	66	77	843	28.1	10.9	68.5	5.0	1.8
July	112	101	1049	33.8	10.3	82.5	2.1	2.4
Aug	104	104	1095	35.3	10.5	89.0	1.6	2.5
Sep	91	102	1053	35.1	10.3	82.8	1.7	2.4
Oct	94	91	1003	32.3	11.0	78.9	2.9	2.2
Nov	95	98	1031	34.3	10.5	83.8	2.0	2.3
Dec	95	108	1084	34.9	10.0	85.3	1.7	2.6
Department of Medicine (covering pediatrics, oncology, geriatrics, and skin ward VD)								

In orthopedics, a hospital stay typically lasts 9–13 days. The reason for this may be that the hospital is long in orthopedic cases and it takes a long time for the bones to heal. The beds are at their best in March and are at their best in most months except May and June. The transition period in March was negative and January, February and June have the lowest bed turnover rate (1.8) (Table 3).

The typical hospital stay in ENT and ophthalmology varies from 5 to 11 days throughout the year; however, it is only 4 days in April. The number of people in the bed is low compared to the upper level. This can be attributed to the severity of patients who are usually treated as outpatients. The annual turnover period is also high. Highest bearing rotation speed (3.3%) is February (Table 4).

Overall average length of stay is good. All beds in this teaching hospital are limited. Total conversion time is more than a good number. This clearly shows that

Table 4 Monthly statistics of ophthalmology each month

Month	Admissions + Deaths	Discharge Days	Hospital bed Occupancy	Average	ALS H/(D−d)	BOR (N/B)*100	TOI	BTR
Jan	75	65	455	15	7	40.7	10.1	1.8
Feb	113	119	586	21	5	58.1	3.5	3.3
Mar	86	89	443	14	5	39.6	7.1	2.4
Apr	83	82	368	12	4	44.3	8.0	2.3
May	88	92	511	16	6	45.7	6.5	2.6
June	71	69	271	9	9	24.2	11.7	1.9
July	108	102	418	13	11	37.5	6.8	2.8
Aug	62	69	261	8	7	23.4	12.3	1.9
Sep	63	60	300	10	10	27.2	13.0	1.7
Oct	109	108	538	18	10	48.2	5.3	3.0
Nov	76	79	324	10	9	30.0	9.5	2.2
Dec	103	95	465	15	10	41.7	6.8	2.6

Table 5 Monthly statistics for whole infrastructure

Month	Hospital Days	Discharges + Deaths	ALS H/(D−d)	BOR (N/B)*100	TOI	BTR
Jan	8693	909	10	57	6.6	1.8
Feb	7810	919	9	57	6.4	1.8
Mar	8917	980	9	59	6.4	1.9
Apr	8102	898	9	54	7.5	1.8
May	8073	964	8	53	7.4	1.9
June	7371	864	9	50	8.5	1.7
July	8929	1089	8	59	5.8	2.2
Aug	7585	1030	7	50	7.1	2.0
Sep	8023	980	8	54	6.8	1.9
Oct	8405	1000	8	55	6.8	2.0
Nov	7408	902	8	50	8.1	1.8
Dec	8043	914	9	53	7.8	1.8

there are many patient beds. Therefore, the beds in the hospital were not used. This may be because the teaching hospital is a private hospital. Access to various departments is optional. This lowers the acceptance rate. Increasing hospital development has led to a decrease in hospital use (Table 5).

4 Proposed Method or Model: A Numerical Model for Plotting Advance Time Using Queues

For the three patient types in the given study, let λ_i be the attendance rate for three types of patients and μ_i be the level of service for type 1 patients, i = any natural number, which denotes the number of visits to each category. LOS follows a negative distribution with input and output from each other. An equation of birth and death can be created that describes the state of the body.

In this model, instead of treating the row as a linear row, we choose a circular row to increase the effectiveness of the model. The circular line is initially divided by n = 0 and then m = 0, i.e., the service fee is 0. Number of beds allocated = 0. During the design process, we chose a multilevel service model from the interconnected system, where beds are available due to the season or lower for various reasons.

4.1 Structure of the Model

When the patient is in line, fasting is observed. If it is an emergency, it is immediately queued from the main job. If the patient is under observation, he is placed at the end of the line so that he can follow the nonurgent phase. The nature of the line for the transition period was chosen as circular.

Let Pn, E$i(t)$ be the likelihood that at time t, in a ward with n patients, E beds of type 1 are filled by emergency, observation, and nonemergency levels.

Assume that m is the cutoff priority expressed in beds. Consequently, out of a total N beds, N − ($m1$ + $m2$) are reserved for type 1 patients. Upon considering the outcomes associated with different individuals occupying beds, the “m” value is computed. Since these equations are created by slightly modifying the familiar “birth-death” formula, we skip the examples here and limit the results to each distribution only.

4.2 Variables

1. Arrival rate of the patient (λi)
2. Service rate, i.e., allotment of beds with minimum time taken (μi)

3. No. of patients = n
4. Cutoff priority = m
5. Total no. of patients = N
6. Emergency level = (E*i*)

4.3 Goals

(i) Despite the fact that hospital bed utilization rates were often low.
(ii) Maximum utilization of hospital beds.
(iii) To provide complete information about the fluctuations during the usage of beds.
(iv) Maximum utilization of cutoff priority and efficient usage of circular queue pattern.
(v) To some parts, the low occupancy rate might be attributed to the vast number of small and large convalescent home, health center, and other facilities clustered in this restricted geographic area.

4.4 Formulation of Goal Constraints

If the “birth-death process” equations are rewritten on the left hand side of[P*n*, E*i*(*t* + Δ*t*) − P*n*, E*i*(*t*)]/Δ*t* and a limit has taken as Δ*t* → 0, we obtain a system of linear-differential equations in t and first order difference equations in n and E*i*, if one assumes that the study state conditions exist.

Define T*n*, E*i* = *Limt* → 0P*n*, E*i*(*t*).

Assuming these limiting probabilities, T*n*, E*i* exists independently of the initial state of the system. We can show that the resultant set of equations is satisfied by T*n*, E*i* with P*n*, E*i* = 0. The system of linear-differential equations for the various cases may be written as follows:

Case 1: $n = 0$, $m = 0$, $E_i = 0$

$$-(\lambda_1 + \lambda_2)T0,0 + \mu_1\left[T_{1,1} + T_{2,2}\right] + \mu_2\left[T_{1,2} + T_{2,1}\right]$$

Case 2: $1 \le n \le m$, $E_i = n$

$$-\left[(l_1 + l_2) + E_i m_2 + (n - E_i)m_1\right]T_{n,E_i} + \left[n + 1 - E_i\right]m_1 T_{n+1,E_i} + \left[E_i + 1\right]m_2 T_{n+1,E_i+1}$$
$$+ l_1 T_{n-1,E_i} + l_2 T_{n-1,E_i-1} = 0$$

Case 3: $[n = m, E_i < m]$

$$-\left[l_1 + \mathrm{E}_i m_2 + \left[m - \mathrm{E}_i\right] m_1\right] \mathrm{T}_{m,\mathrm{E}_i} + \left[m + 1 - \mathrm{E}_i\right] m_1 \mathrm{T}_{m+1,\mathrm{E}_i} + \left[\mathrm{E}_i + 1\right] m_2 \mathrm{T}_{m+1,\mathrm{E}_i+1}$$
$$+ l_1 \mathrm{T}_{m-1,\mathrm{E}_i} + l_2 \mathrm{T}_{m+1,\mathrm{E}_i-1} = 0$$

Case 4: $m < n < \mathrm{N}, \mathrm{E}_i < n$

$$-\left[\lambda_1 + \mathrm{E}_i \mu_2 + \left[n - \mathrm{E}_i\right] \mu_1\right] \mathrm{T}_{m,\mathrm{E}_i} + \left[n + 1 - \mathrm{E}_i\right] \mu_1 \mathrm{T}_{m+1,\mathrm{E}_i} + \left[\mathrm{E}i + 1\right] \mu_2 \mathrm{T}_{m+1,\mathrm{E}_i+1} = 0$$

Case 5: $[n = \mathrm{N}, \mathrm{E}_i \le m, n \ne \mathrm{N}]$

$$-\left[\mathrm{E}_i \mu_2 + \left[\mathrm{N} - \mathrm{E}_i\right] \mu_1\right] \mathrm{T}_{\mathrm{N},\mathrm{E}_i} + \lambda_1 \mathrm{T}_{m-1,\mathrm{E}_i} = 0$$

Case 6: $[\mathrm{E}_i = n \text{ or } \mathrm{E}_i = m \text{ or } \mathrm{E}_i = \mathrm{N}]$

$$-\left[\lambda_1 + m\mu_2\right] \mathrm{T}_{m,n} + \mu_1 \mathrm{T}_{m,+1,mi} + \lambda_2 \mathrm{T}_{m,+1,m-1} = 0$$

$$-\left[\lambda_1 + m\mu_2\right] + \left[n - m\right] \mu_1 \mathrm{T}_{n,m} + \left[n + 1 - m\right] \mu_1 \mathrm{T}_{m+1,mi} + \lambda_1 \mathrm{T}_{m-1,m} = 0$$

5 Result and Discussion

Using the information given to the patients and service personnel of different beds, the design was solved using QSB+ computer software. It was observed that the solution time increased rapidly as the number of beds in the service increased. The large increase in computation time required to solve the problem is directly due to the business matrix inversion subroutine, which uses an efficient inversion subroutine that takes advantage of the approximately three diagonal structures of the matrix. This time can be greatly reduced by using an advanced model with a larger bearing. The results for the eight-bed ward are summarized in Table 6.

Table 6 Results obtained

Cutoff occupancies	Shortage probabilities		Expected no. of beds		Total
	$Ø_1$	$Ø_2$	$\sum_{n=0}^{M}\sum_{n2=0}^{n}(N-m)Tn,n2$	$\sum_{n=\mathrm{m}+1}^{N}\sum_{n2=0}^{n}(N-m)Tn,n2$	
1	0.0935	0.8122	1.2189	2.1176	3.335
2	0.0960	0.7000	1.7801	1.5086	3.2887
3	0.0990	0.5716	2.3050	0.9567	3.2617
4	0.1026	0.4256	2.7692	0.4923	3.2615
5	0.1068	0.2664	3.1380	0.1596	3.2976
6	0.1072	0.1072	3.4857	0.000	3.4857

Table 7 Values of objective function for various holding and shortage costs

	C(m, N)				
Value of cutoff occupation	Case 1	Case 2	Case 3	Case 4	Case 5
M	$C_1 = 1$ $C_2^1 = 1.5$ $C^2_2 = 1$	$C_1 = 1$ $C_2^1 = 2$ $C^2_2 = 1$	$C_1 = 1$ $C_2^1 = 8$ $C^2_2 = 2$	$C_1 = 1$ $C_2^1 = 6$ $C^2_2 = 3$	$C_1 = 1$ $C_2^1 = 8$ $C^2_2 = 4$
1	4.2695	4.7184	6.7021	6.5926	7.6847
2	4.1707	4.3231	6.5862	6.3919	7.4262
3	4.0797	4.2383	6.4820	6.1914	7.1679
4	4.0095	4.1735	6.3997	5.9924	6.9097
5	3.9706	4.1409	6.350	5.8277	6.6700
6	4.0667	4.2361	6.3564	5.7369	6.4557

We obtained the values of the objective function C(m, N) given in Table 7 by assigning relative values of shortage weights (C^1_2, C^2_2) and holding cost C_1. The data presented in Table 7, reveals that the most favorable operational costs of the ward in cases 1, 2, and 3 are achieved by implementing a cutoff priority at seven beds. In cases 4 and 5, when shortage cost for type 2 patients is increased, the optimum cutoff priority takes a value of eight beds. That is, no bed is returned for type 1 patients. This is what we expect if the relative penalty cost for type 2 is increased beyond a certain value, which in our case appears to be 2.

6 Conclusion

In this study on the use of beds in hospitals, although beds are often low, it was found that more than 10% of beds were available in some wards in some months, indicating that beds are scarce when available. In other months, the price is lower. As the hospitals demonstrate, nursing homes large and small focus on this small area as part of their responsibility to the limited population. A monthly analysis allows leaders to plan ahead for such events and make necessary adjustments for a particular church or a particular season. Again, considering the number of beds or the situation in some wards, the change time is very high in some cases.

References

1. Hammond, C. L., Pinnington, L. L., & Phillips, M. F. (2009). A qualitative examination of inappropriate hospital admissions and lengths of stay. *BMC Health Services Research, 9*, 1–9.
2. Greene, J. K., & Metwalli, A. (2001). The impact of activity based cost accounting on health care capital investment decisions. *Journal of Health Care Finance, 28*(2), 50–64.
3. Greenfield, H. I. (1973). *Hospital efficiency and public policy*. Praeger Publishers.
4. Harrison, J. P., Coppola, M. N., & Wakefield, M. (2004). Efficiency of federal hospitals in the United States. *Journal of Medical Systems, 28*, 411–422.

5. Chen, M. M., & Bush, J. W. (1971). *A mathematical programming approach for selecting an optimum health program case mix*. Presented before the Operations Research Society of America.
6. Durbin, R. L., & Springall, W. H. (1974). *Organization and administration of health care: Theory, practice, environment*. Mosby Incorporated.
7. Robinson, J. C. (1996). The dynamics and limits of corporate growth in health care. *Health Affairs, 15*(2), 155–169.
8. Knobel, R. J., & Longest, B. B. (1972). Cost-benefit assumptions in analysis for hospitals. *Financial Management, 1*(1), 63.
9. Ross, T. K. (2004). Analyzing health care operations using ABC. *Journal of Health Care Finance, 30*(3), 1–20.
10. Levey, S., & Loomba, N. P. (1973). *Health care administration: A managerial perspective*. J.P. Lippincott Co.

Early Detection of Diabetic Retinopathy Using Deep Learning

Tanmay Patil, Rushikesh Kundkar, Sarvadnya Pande, Yash Katkamwar, Amit Joshi, and Suraj Sawant

Abstract Diabetic retinopathy is a major cause of blindness in diabetic individuals aged 25–65, where lesions on the retina caused by weakened blood vessels can lead to visual loss and even total blindness. Current manual grading methods for detecting diabetic retinopathy are time-consuming and error-prone. Convolutional neural networks have shown significant promise for automating identification and categorization of diabetic retinopathy. Google AI, IDx-Diabetic Retinopathy, Eyenuk, and VoxelCloud are among the current approaches using convolutional neural networks to detect diabetic retinopathy with great accuracy and speed, saving time and money over manual diagnosis. This study aims to develop a convolutional neural network-based model using retinal fundus pictures for detecting diabetic retinopathy. A sizable dataset of retinal pictures will be used to train the model on a high-powered computing architecture provided by a graphics processing unit. The model's performance will be evaluated using multiple performance metrics such as accuracy, sensitivity, and specificity to increase performance in classification tasks. The proposed solution has the potential to improve diabetic retinopathy detection and enable faster and more reliable interpretation of retinal images, saving time and money over manual diagnosis. However, further research is required to improve these models and incorporate them into clinical practice.

1 Introduction

Excessive blood sugar levels are a defining feature of the chronic disease called diabetes. By 2021, there will be 537 million diabetics globally, with type 2 diabetes accounting for more than 90% of cases, according to the International Diabetes

---

T. Patil (✉) · R. Kundkar · S. Pande · Y. Katkamwar · A. Joshi · S. Sawant
Department of Computer Science & Engineering,
COEP Technological University (COEP Tech), Pune, India
e-mail: patilts19.comp@coep.ac.in

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_8

Federation (IDF). Diabetes also kills around 4.2 million people each year, making it a significant global health issue. Diabetes-related retinopathy (DR) is an eye-related consequence of diabetes. The blood vessels in the retina, the part of the eye that receives light and sends information to the brain, are damaged when blood sugar levels are high. The injured blood vessels may leak fluid or blood over time, producing swelling and impairing eyesight. In certain situations, aberrant blood vessels may form on the retina, causing scarring and retinal detachment and, if ignored, significant vision loss or blindness. DR may affect anybody with diabetes, type 1 or type 2, and it is more common in persons who have had diabetes for a long period or have poorly regulated blood sugar levels. Early identification and treatment of DR are critical for preventing or slowing visual loss. Diabetic patients should get a full dilated eye exam at least once a year to detect indicators of DR [1].

DR is divided into four phases:

Mild nonproliferative DR (NPDR): Little regions of edema or microaneurysms form in the retinal blood vessels at this early stage. Certain blood veins may leak fluid or blood into the retina, producing abnormalities in vision. Nonetheless, at this time, there are generally no symptoms [2].

Moderate nonproliferative DR (NPDR): This occurs when the disease progresses and blood arteries that feed the retina become partially clogged. This can deplete the retina of oxygen and nutrients, triggering the formation of new vascular structures in the retina.

Severe NPDR: At this stage, the retina's blood flow is drastically reduced as many blood vessels get blocked. This may lead to the development of new, abnormal blood vessels on the retina, which might cause further problems.

Proliferative DR (PDR): This is the most advanced stage of DR, where new blood vessels are formed on the surface of the retina, which are fragile and quickly bleed blood into the vitreous humor. The creation of scar tissue caused by the establishment of these new blood vessels can result in retinal detachment, macular edema, and significant vision loss or blindness [3].

Deep learning-based techniques are particularly well-suited for DR detection because they can learn features from huge datasets automatically and provide precise and reliable predictions of DR from fundus pictures [4]. Images of the retina have been successfully analyzed using deep learning algorithms [5]. Nevertheless, constructing an effective deep learning-based DR detection system necessitates a large dataset, strong feature extraction methods, and good classification algorithms [6].

A type of deep learning method called convolutional neural networks (CNNs) uses many layers of artificial neurons to extract characteristics from pictures automatically. CNNs have shown tremendous potential in identifying and diagnosing DR from retinal pictures [7].

CNNs are trained on massive datasets of retinal images to learn disease-related patterns and characteristics [8]. The RA-EfficientNet model created by Kunming University researchers [9] is one of the most successful instances of a CNN for DR detection. This algorithm was trained on a large dataset of retinal pictures and showed a high level of accuracy in recognizing DR and its severity.

Overall, CNNs might greatly increase the precision and effectiveness of DR diagnosis, possibly resulting in earlier patient identification and better treatment outcomes [10].

There are five sections to the study. The introductory part offers a summary of diabetic retinopathy (DR) disease and the current technologies used to identify it, such as convolutional neural networks and deep learning. The second portion is a literature review, which examines existing models for DR detection.

The approach employed in the study is presented in the third part. It contains the proposed model, the dataset utilized, as well as the preparation and training techniques. The experimental setup, performance measures, and outcome comparison are all included in the fourth section, which is the discussion and conclusion.

Lastly, in the fifth segment, the research is concluded with closing views on the study and its future scope.

2 Literature Review

The main factor in blindness worldwide is diabetic retinopathy (DR), a common chronic diabetes complication. Early identification and treatment of DR can help to avoid visual loss. Deep learning-based techniques to detect and diagnose DR using retinal pictures have shown potential. Numerous researchers have looked at using deep learning-based algorithms to classify DR from retinal pictures.

Sabeenian et al. recommended modifying the original AlexNet design by raising the number of output classes from 1000 to 5, corresponding to the various stages of DR. For DR classification, the modified AlexNet architecture outperformed other cutting-edge deep learning techniques, with an accuracy of 89.6% on the testing set. The model's accuracy for the IDRiD dataset [11] was 95.60%.

Similarly, Sarwinda et al. proposed using the GoogleNet architecture to identify DR in fundus pictures. The authors preprocessed the fundus pictures and applied transfer learning to fine-tune the GoogleNet architecture for DR detection. For the DIARETDB1 dataset, the suggested architecture has an accuracy of 97.42%. The work emphasizes the potential of transfer learning and the GoogleNet architecture for accurate and efficient DR detection from fundus pictures [12].

Aatila et al. looked at using pretrained ResNet50 and VGG-16 networks for DR classification. The suggested method attained an accuracy of 93.6% on the testing set using the ResNet50 model and 91.6% using the VGG-16 model. According to the authors, ResNet50 outperformed VGG-16 in terms of sensitivity, accuracy, and specificity [13].

For the Messidor-2 dataset, Yadav et al. employed an Inception-V3 model to identify DR from fundus pictures, and 96.20% accuracy was attained. The Inception-V3 model used in this work serves as evidence of the advantages of using deep learning models in DR designed specifically for image recognition tasks [14].

A deep learning model for DR detection is created by Pavate et al., from fundus pictures based on the MobileNet architecture. The researchers suggested a binary classifier that can tell the difference between normal and DR fundus pictures. Using a dataset of 400 fundus images, the proposed model was trained and assessed, attaining a testing accuracy of 94.5%. The work shows that the binary classifier method to MobileNet architecture can provide a simple and effective tool for first diabetes screening [15].

Using a mix of DenseNet-65 and Faster-RCNN architectures, for the identification and detection of lesions connected to DR, Albahli et al. proposed a deep learning approach. A dataset of 2000 fundus images was used to train and test the recommended model, and its DR detection accuracy was 93.6 [16].

A deep learning model was proposed by Yi et al., based on the RA-EfficientNet architecture for DR diagnosis. A dataset of 46,506 fundus images served as the training and evaluation data for the recommended model, and it achieved a DR diagnostic accuracy of 96.97%. The study shows that the RA-EfficientNet model outperforms existing deep learning models including VGG16, Inception-v3, and ResNet50 in terms of efficiency and accuracy for DR diagnosis [17].

Jindal et al. suggested a method for identifying diabetic retinopathy in 2D color fundus retina images using an ensemble-based approach. The authors started by extracting features with deep learning-based methods and handmade features. The authors tested their suggested technique on a publically accessible dataset and discovered that it performed better than alternative approaches in terms of sensitivity, accuracy, area under the receiver operating characteristic curve, and specificity [18].

Apart from using deep learning to detect DR, several research have focused on detecting particular traits linked with DR. Hassan et al. looked into the application of fractal analysis to detect DR-related alterations in the retinal vasculature [19]. To detect and diagnose DR, Röhlig suggested a visual analytics technique that integrated deep learning with interactive visualization [20]. Jena et al. [21] employed asymmetric deep learning features for DR screening. For DR detection, Band et al. employed Bayesian deep learning [22].

Some investigations have concentrated on segmenting retinal pictures to identify DR. Pekala et al. proposed a deep learning-based method for segmenting retinal OCT images into different retinal layers with a mean dice similarity coefficient (DSC) of 0.925. According to the authors, their model was highly accurate in segmenting retinal layers and recognizing DR characteristics [23]. Hao et al. developed a multitask learning strategy based on voting for detecting retinal structures in OCT angiography (OCTA) pictures [24]. Using deep learning approaches, in ultra-widefield fundus images, Wang et al. offer an intelligent diagnostic approach for locating multiple peripheral retinal abnormalities. Deep learning algorithms can accurately segment retinal pictures and detect DR characteristics, according to the experiments shown above [25].

3 Proposed Methodology

This section discusses the study's objectives as well as the model's properties using the dataset used. After that, we discuss data preparation and training approaches.

The objectives to which we are referring are as follows:

Diagnosis: CNNs may be trained to differentiate diabetic retinopathy from other retinal diseases or normal retinas. This can help to produce more accurate and reliable diagnoses.

Diabetic retinopathy screening in a large population can be a challenging and time-consuming process. CNNs might be used to construct an automated screening system that can quickly and accurately identify patients who need further evaluation.

3.1 Model

This model is a convolutional neural network (CNN) shown in Fig. 1 designed to categorize images. It is made up of a number of convolutional layers, followed by batch normalization and pooling layers. The first layer takes a 224 × 224 × 3 input image and applies 16 3 × 3 filters with ReLU activation, generating 16 feature maps. By picking the largest value in each 2 × 2 region, the max pooling layer reduces the dimensionality of the feature maps. To normalize the activations of the prior layer, batch normalization is utilized.

With each consecutive convolutional layer, this method is repeated with an increasing number of filters, resulting in a more comprehensive and sophisticated representation of the source image. The final layer of convolution produces 256 feature maps that are flattened into a one-dimensional array.

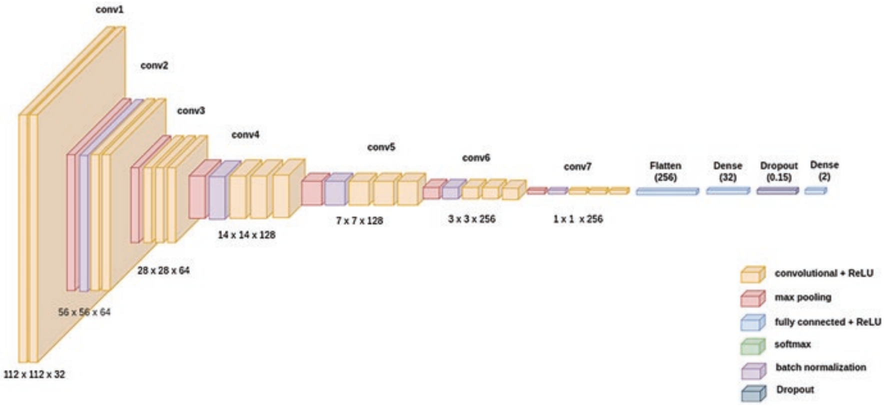


Fig. 1 CNN layers

3.2 Dataset

We used the DR dataset [26] from Kaggle, which is a collection of fundus images from patients with DR and healthy controls. It was initially obtained through EyePACS, a nonprofit organization dedicated to increasing access to eye care for marginalized communities.

The collection contains 88,702 fundus photos, of which 28,709 are diagnosed with DR and 60,993 are controls. The photographs were taken with several types of cameras and show a variety of quality and lighting situations. The collection is unbalanced, with more photos of healthy eyes than DR eyes.

Each image in the collection has been labeled by trained human graders dependent on how severe DR is, using a five-level scale:

The collection is distributed in 9 zip files, each containing around 10,000 photos. The photos are identified by a unique ID, and their labels are given in a separate CSV file.

The dataset has been heavily used to create and test machine learning methods for automated DR identification along with grading. Its huge size and wide variety of pictures make it a significant resource for medical imaging researchers.

3.3 Preprocessing

Image preprocessing is an important stage in the diabetic retinopathy diagnosis procedure utilizing CNN. Because the lighting, contrast, and noise levels in the input photos may vary, preprocessing is required to improve image quality and model performance.

Preprocessing stages often include scaling the pictures to a predetermined size, commonly 224×224 , to match the CNN model's input size. To lessen the influence of fluctuations in lighting and contrast, with a mean of 0 and a standard deviation of 1, the images can be normalized.

The following stages are included in image preprocessing:

Image resizing: The initial retinal pictures might be rather huge. Image resizing can enhance processing speed and minimize the amount of computer resources required. The process of resizing the image is depicted in the following figures, Figs. 2 and 3, even if there is only a change in dimension.

Image filtering: Retinal scans contained noise, artifacts, and other undesired traits that might impede further investigation. To eliminate these imperfections and improve image quality, we employed Gaussian filtering techniques. The image filtering method is best shown in Figs. 4, 5 and 6.

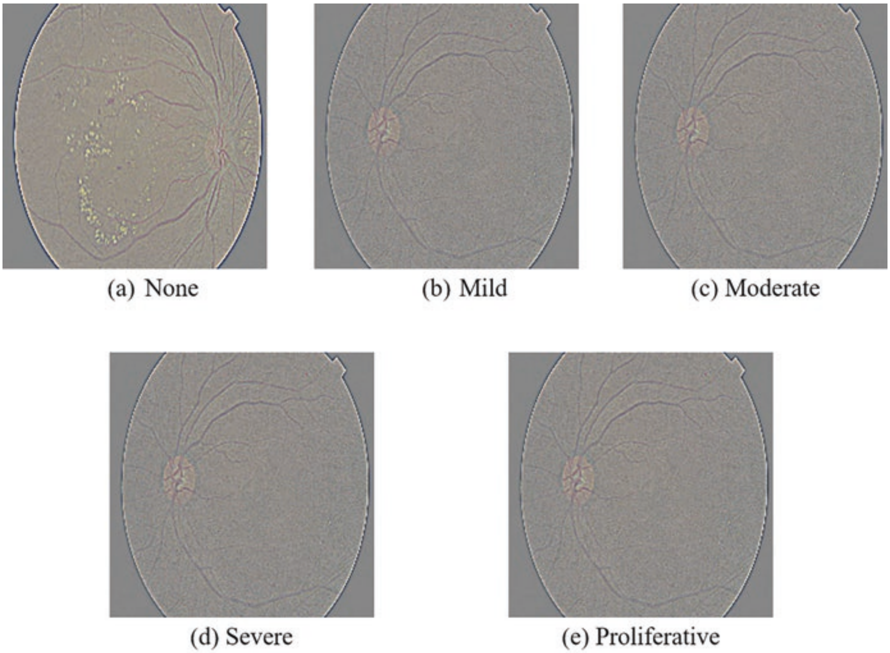
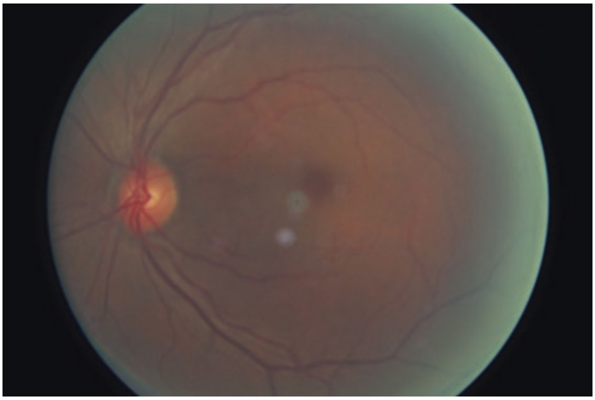

(a) None (b) Mild (c) Moderate (d) Severe (e) Proliferative

Fig. 2 Images corresponding to different severity of DR

Fig. 3 Actual fundus image



3.4 Training

Using a collection of retinal pictures from diabetic patients with varied degrees of diabetic retinopathy, we trained our CNN architecture. All photos were scaled to 224 × 224 × 3 pixels and normalized with pixel values ranging from 0 to 1.

This study used the Adam optimization strategy to train the network for 80 epochs with a batch size of 32 and a learning rate of 0.00001. We used early

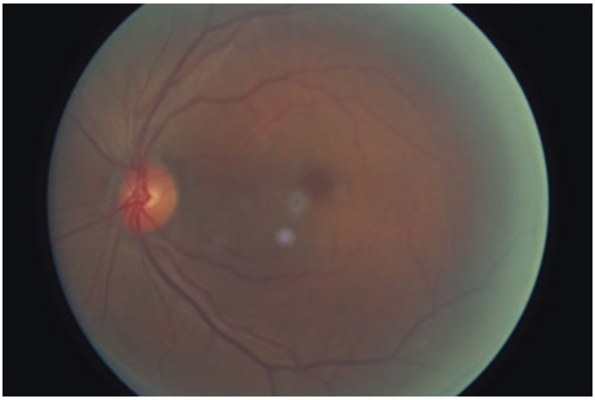


Fig. 4 Resized fundus image

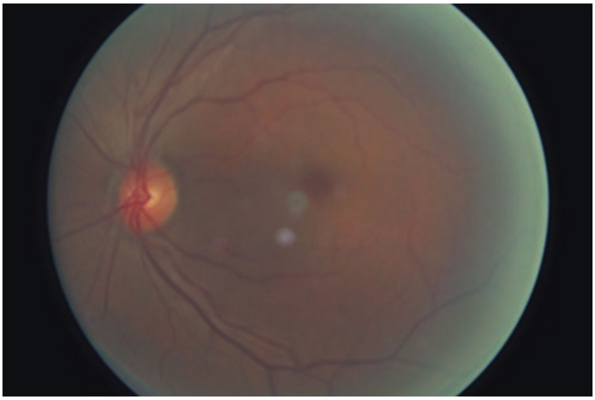


Fig. 5 Resized fundus image

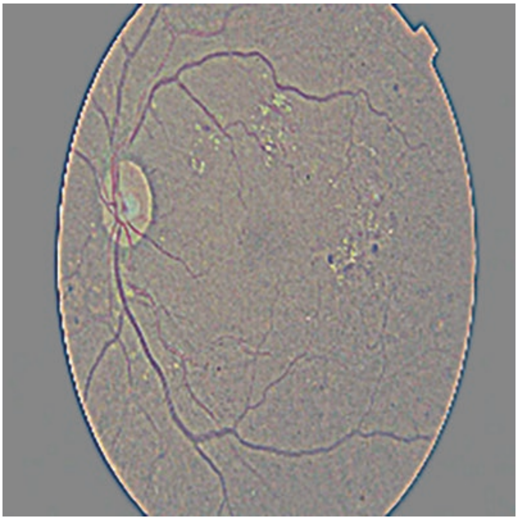


Fig. 6 Filtered fundus image

stopping, monitored the validation loss, and stopped training if it did not improve for ten consecutive epochs.

We also used data augmentation techniques like random rotations, flips, and zooms during training to boost the variety of the training set and minimize overfitting. The loss function was categorical cross-entropy, and the assessment metrics were area under the receiver operating characteristic curve (AUC-ROC), sensitivity, specificity, and accuracy.

To avoid overfitting, we applied 0.15 dropout regularization before the final thick layer. To increase convergence and avoid overfitting, we also applied batch normalization after each convolutional layer.

Many hyperparameters, including the count of filters in each convolutional layer, were investigated with this study along with filter size and density layer unit count. The final hyperparameters were chosen based on their performance on the validation set.

Following training, we examined CNN's performance on the test set and compared it to state-of-the-art approaches for diabetic retinopathy identification.

4 Results and Discussion

This part will go over the experimental setup needed for our model, the performance metrics employed in model evaluation, and the comparison of outcomes for several models.

4.1 Experimental Setup

In this study, we aim to develop an accurate and efficient CNN model for DR detection. The experimental setup involves several stages, including dataset collection, data preprocessing, dataset splitting, model architecture design, training, and evaluation.

As the main dataset in this work, the Kaggle DR detection dataset is used to construct a CNN-based DR detection algorithm. The dataset is initially preprocessed by shrinking the photos to a standard size and using Gaussian filters to normalize pixel intensities. Circular cropping and other picture alterations are also used to improve image quality. Thereafter, training, validation, and testing sets are created from the dataset. Although the validation set is used to adjust hyperparameters and prevent overfitting, the training set is utilized to train the CNN model. To rate the effectiveness of the finished model, the testing set is utilized.

The CNN model architecture is intended to be precise enough to distinguish between diabetic and nondiabetic retinal pictures. The architecture is deep enough to capture the complexities of retinal pictures, but not so deep that overfitting occurs. After being trained on the training set, the model is then utilized to adjust the hyperparameters such as learning rate and batch size on the validation set. The training is done on a GPU with four Nvidia Tesla V-100 cards, each having 32,505 MB of RAM and a computational capacity of 7.0.

On the testing set, the trained model's performance is evaluated using measures like accuracy, F1-score, AUC, and Cohen kappa coefficient. These measures are used to evaluate the model's performance as well as its capacity to generalize to new, previously unknown data. Ultimately, the goal is to create a comprehensive and reliable DR detection model that doctors may utilize to help them diagnose and treat diabetic retinopathy.

4.2 Performance Metrics

The efficiency of a CNN model for DR detection may be assessed using a variety of performance criteria. Here are some of the commonly used performance metrics:

1. Accuracy

   The most used statistic for assessing a classification model's performance is accuracy, including CNN models for DR detection. Accuracy measures the percentage of correctly classified images. Our model gives an accuracy of 94.72%.
2. F1-Score

   The accuracy of a model is assessed using the F1-score, which also considers precision and recall. It is calculated using the accuracy and recall harmonic means:

$$F1\,\text{Score} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \tag{1}$$

where the ratio of true positive results to true positive and false positive results is known as accuracy and the ratio of true positive results to true positive and false negative results is known as recall.

In other words, the F1-score is a measure of the balance between precision and memory, and it ranges from 0 to 1, wherein 1 denotes absolute accuracy and recall. Our model gives an F1-score of 0.94.

3. Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

   AUC-ROC measures the ability of the model to distinguish between positive and negative cases. AUC-ROC provides a single number that summarizes the overall performance of the model across different thresholds. Our model gives an AUC value of 0.97.

$$AUC\_ROC = \int_{1}^{0} TPR\left(FPR^{-1}(u)\right) du \tag{2}$$

where *TPR* and *FPR* stand for true positive rate and false positive rate, respectively. *TPR* is the proportion of real positive samples that the model properly identifies as positive, whereas *FPR* is the proportion of real negative samples that the model erroneously identifies as positive.

4. Cohen Kappa Coefficient (κ).

   A statistical indicator of inter-rater agreement between two raters or judges is the Cohen kappa coefficient. It is used to evaluate the agreement beyond chance that two raters have when they are rating or categorizing the same set of items. Our model gives a coefficient value as 0.88.

   Cohen kappa coefficient is calculated using the formula:

$$\kappa = \frac{\left(p_o - p_e\right)}{\left(1 - p_e\right)} \tag{3}$$

$$p_o = \frac{\left(a + d\right)}{\left(a + b + c + d\right)} \tag{4}$$

$$p_e = \left(\frac{\left(a+b\right)}{\left(a+b+c+d\right)} \cdot \frac{\left(a+c\right)}{\left(a+b+c+d\right)}\right) + \left(\frac{\left(c+d\right)}{\left(a+b+c+d\right)} \cdot \frac{\left(b+d\right)}{\left(a+b+c+d\right)}\right) \tag{5}$$

where

a denotes number of items that both raters rated positively.
d is for items that both raters rated negatively.
b is count of items that rater 1 rated positively and rater 2 rated negatively.
c is the count rater 1 rated negatively and rater 2 rated positively.

4.3 Results

The performance of three separate convolutional neural network (CNN) models, VGG-16, Densenet65, and a proposed model, was evaluated for diabetic retinopathy detection using four different metrics: accuracy, F1-score, Cohen kappa coefficient, and area under the receiver operating characteristic curve (AUC-AOC).

In terms of F1-score, accuracy, and Cohen kappa coefficient, the recommended model performs better than both Densenet65 and VGG-16.

Table 1 provides a quick summary of performance measures for various models, as well as a comparison measure between well-known accessible models and the one suggested in the study.

The Densenet65 model has a 93.6% accuracy and an F1-score measured to be 0.83 compared to the VGG-16 model's 91.6% accuracy and 0.95 as F1-score. VGG-16 and Densenet65 were outperformed by the proposed model, with 94.72% accurate results and a decent 0.94 as F1-score. Furthermore, when compared to

Table 1 Performance metrics comparison for well-known models

Model	Accuracy (%)	F1-score	AUC-AOC	Cohen kappa coefficient
VGG-16	91.60	0.95	0.98	NA
Densenet-65	93.60	0.83	0.97	0.76
Proposed solution	94.72	0.94	0.97	0.88

VGG-16 and Densenet65, the suggested model acquired a higher Cohen kappa coefficient of 0.88, showing greater agreement between projected and actual labels.

In terms of AUC-AOC, all three models performed similarly, with the VGG-16 model getting the highest AUC-AOC of 0.98, closely followed by Densenet65 and the suggested model, all of which reached 0.97. While not having the maximum AUC-AOC value, the suggested model obtained a decent mix of sensitivity and specificity.

In terms of Cohen kappa coefficient, accuracy, and F1-score, the model proposed outperformed VGG-16 and Densenet65 while keeping comparable performance in terms of AUC-AOC. These findings indicate that the suggested model might be a potential method for detecting diabetic retinopathy, which is critical for early identification and prevention of vision loss in diabetic patients. Further research is needed, however, to validate the suggested model on bigger datasets and in real-world clinical settings.

5 Conclusion and Future Work

In this study, we trained a CNN model to recognize diabetic retinopathy using retinal fundus images. As compared to the commonly used VGG-16 and Densenet-65 models, our suggested technique attained a much better accuracy of 94.72%. Our model's strong F1-score suggests that it is trustworthy in recognizing positive instances while limiting false positives, making it suited for medical applications where false positives might have serious repercussions.

Despite the fact that our proposed strategy produced promising outcomes, there is yet space for improvement. One possible future route for research is to integrate additional preprocessing techniques, such as contrast normalization and sharpening, to increase picture quality and model performance. Another potential path is to investigate other CNN designs and fine-tune hyperparameters to increase the model's accuracy.

Furthermore, because detecting diabetic retinopathy is crucial for preventing vision loss, the practical use of our suggested technique should be examined further. To assess the model's utility in a real-world setting, future research may make use of larger datasets and a wider range of participants. Furthermore, the model's interpretability might be improved by employing techniques such as attention mechanisms to emphasize portions of the image that are most suggestive of diabetic retinopathy, assisting ophthalmologists in their diagnosis.

Acknowledgments We are thankful to the Department of Computer Science & Engineering, COEP Tech, for providing the GPU Server Facility for our work which is established under TEQTEQIPIP III (a World Bank project).

Declaration The authors do not have conflict of interests for the proposed work. The work has not received funding from any agency. The authors have not executed any experimentation that causes harm to any unit of nature.

References

1. Stitt, A. W., Curtis, T. M., Chen, M., Medina, R. J., McKay, G. J., Jenkins, A., et al. (2016). The progress in understanding and treatment of diabetic retinopathy. *Progress in Retinal and Eye Research, 51*, 156–186.
2. Sabanayagam, C., Banu, R., Chee, M. L., Lee, R., Wang, Y. X., Tan, G., et al. (2019). Incidence and progression of diabetic retinopathy: A systematic review. *The Lancet Diabetes & Endocrinology, 7*(2), 140–149.
3. Aiello, L. M. (2003). Perspectives on diabetic retinopathy. *American Journal of Ophthalmology, 136*(1), 122–135.
4. Ramos, J. M. A., Perdómo, O., & González, F. A. (2022). Deep semi-supervised and self-supervised learning for diabetic retinopathy detection. arXiv preprint arXiv:2208.02408.
5. Badar, M., Haris, M., & Fatima, A. (2020). Application of deep learning for retinal image analysis: A review. *Computer Science Review, 35*, 100203.
6. Rahim, S., Sabri, K., Ells, A., Wassyng, A., Lawford, M., Chu, L., & He, W. (2023). Novel Fundus image preprocessing for Retcam images to improve deep learning classification of retinopathy of prematurity. arXiv preprint arXiv:2302.02524.
7. Luo, X., Wang, W., Xu, Y., Lai, Z., Jin, X., Zhang, B., & Zhang, D. (2023). A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence. *CAAI Transactions on Intelligence Technology*.
8. Bajwa, A., Nosheen, N., Talpur, K. I., & Akram, S. (2023). A prospective study on diabetic retinopathy detection based on modify convolutional neural network using fundus images at Sindh Institute of Ophthalmology & Visual Sciences. *Diagnostics, 13*(3), 393.
9. Yi, S. L., Yang, X. L., Wang, T. W., She, F. R., Xiong, X., & He, J. F. (2021). Diabetic retinopathy diagnosis based on RA-EfficientNet. *Applied Sciences, 11*(22), 11035.
10. Nazir, T., Nawaz, M., Rashid, J., Mahum, R., Masood, M., Mehmood, A., et al. (2021). Detection of diabetic eye disease from retinal images using a deep learning based CenterNet model. *Sensors, 21*(16), 5283.
11. Salamat, N., Missen, M. M. S., & Rashid, A. (2019). Diabetic retinopathy techniques in retinal images: A review. *Artificial Intelligence in Medicine, 97*, 168–188.
12. Shanthi, T. Modified Alexnet architecture for classification of diabetic retinopathy images. *Computers & Electrical Engineering, 76*, 56–64.
13. Salma, A., Bustamam, A., & Sarwinda, D. (2021). Diabetic retinopathy detection using GoogleNet architecture of convolutional neural network through fundus images. *Nusantara Science and Technology Proceedings*, 1–6.
14. Aatila, M., Lachgar, M., Hrimech, H., & Kartit, A. (2021). Diabetic retinopathy classification using ResNet50 and VGG-16 pretrained networks. *International Journal of Computer Engineering and Data Science (IJCEDS), 1*(1), 1–7.
15. Yadav, S., & Awasthi, P. (2022). Diabetic retinopathy detection using deep learning and inception-v3 model. Int. Res. J. Mod. Eng. Technol. Sci, 4, 1731–1735.
16. Pavate, A., Mistry, J., Palve, R., & Gami, N. (2020). Diabetic retinopathy detection-MobileNet binary classifier. *Acta Scientific Medical Sciences, 4*(12), 86–91.

17. Albahli, S., Nazir, T., Irtaza, A., & Javed, A. (2021). Recognition and detection of diabetic retinopathy using densenet-65 based faster-RCNN. *Computers Materials & Continua, 67*, 1333–1351.
18. Jindal, H., Jain, S., & Aggarwal, A. (2023). *Ensemble diabetic retinopathy detection in 2-D Color fundus retina scan*.
19. Hassan, M., Zhang, H., Ameen, A. F., Zeng, H. W., Ma, S., Liang, W., ... & Qin, P. W. (2023). Futuristic variations and analysis in fundus images corresponding to biological traits. arXiv preprint arXiv:2302.03839.
20. Röhlig, M., Stachs, O., & Schumann, H. (2022). Visual analytics for early detection of retinal diseases. arXiv preprint arXiv:2212.10566.
21. Jena, P. K., Khuntia, B., Palai, C., Nayak, M., Mishra, T. K., & Mohanty, S. N. (2023). A novel approach for diabetic retinopathy screening using asymmetric deep learning features. *Big Data and Cognitive Computing, 7*(1), 25.
22. Band, N., Rudner, T. G., Feng, Q., Filos, A., Nado, Z., Dusenberry, M. W., ... & Gal, Y. (2022). Benchmarking Bayesian deep learning on diabetic retinopathy detection tasks. arXiv preprint arXiv:2211.12717.
23. Pekala, M., Joshi, N., Liu, T. A., Bressler, N. M., DeBuc, D. C., & Burlina, P. (2019). Deep learning based retinal OCT segmentation. *Computers in Biology and Medicine, 114*, 103445.
24. Hao, J., Shen, T., Zhu, X., Liu, Y., Behera, A., Zhang, D., et al. (2022). Retinal structure detection in OCTA image via voting-based multitask learning. *IEEE Transactions on Medical Imaging, 41*(12), 3969–3980.
25. Wang, T., Liao, G., Chen, L., Zhuang, Y., Zhou, S., Yuan, Q., et al. (2023). Intelligent diagnosis of multiple peripheral retinal lesions in ultra-widefield fundus images based on deep learning. *Ophthalmology and therapy*, 1–15.
26. Kaggle. (n.d.). *Diabetic retinopathy detection*. Retrieved September 12, 2021, from https://www.kaggle.com/competitions/diabetic-retinopathy-detection/data

Performance Analysis of Memory-Efficient Vision Transformers in Brain Tumor Segmentation

Atharva Marathe, Vishwajit Kadam, Aarya Chaumal, Sairaj Kodilkar, Amit Joshi, and Suraj Sawant

Abstract Convolutional neural networks are the most in demand solution for computer-aided medical image-segmentation applications and have attained excellent results. In recent years, transformer architectures, conventionally used in natural language processing tasks, have been adapted for computer vision tasks. These Vision Transformers have demonstrated propitious results in image processing and have recently been extended to medical image segmentation. Vision Transformers can scale on high-resolution images and incorporate more global information by self-attention as compared to convolutional neural networks. However, as they are known for their high memory consumption, several memory-efficient transformer architectures have been proposed to reduce memory requirements. These efficient models can process the images with remarkable accuracy while maintaining a low memory footprint. This work evaluates the performance and analyzes the memory utilization of memory-efficient transformer architecture, such as Efficient Transformer for Visual Recognition, i.e., ResT, on a brain tumor segmentation. The work attempts to illustrate that the proposed memory-efficient transformers can achieve comparable performance to existing convolutional neural networks and the original Vision Transformer and have the potential to be a potent alternative in computer-aided medical image-segmentation tasks, particularly in resource-constraint settings.

1 Introduction

Fractionalization of tumors is an important step in detecting and treating brain tumors. Techniques like magnetic resonance imaging (MRI) are broadly utilized in the discovery of brain tumors. In the past few years, brain tumor segmentation

---

A. Marathe (✉) · V. Kadam · A. Chaumal · S. Kodilkar · A. Joshi · S. Sawant
Department of Computer Science & Engineering, COEP Technological University (COEP Tech), Pune, India
e-mail: maratheas19.comp@coep.ac.in

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_9

(BTS) has been a live area in research, and many medical image processing-based approaches have been proposed [1]. These image processing techniques include thresholding, region growing, and level set methods. In order to find the difference between the tumors from the neighboring tissue, it uses several features of the images, like texture and intensity. However, due to the complicated and diverse nature of the tumors present, these techniques are not able to segment them precisely.

In the last few years, due to the progress in computer vision, specifically convolutional neural network (CNN) model, there has been much improvement in the implementation of BTS [2]. These methods use many annotated MRI scans to train a neural network to segment tumors accurately. This trained network can then segment tumors in new MRI scans. The CNN-based models have demonstrated to have segmented the tumors from the MRI scans with remarkable accuracy and outperformed previous approaches. Along with these techniques, other neural network architectures, such as generative adversarial networks (GANs) and recurrence-based networks (RNNs), have also been used for this segmentation task [3]. GANs have been used to generate synthetic MRI scans with tumor annotations, which can be used to train CNNs. RNNs are used to model the temporal progress of brain tumors over time and increase accuracy [4].

Due to the intricate and heterogeneous nature of brain tumors, BTS is a complicated task. However, recent advances in deep learning techniques have shown favorable results for BTS in MRI scans. Further research is needed to develop more accurate and efficient models for BTS. Recently, Vision Transformers have been considered a promising approach for BTS tasks. Vision Transformers are a type of deep learning model inspired by the transformer architecture based on the attention mechanism, initially proposed for natural language processing tasks. Transformers are able to encapsulate dependencies and context in the data in contrast to CNNs, which are based on local convolutional operations [5].

Vision Transformers have been applied to medical and multimodal imaging in BTS task. The training of the models is done on large datasets having MRI scans and can automatically learn features from the data. Vision Transformers can achieve good performance in BTS and outperform traditional image processing methods and CNNs. In multimodal imaging, Vision Transformers have been used to incorporate information from multiple imaging modalities, such as structural MRI, functional MRI, and diffusion-weighted imaging. The models are trained on large datasets of multimodal imaging. They can automatically learn how to combine information from different modalities to refine the performance of BTS. Some studies have shown that using Vision Transformers in multimodal imaging rather than single modality can improve the accuracy of the segmentation.

Vision Transformers are a promising approach for BTS. The ability to process global information of Vision Transformers allows them to capture dependencies that have very long range and understand the context better, which is essential for accurately segmenting brain tumors. Several memory-efficient transformer architectures have been introduced in recent years, and more research is required to measure the performance of the transformers compared to other machine learning-based methods and to explore their potential for other medical imaging tasks.

The rest of this chapter is categorized into different sections. Section 2 concisely discusses related works. The models under comparison, dataset, and performance metrics are mentioned in Sect. 3. Section 4 analyzes the results, and conclusion, as well as future scope, is explained in Sect. 5.

2 Literature Review

Segmentation of tumors is a pivotal step in the identification and treatment of cancer. Due to advancements in technologies like magnetic resonance imaging (MRI) and the good soft tissue contrast of pictures, computer-aided BTS investigations which are noninvasive are gaining interest in recent years. The revolutionary methods using computer-aided techniques for segmenting brain tumors have evolved over 20 years and are now getting closer to being used in ordinary clinical settings. These techniques utilize classification and clustering methods by using various features and considering spatial information in a local neighborhood. The techniques assist in forming a preliminary diagnosis and treatment recommendation for the doctors [1, 6].

Artificial intelligence (AI) methods have demonstrated excellent results in BTS tasks using medical imaging data by applying techniques based on supervised, unsupervised, and deep learning (DL) [7]. Zhao et al. suggested a solution that uses both the gradient and context-sensitive features to recognize and segment brain tumors accurately. They trained a support vector machine (SVM) using the extracted features to classify each voxel into tumor and nontumor regions and evaluated it using a publicly available dataset of brain MRI scans. The model performed better than the previous best in terms of accuracy and sensitivity [8]. Another DL model proposed by Zhao et al. used fully convolutional neural networks (FCNNs) along with conditional random fields (CRFs) to improve accuracy [9]. Gruber et al. proposed another DL-based method for the computerized liver as well as tumor segmentation in abdominal CT images. The method was based on a joint learning approach combining two CNNs. It achieved high accuracy for both tasks, showing higher dice similarity coefficient (DSC) and volumetric overlap error (VOE) [10]. Mlynarski proposed a method combining fully and weakly supervised learning to enhance the model's accuracy and generalization capability. The method employed a fully supervised CNN to train on the part of labeled data and a weakly supervised CNN that learns from a more extensive set of unlabeled data with partial annotations. The weakly supervised CNN used a unique self-supervised learning mechanism that leverages the spatial and contextual information of the MRI images to create the pseudo-labels for the data that is unlabeled. The mixed supervision approach effectively leveraged the limited labeled data and the unlabeled data which is relatively large, improving performance [11]. Kamnitsas et al. and Pereira et al. recommended a 3D CNN and FCN, respectively, for BTS that achieved sotA performance on the BraTS challenge dataset [2, 12]. Peiris et al. suggested a novel approach to segment tumors in 3D medical images using a DL model called

VTranSeg that utilized a volumetric transformer network to learn a nonrigid deformation field between the given image and the expected truth segmentation. This allowed the model to handle anatomical variations better and boost the segmentation score [13].

CNNs are the models of choice in biomedical image segmentation, but there is a growing interest in using transformer models because of their capability to perceive global dependencies and learn the context of the information. Hatamizadeh et al. implemented a new model called Unetformer, a unified Vision Transformer that combined the benefits of both CNNs and transformers. The model used a variant of the U-Net architecture, a famous CNN model architecture used for biomedical image segmentation, and replaced the convolutional layers with transformer layers [14]. Hatamizadeh et al. contended that while traditional U-Net models have been commonly used for this task, their performance can be limited because of their inability to recognize long-range information and learn global context. They proposed a new approach to BTS using a Swin Transformer-based U-Net model. Swin U-NetR is an altered version of the U-Net model that uses Swin Transformer as the backbone network. The Swin Transformer architecture divides the image into patches that do not overlap, which are then processed hierarchically using multiple layers of self-attention and feedforward networks [15]. Vitbis is a modified version of the ViT model that uses a hybrid 2D-1D input format to handle 2D and 3D biomedical images. The model consists of a series of self-attention layers followed by feedforward layers and is trained using a combination of supervised and semi-supervised learning [16]. Recent studies have demonstrated that using MRI scans, ViTs can be effective for BTS. The use of hybrid architectures that combine ViTs with traditional 3D CNNs or U-Nets can further improve the performance of the networks. Using attention mechanisms and spatial-temporal transformers can also help to model the dependencies having long ranges in the images and improve accuracy [17].

3 Proposed Methodology

3.1 Model

We use UNETR, which is a variant of U-Net architecture that incorporates the Transformer model [2]. The UNETR model follows the same basic architecture as U-Net, but it replaces the traditional convolutional layers with Transformer blocks. The Transformer models are a neural network architecture that is primarily used for natural language processing tasks. Recently, they have been successfully used to image processing as well. By using Transformer blocks in UNETR, the model can find long-range dependencies in MRI scans. This helps in the segmentation of medical images where contextual information is critical.

However, the memory consumption-related disadvantages of Vision Transformer (ViT) can make them challenging to use in resource-constrained environments or real-time applications. We instead propose to use memory-efficient transformer architectures for our segmentation. ResT (Memory-Efficient Vision Transformer) is a variant of Vision Transformers that addresses some of the memory consumption-related disadvantages of traditional ViTs [18]. ResT uses a memory-efficient design that reduces the memory utilization and cost involved in computations of the self-attention mechanism, making it more suitable for processing high-resolution images in memory-constrained environments. Efficient self-attention in ResT is given in Eq. 1.

$$\mathrm{EMSA}(Q,K,V) = \mathrm{IN}\left(\mathrm{Softmax}\left(\mathrm{Conv}\left(\frac{QK^T}{\sqrt{d_k}}\right)\right)\right)V \tag{1}$$

The ResT transformer is modified to process a 3D volume of the scans instead of 2D image data. In the proposed model, the input scans are fed to this modified transformer which processes them and passes them to the next transformer. A number of such transformers stacked after one another compose the encoder block. After processing through the encoder, intermediate outputs are passed to the decoder, which consists of CNN blocks. Residual connections are used between the layers of encoder and decoders. Finally, the last layer of the decoder outputs a 3D volume having three channels representing the segmented area.

We propose that by incorporating ResT into the UNETR architecture, it is possible to achieve comparable accuracy in BTS while simultaneously reducing the memory consumption and computational power needed for the self-attention mechanism. The training of the model is carried on a dataset of 3D MRI scans with corresponding segmentation masks. During training, to optimize the model, binary cross-entropy loss, as well as dice coefficient loss, is used. We also employ data augmentation methods like rotations and flips for generalization.

3.2 *Dataset*

The BraTS dataset is an open-source dataset that contains MRI scans of the brain with corresponding segmentation masks of four distinct types of the brain tumors, which include glioma, astrocytoma, oligodendroglioma, and meningioma [19]. The BraTS dataset has four separate types of scans – T1, contrast-enhanced T1, T2 weighted, and FLAIR. These scans are categorized as training and evaluation sets, respectively. The training split comprises of 369 MRI instances and their segmentation masks, while the evaluation set has 125 MRI scans without segmentation masks. The BraTS dataset is widely used in training and evaluating models for BTS tasks. It provides a diverse range of brain tumor types and imaging modalities, which allows researchers to develop and evaluate segmentation algorithms on a large and diverse dataset.

3.3 Training

We use the PyTorch module as well as Medical Open Network for AI (MONAI) for building our model. MONAI library is an openly available framework that provides various tools and capabilities for medical image analysis and machine learning. It is built on top of PyTorch and provides an extensive collection of pre-built components, such as data loaders, transforms, and neural network models, that can be easily customized and combined to create medical image analysis pipelines. The model is run on NVIDIA's V100 GPU for 20 epochs, having a batch size of 4, 0.0001 initial learning rate, and AdamW optimizer.

3.4 Performance Metrics

BTS is important in medical image analysis and involves identifying and delineating the region of a tumor in a brain image. There are several performance metrics that can be used to compute the precision of tumor segmentation algorithms. We apply the following techniques to measure the accuracy of our model.

Dice Similarity Coefficient (DSC): It is a metric that evaluates the actual and predicted segmentation described in Eq. 2:

$$\frac{2x}{2x+y+z} \tag{2}$$

where x is true-positive, y is false-positive, and z is false negative pixels in the image.

Hausdorff Distance: The Hausdorff distance computes the largest distance between the predicted segmentation and the actual segmentation. It is defined as max(hd(GT, P), hd(P, GT)), where hd(GT, P) is the Hausdorff distance between the required segmentation (GT) and predicted segmentation (P) and hd(P, GT) is the Hausdorff distance between P and GT.

These metrics are used to determine the performance of distinct BTS models and analyze their accuracy. It is important to note that no single metric can fully capture the performance of an algorithm, and a combination of metrics should be used for a more comprehensive evaluation.

Table 1 Comparison of UNETR architecture with the previous transformer-based and CNN-based architectures

Model	Dice score	#params
2D U-Net architecture	0.86	16.21M
TransBTS	0.91	15.14M
Swin UNETR	0.92	61.98M
UNETR architecture with ResT	0.82	10.49M

Note: Reported dice scores correspond to the whole tumor

4 Results

Memory-efficient transformer architecture UNETR using ResT transformer is compared with various methods, including TransBTS and Swin UNETR [15, 20]. It is observed that compared to previous state-of-the-art models, the suggested model achieves comparable score with a significant reduction in memory requirement. Table 1 shows that mentioned architecture gains comparable performance to the rest of the architectures while having a significantly smaller number of parameters.

5 Conclusion and Future Work

This work has evaluated the performance and memory utilization of various memory-efficient transformer architectures for brain tumor segmentation task on the BraTS dataset. The results indicate that memory-efficient transformer architectures like ResT can achieve comparable performance to existing convolutional neural networks and the original Vision Transformer though the number of parameters is reduced by 69.2%. These memory-efficient transformer architectures have the potential to be a good alternative in medical image-segmentation tasks in resource-constraint settings. As transformer architectures continue to gain popularity in computer vision tasks, this research provides valuable insights for researchers and practitioners to explore memory-efficient transformer architectures for medical image segmentation.

Acknowledgment We are thankful to the Department of Computer Science & Engineering, COEP Tech, for providing the GPU Server Facility for our work which is established under TEQIP III (a World Bank project).

Declarations The authors do not have conflict of interest for the proposed work. The work has not received funding from any agency. The authors have not executed any experimentation that causes harm to any unit of nature.

References

1. Liu, J., Li, M., Wang, J., Wu, F., Liu, T., & Pan, Y. (2014). A survey of MRI-based brain tumor segmentation methods. *Tsinghua Science and Technology, 19*(6), 578–595.
2. Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging, 35*(5), 1240–1251.
3. Chen, H., Qin, Z., Ding, Y., & Lan, T. (2019, May). Brain tumor segmentation with generative adversarial nets. In *2019 2nd international conference on artificial intelligence and big data* (ICAIBD) (pp. 301–305). IEEE.
4. Thayumanavan, M., & Ramasamy, A. (2022). Recurrent neural network deep learning techniques for brain tumor segmentation and classification of magnetic resonance imaging images. *Journal of Medical Imaging and Health Informatics, 12*(3), 247–254.
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., …, Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv*:2010.11929.
6. Tiwari, A., Srivastava, S., & Pant, M. (2020). Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. *Pattern Recognition Letters, 131*, 244–260.
7. Ranjbarzadeh, R., Caputo, A., Tirkolaee, E. B., Ghoushchi, S. J., & Bendechache, M. (2022). Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools. *Computers in Biology and Medicine, 106405*.
8. Zhao, J., Meng, Z., Wei, L., Sun, C., Zou, Q., & Su, R. (2019). Supervised brain tumor segmentation based on gradient and context-sensitive features. *Frontiers in Neuroscience, 13*, 144.
9. Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Medical Image Analysis, 43*, 98–111.
10. Gruber, N., Antholzer, S., Jaschke, W., Kremser, C., & Haltmeier, M. (2019, July). A joint deep learning approach for automated liver and tumor segmentation. In *2019 13th international conference on sampling theory and applications* (SampTA) (pp. 1–5). IEEE.
11. Mlynarski, P., Delingette, H., Criminisi, A., & Ayache, N. (2019). Deep learning with mixed supervision for brain tumor segmentation. *Journal of Medical Imaging, 6*(3), 034002.
12. Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis, 36*, 61–78.
13. Peiris, H., Hayat, M., Chen, Z., Egan, G., & Harandi, M. (2022, September). A robust volumetric transformer for accurate 3d tumor segmentation. In *Medical image computing and computer assisted intervention–MICCAI 2022: 25th international conference*, Singapore, September 18–22, 2022, Proceedings, Part V (pp. 162–172). Springer Nature.
14. Hatamizadeh, A., Xu, Z., Yang, D., Li, W., Roth, H., & Xu, D. (2022). Unetformer: A unified vision transformer model and pre-training framework for 3d medical image segmentation. *arXiv preprint arXiv*:2204.00631.
15. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2022, July). Swin unetr: Swin transformers for semantic segmentation of brain tumors in MRI images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event*, September 27, 2021, Revised Selected Papers, Part I (pp. 272–284). Springer.
16. Sagar, A. (2021). Vitbis: Vision transformer for biomedical image segmentation. In *Clinical image-based procedures, distributed and collaborative learning, artificial intelligence for combating COVID-19 and secure and privacy-preserving machine learning: 10th Workshop, CLIP 2021, Second Workshop, DCL 2021, First Workshop, LL-COVID19 2021, and First Workshop and Tutorial, PPML 2021*, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27 and October 1, 2021, Proceedings 2 (pp. 34–45). Springer.

17. Parvaiz, A., Khalid, M. A., Zafar, R., Ameer, H., Ali, M., & Fraz, M. M. (2022). Vision transformers in medical computer vision – A contemplative retrospection. *arXiv preprint arXiv*:2203.15269.
18. Zhang, Q., & Yang, Y. B. (2021). Rest: An efficient transformer for visual recognition. *Advances in Neural Information Processing Systems, 34*, 15475–15485.
19. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging, 34*(10), 1993–2024.
20. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., & Li, J. (2021). Transbts: Multimodal brain tumor segmentation using transformer. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24* (pp. 109–119). Springer.

Unlocking New Possibilities in Drug Discovery: A GAN-Based Approach

Vaishnavi Rathod, Jayesh Gadilohar, Shalaka Pawar, Amit Joshi, and Suraj Sawant

Abstract Drug discovery refers to the process of identifying and developing new chemical compounds, to create medications that can treat or cure diseases. In the process of drug discovery, one of the major obstacles is designing a molecule with the necessary properties, as it requires numerous chemical and structural optimizations. Drug discovery is also a time-consuming process, as the typical length of time it takes for a drug to be tested and approved ranges from 10 to 15 years or more to complete all three phases of clinical trials before the licensing stage. In recent years, deep learning has shown great promise in various fields, including drug discovery. Generative adversarial networks can learn representation from chemical structures and the properties of drugs from the dataset to produce new chemical structures that have similar properties to the training set. Generative adversarial networks have the potential to accelerate drug discovery by generating novel compounds with desirable properties, thus reducing the time and cost required for traditional drug discovery methods. This paper explores the use of generative adversarial networks in drug discovery, including the challenges and opportunities associated with their use, and highlights potential future directions for research in this field.

1 Introduction

The process of finding novel drugs to treat or prevent diseases is known as drug discovery. It involves the identification of chemical compounds that have the potential to be developed into drugs, followed by a series of steps to optimize the compounds for safety, efficacy, and pharmacokinetic properties. The procedure for introducing a new drug to the market is a long one. The average time required from

---

V. Rathod (✉) · J. Gadilohar · S. Pawar · A. Joshi · S. Sawant
Department of Computer Engineering and Information Technology, COEP Technological University (COEP Tech), Pune, India
e-mail: rathodvs19.comp@coep.ac.in

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_10

the discovery of a chemical compound to bringing it to the market is about 13 years [1]. The identification of drug targets, which are particular molecules or biological processes that can be altered by a drug to cure a disease, is another crucial aspect of drug development. Recent developments in technology and the accessibility of high-throughput screening techniques have completely changed the way drugs are discovered.

Weliniger et al. [2] coined the term Simplified Molecular Input Line Entry System (SMILES). SMILES is a method of displaying the medicinal molecule's chemical composition. It is a collection of strings and numbers that are used to represent the atoms and bonds of a molecule. The atomic symbol for each atom serves as a representation, while dashes, dots, and other symbols are used to symbolize bonds. Similarity-based methods used this representation as a way to show that structurally similar drugs possess similar properties, operate similarly, and engage in comparable medication interactions [3]. The atom at one end of the molecule marks the start of the SMILES string, and the atom at the other end marks its conclusion. Simple organic molecules as well as more complicated molecules like proteins and DNA can be represented using the SMILES notation. It is a crucial instrument for the storage, retrieval, and data mining of chemical information. It can be used for analyzing the molecular structure of drug candidates. This chemical composition decides the property and usage of a drug molecule.

Drug discovery is a critical activity in the field of medicine as well as healthcare. It entails the identification, discovery, and creation of new medications to treat or cure diseases. The task of designing a drug molecule is complex and involves combining various chemical structures. It is a lengthy and iterative process of identifying potential drug targets, screening natural or synthetic compounds for activity against those targets, and optimizing lead compounds to create safe and effective drugs. Every promising molecule must go through clinical trials to be assessed for its safety and effectiveness in humans after being found. Clinical trials are carried out in stages and evaluate the medication on progressively larger patient populations. If the drug demonstrates safety and efficacy in clinical trials, it can be submitted to regulatory agencies for approval.

In addition, the drug discovery process has ethical implications, particularly in light of the use of animal experimentation and the possible drawbacks and advantages for patients. As a result, attempts are being made to improve the efficacy and moral conduct of drug discovery, for example, by utilizing alternative testing techniques and increasing clinical trial transparency.

For more than three decades, computer-aided drug design (CADD) has been a key factor in the creation of medically significant small molecules [4]. It is a multidisciplinary field that integrates chemistry, biology, and computer science to accelerate the drug discovery process. It helps provide a speedy, accurate, and cost-effective approach to identifying drug candidates. DeepPurpose is a deep learning library implementing an encoder-decoder model that gives exceptional results on well-known datasets and the COVID dataset as well [5]. The identification of numerous new drug targets as a result of improvements in our knowledge of the biochemical causes of diseases has opened up possibilities for the creation of innovative treatments.

Over the years, deep learning models have significantly sped up the pharmaceutical discovery process by cutting down the number of molecules that are experimentally tested and by offering insightful information on how medicines interact with target protein. It is currently employed in screening activities since it is faster than traditional computational chemistry methods and can generate rapid and precise attribute predictions for a variety of relevant qualities [6]. In our approach, we use the generative adversarial networks (GANs) for drug discovery. We look at different approaches for training GANs for drug discovery. We also discuss the various challenges and strategies applied to overcome them. New compounds with desirable characteristics, such as high efficacy against a particular target and minimal toxicity, can be created using GANs. The architecture of GAN includes two neural networks, a generator and a discriminator. The discriminator network undergoes training to discern between actual molecules and molecules that have been made, while the network of generators is trained to create molecules with these characteristics.

Overall, this study summarizes their use and emphasizes how GANs could transform the process of finding new drugs. GANs can assist in accelerating the drug discovery process, cut down on the cost and time associated with drug development, and ultimately lead to the discovery of novel compounds with distinct therapeutic benefits by utilizing the power of artificial intelligence and deep learning.

The remainder of the essay is organized like below. GAN is introduced in Sect. 1.1. Review of the literature is described in Sect. 2. Section 3 contains the proposed methodology for our project. It describes the dataset, encoding, and details on GAN training. Section 4 tackles the results of our model. Section 5 describes the conclusion of the study and its future scope.

1.1 GAN

In the year 2014, Ian Goodfellow and his colleagues proposed a new framework for training generative models using adversarial training [7]. A generator, a discriminator, and an adversarial network make up the three main parts of GANs where the adversarial network makes them compete against each other. In GANs, the two networks are trained iteratively, where the generator *G(z)* is trained to improve its performance and the generation of similar synthetic data that is identical to the actual data, while a learned discriminator network terms the results of the generator as fake. The discriminator *D(x)* can be coined as a classifier that tries to identify the generated compounds as real or fake. The training process of the model begins with initializing the weights of the two networks randomly. The generator network produces fresh data by using random noise as an input. At this stage, the output of the generator is not up to the mark, but with every sample marked as fake by the discriminator, the generator improves and updates its parameters. Now the discriminator network takes both real and generated data as input and outputs a probability that the data is real.

The discriminator network assesses the created data as well as a randomly selected fraction of real data from the training dataset during each iteration. The generator network generates fresh data during each iteration. The generator network changes its weights based on its capacity to produce data that can deceive the discriminator network, whereas the discriminator network updates its weights based on its capacity to correctly discern between actual and produced data. GANs generate huge amounts of data very quickly and effectively. This is a big advantage in drug discovery as the amount of training dataset is limited and expensive. GAN generates new medical compounds with desired properties such as high potency and low toxicity.

However, GAN usage possesses many challenges to be overcome. One limitation is that the generated molecules may not be as efficient as the existing compounds. One way to overcome the difficulties coming with GAN was to use different optimization techniques and hyperparameters.

2 Literature Review

Drug discovery is a challenging procedure that involves the identification and generation of novel drug compounds to treat a variety of diseases. Deep learning has been a powerful tool for finding new drugs in recent years. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs) are a few examples of deep learning algorithms that have been used by many academics to hasten the process of drug discovery.

Chen et al. [8] showed that deep neural networks (DNN) could routinely produce more accurate future predictions on large datasets than RF. A DNN was utilized on a big number of 2D topological descriptors on the Merck Kaggle competition database. The results indicated that the DNN outperformed the standard RF method slightly in 13 out of 15 targets.

SMILES (Simplified Molecular Input Line Entry System), a string-based nomenclature structure, is used to express the chemical makeup of a medicine. It represents the molecular structure of a molecule. Ramesh et al. [9] proposed a system to encode these SMILES using autoencoders and then passing them to a GAN to generate similar compounds. This network produced 39 novel structures, out of which 15 were successfully able to engage at least 1 of the target receptors.

Lukaz et al. [10] propose a Mol-CycleGAN-based approach that produces optimal compounds having a strong structural resemblance to that of the original molecule. They propose to use variational autoencoders that use an encoder-decoder model in which the encoder is responsible to encode the molecules in a high-dimensional vector representation and the decoder is a different neural network that is used for decoding an atom from its encodings in the latent space.

Bjerrum et al. [11] suggested taking as input the SMILES string for LSTM RNNs to model predictions, obviating the necessity of producing molecular descriptors. To increase the dataset size, they utilized the fact that a single molecule can have numerous SMILES, augmenting the data. The augmented dataset was 130 times larger than the original, and the model trained on this dataset displayed superior performance on the test set in comparison with a model generated using only a single canonical SMILES string per molecule.

Blanchard et al. [12] suggested a method of training GANs that encourages gradual exploration and mitigates the effects applying ideas from genetic algorithms to study the impacts of mode collapse. It replaced samples from the training data with valid samples from the generator. It is shown that updates to the training data are demonstrated to significantly outperform the conventional method, expanding the potential uses for GANs in drug discovery.

Weininger et al. [2] trained GAN for generating new molecules with similar properties as the dataset of known drugs. They targeted the HIV protease and SARS-CoV-2 main protease. The paper proposed a framework including integrating an autoencoder, GAN, and LSTM to produce a new set of molecules that were more accurate.

Shao et al. [13] described and equate the prediction of drug-target interactions to a binary classification issue. They proposed a hybrid model that combined two different neural networks: a spectral-based graph convolutional network (GCN) and a convolution neural network (CNN). This approach was able to extract characteristics from drug and target expression profiles. The implicit relationship between a drug and its target is also taken into account in this article.

Bajorath et al. [14] provide a brief outline of the computer-aided drug design (CADD) field in their paper published in 2022. He states that CADD has grown and demonstrated various scientific developments in recent times. Deep learning applied to macromolecular or tiny-molecular design is the main driver of most of these breakthroughs.

In a study put forward by Blaschke T et al. [15] in the year 2017, generative adversarial autoencoders were used to generate novel chemical structures. Blaschke T et al. utilized four different autoencoder architectures for structure generation. The result suggests that the ability to generate new sequences for compounds depends not only on the design of the decoder but also on how the encoder's latent vectors are distributed.

In a comparison study, Putin et al. [16] demonstrated that ORGANIC, the first DNN-based equivalent, is outperformed by RANC, learned on the SMILES description of the chemicals, on a number of important drug discovery metrics, including the number of unique structures, passing Muegge criteria, producing high QED ratings, clearing medicinal chemistry filters (MCFs). Additionally, it suggested a novel deep neural network (DNN) architecture called RANC (Reinforced Adversarial Neural Computer) based on reinforcement learning and generative adversarial networks (GANs) for designing novel small-molecule organic structures (RL).

3 Proposed Methodology

3.1 Data

The dataset used for our paper was obtained from Kaggle [18]. The dataset consists of two columns (SMILES, id). SMILES is a way of representing the chemical composition of drug molecules. The SMILES provides sufficient details on molecular structure utilizing short ASCII strings [2]. The dataset was preprocessed by removing duplicates and filtering out compounds with incomplete information or known toxicity. The data consisted of 104 unique smile strings. The following is an example of a molecule in SMILE representation:

“ClC1 = CC(NC(= O)CSC2 = NC = CC(= N2)C2 = CSC(= N2)C2 = CC = CC = C2) = CC(Cl) = C1”

3.2 Encoding

The GAN model needs an encoded tensor to train on and generate new molecules. To get these encodings, we have written a custom autoencoder that can convert a smile to a given representation. These representations are in the form of tensors of shape as [79,latent_dim_shape]. The molecule’s SMILES structure was changed into a latent vector by the encoder. The latent vector of the produced molecules was changed back to SMILES using the decoder.

3.3 Architecture of Generative Adversarial Network (GAN)

Generative adversarial network (GAN) consists of two neural networks, a generator and a discriminator as shown in Fig. 1. GANs can be successfully applied to different tasks including image synthesis, style transfer [19], and data augmentation [20, 21]. The GAN architecture can be modified to suit a particular task. To make use of GAN for generating new chemical compounds, the following alterations were made:

Generator In a fully connected neural network architecture, the generator is made up of four hidden layers, with the final layer serving as the output layer. Each level of the four levels is a linear layer, and a batch normalization layer follows so that the output is restricted to 0 and 1, and a Leaky Rectified linear activation unit (ReLU) activation function. Leaky ReLU activation procedure is the fifth completely linked layer. The output of the generator has the shape (1, 79) each.

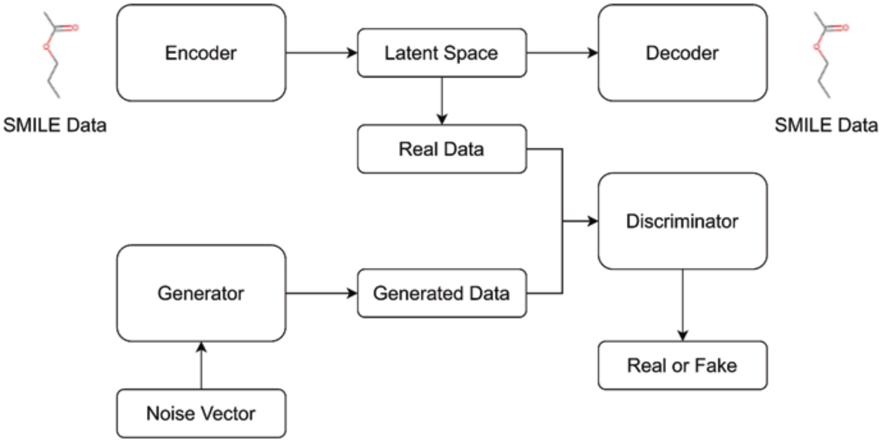


Fig. 1 Architectural diagram for model

Discriminator The discriminator model takes a tensor of size 79 as input and outputs a latent value, i.e., 0 or 1. It consists of eight fully linked layers, all of which have dropout layers between them to prevent overfitting and Leaky ReLU activation functions (except for the final layer). The last layer makes use of a sigmoid activation function, which generates a value between 0 and 1 that denotes the likelihood that the input data is real. To reduce the dimensionality of data as it progresses through the network, the number of neurons was decreased from 256 with every layer to 8.

3.4 Training

To extract the latent vector representations of 104 molecules, their SMILES structures were run through a pretrained encoder. These latent vectors were used to train the GAN. The input data were divided into batches of size 8 and trained for 150 epochs to successfully train the generator and discriminator. Optimization on both the generator and the discriminator is done. Adam optimizer which is a stochastic gradient descent (SGD) variant is used for this purpose. The learning rate was kept at 0.0001. It proves effective in training GANs due to its ability to handle nonstationary objectives and noisy gradients. For calculating the loss after every epoch, Binary Cross Entropy Loss (BCE Loss) function is utilized.

We conducted experiments using various loss functions and regularization techniques that could improve the stability and results of the GAN. The output of the GAN model was newly generated drug molecules. These generated new compounds possessed properties such as high potency and low toxicity. For the evaluation of our model, their activity was assessed against specific targets and potential toxicity. We also compared the properties of the generated compounds to known drugs and other compounds.

4 Results

In this research, we employed a generative adversarial network (GAN) to discover novel drug molecules with potential therapeutic effects. Our GAN model was trained on a dataset of over 104 known drug molecules and their corresponding activities against a specific target protein.

After extensive training, the GAN generated a total of 35 novel drug-like molecules. While generating, the discriminator loss was found to be 0.26. The generator loss was found to be 0.87 as shown in Table 1.

The model proposed using LatentGAN [22] generated some molecules out of which 86% molecules were legitimate and 56% molecules were unique. Out of these generated unique molecules, 97% were novel and they were active against the target to a 71% degree. The RNN model suggested in [23] produced a set of molecules, wherein 96% of the molecules were deemed valid, while 46% of them were distinct. Among the generated unique molecules, 95% were identified as novel, with 65% exhibiting activity against the intended target.

5 Conclusion and Future Work

The paper suggests the use of GAN for drug discovery. As the GAN technology advances, more diverse and high-quality drugs will be generated. This leads to better chemical compounds formed with training. Also, the GAN model could be integrated with other models to achieve better results. An exciting field in drug discovery is the proposal of a personalization approach to drug generation. The generated compounds will be trained based on an individual's personal preferences and genetic information. This could boost patient outcomes and lower the chance of negative effects. This method of using computer-aided approaches reduces the need for animal testing and accelerates the drug discovery procedure. The reduction in time and savings in cost will lead to an efficient and sustainable drug development process. In the years to come, this methodology will benefit patients and society as a whole.

In conclusion, leveraging GANs for drug discovery is a promising strategy that has the potential to transform the industry completely. GANs may quickly produce new medicinal compounds with specified qualities by utilizing machine learning and generative models, which saves time and resources in comparison with conventional drug development techniques. Our paper has given outstanding results by creating compounds with features equal to or better than existing medications. At the same time, it also promotes the usefulness of GANs for drug creation.

Table 1 Results of GAN training

Novel drug molecules generated	35
Generator loss	0.87
Discriminator loss	0.26

However, there are still issues that must be resolved such as the requirement for more diverse and high-quality training data, optimization of GAN hyperparameters, and the validation of produced drugs. Yet, the advantages of using GAN for drug discovery are evident. Our paper proposes that this technology plays a promising role in the creation of fresh medications to treat and prevent diseases.

Acknowledgment We are grateful to the Department of Computer Science & Engineering, COEP Tech, for providing the resources required for our project.

Declarations The authors do not have a conflict of interest for the proposed work. The work has not received funding from any agency. The authors have not executed any experimentation that causes harm to any unit of nature.

References

1. Elton, D. C., Boukouvalas, Z., Fuge, M. D., & Chung, P. W. (2019). Deep learning for molecular design – A review of the state of the art. *Molecular Systems Design & Engineering, 4*(4), 828–849.
2. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences, 28*(1), 31–36.
3. Ma, T., Xiao, C., Zhou, J., & Wang, F. (2018). Drug similarity integration through attentive multi-view graph auto-encoders. In *IJCAI*.
4. Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological Reviews, 66*(1), 334–395.
5. Huang, K., Fu, T., Glass, L. M., Zitnik, M., Xiao, C., & Sun, J. (2020). Deep Purpose: a deep learning library for drug–target interaction prediction. *Bioinformatics*.
6. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. *Machine learning for molecular and materials science*. https://doi.org/10.1038/s41586-018-0337-2
7. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative adversarial networks*. https://doi.org/10.48550/arXiv.1406.2661
8. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today, 23*(6), 1241–1250.
9. Ramesh, A., Rao, A. S., Moudgalya, S., & Srinivas, K. S. (2021, December). GAN-based approach for drug design. In *2021 20th IEEE international conference on machine learning and applications* (ICMLA) (pp. 825–828). IEEE.
10. Maziarka, Ł., Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., & Warchoł, M. (2020). Mol-CycleGAN: A generative model for molecular optimization. *Journal of Cheminformatics, 12*(1), 1–18.
11. Bjerrum, E. J. (2017). SMILES enumeration as data augmentation for neural network modeling of molecules. *arXiv preprint arXiv*:1703.07076.
12. Blanchard, A. E., Stanley, C., & Bhowmik, D. (2021). Using GANs with adaptive training data to search for new molecules. *Journal of Cheminformatics, 13*(1), 1–8.
13. Shao, K., Zhang, Z., He, S., & Bo, X. (2020, November). DTIGCCN: Prediction of drug-target interactions based on GCN and CNN. In *2020 IEEE 32nd international conference on tools with artificial intelligence* (ICTAI) (pp. 337–342). IEEE
14. Bajorath, J. (2022). Deep machine learning for computer-aided drug design. *Frontiers in Drug Discovery, 2*. https://doi.org/10.3389/fddsv.2022.829043

15. Blaschke, T., Olivecrona, M., & Engkvist, O. (2017). Application of generative autoencoder in De Novo molecular design.
16. Putin, E., Asadulaev, A., Ivanenkov, Y., Aladinskiy, V., Sanchez-Lengeling, B., Aspuru-Guzik, A., & Zhavoronkov, A. (2018). A reinforced adversarial neural computer for de novo molecular design. *Journal of Chemical Information and Modeling, 58*(6), 1194–1204.
17. Dataset: https://www.kaggle.com/datasets/divyansh22/drug-discovery-data
18. Liao, Y. M., & Huang, Y. F. (2022). Deep learning-based application of image style transfer. *Mathematical Problems in Engineering, 2022*, Article ID 1693892, 10 pages.
19. Sandfort, V., Yan, K., Pickhardt, P. J., et al. (2019). Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports, 9*, 16884.
20. Motamed, S., Rogalla, P., & Khalvati, F. (2021). Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. *Informatics in Medicine Unlocked, 27*.
21. Prykhodko, O., Johansson, S. V., Kotsias, P.-C., Arus-Pous, J., Bjer-rum, E. J., Engkvist, O., & Chen, H. (2019). A de novo molecular generation method using latent vector based generative adversarial network. *Journal of Cheminformatics, 11*(1), 1–13.
22. Segler, M. H., Kogej, T., Tyrchan, C., & Waller, M. P. (2018). Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science, 4*(1), 120–131.

A Systematic Review on ECG and EMG Biomedical Signal Using Deep-Learning Approaches

Aarti Chugh and Charu Jain

Abstract Biomedical signals play an indispensable role in many medical applications like diagnosis, prognosis, and defining treatment procedures. Recent advancements in artificial intelligence and computation speed have intensified biomedical signal research. This article conducts a systematic and exhaustive overview of the latest research literature on deep-learning methods for the analysis of biomedical signal results, such as electrocardiograms (ECGs) and electromyograms (EMG). ECG and EMG techniques are among some of the most imperative biomedical signals in the diagnosis and activity recognition of the subject. Additionally, the review will explore various well-known databases and discuss numerous contemporary methods with results published between January 2018 and December 2020. We mainly studied the key parameters in the collected paper: deep-learning model and training architecture, medical tasks, dataset sources, and medical application. These are the essential parameters that influence performance. This paper will also discuss and conclude by highlighting critical research gaps and possible future scope to directly build intelligent computational models from biomedical signals.

1 Introduction

Biomedical signals (BS) describe the electrical activity generated by different human body cells like muscular, cardiac, etc. BS evidence in the form of 1D signals are time-domain data, in which sample data points are acquired over time [1]. These signals are constantly changing and reflect the health of the patient's psyche. BS data categories include electromyogram (EMG), data about alterations to skeleton muscle tissue, and electrocardiogram (ECG), data about changes in heartbeat or

---

A. Chugh (✉)
SGT University, Gurugram, India

C. Jain
Xebia Academy, Gurgugram, India

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_11

rhythm. Some signals are systematic and spontaneous activities of the human psyche, such as electroglottography EEG or ECG. In contrast, others, such as the signal of visual evoked potential, are responses to external stimuli. Additionally, BS analysis practices are often shared. The behavior and essential aspects of generating such signals have to be understood, considering the end objective to retrieve the desired information. However, the form of analysis can alter depending on the analyzed signals and the data to be retrieved. Figures 1 and 2 show the raw EMG sample and raw ECG signal. Deep neural network (DNN) is already increasingly used in various disciplines for prediction and classification. In recent time, DNNs are rapidly developing and significantly impacting generalization ability for a vast range of medical tasks.

We assembled papers via PubMed's search tool with keyword combination: "deep-learning," "deep-learning electromyogram," and "deep-learning electrocardiogram." We found 87 recent research work published between January 2018 and December 2020 inclusive from numerous academic journals and publishers. The predominant objective of this survey is to address a wide range of DLM application in BS research. Next section shows literature review followed by Sect. 3 for the most popular deep training models and training architecture employed in the studied articles. Then, in Sect. 4, we provide a summary of deep-learning deployment in specific signal categories and tabulated representation for quick reference. In tables, we only compare the key parameters within studied paper such as medical application, medical task, variants of different DLM, dataset used, and performance of that paper. The Sect. 5 is for discussion and the final section conclude the key gaps and foreseeing research direction.

2 Related Works

Deep-learning approaches in biomedical signaling application evidence between January 2018 and December 2020 provide two forms of scientific research. The first is oriented toward medical fields, such as a taxonomy focusing on medical activities

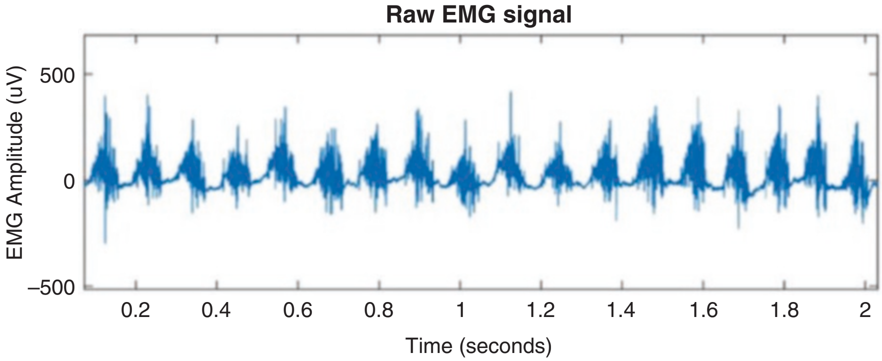


Fig. 1 Raw EMG sample [3]

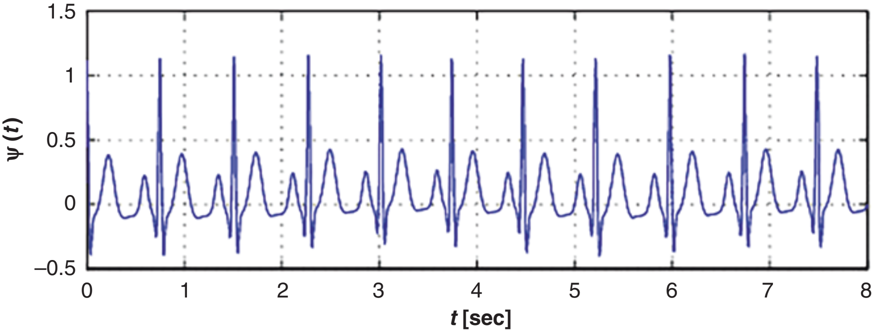


Fig. 2 Raw ECG sample [4]

such as computer-aided detection, illness detection, etc. or a taxonomy focusing on anatomy implementation areas such as the heart, chest, abdomen, eyes, and liver, among others.

The author [5] has accumulated 53 scientific publications on deep-learning strategies for BS processing conducted from 2012 to 2017. This research initially used DLM techniques such as RNN, auto-encoder (AE), DBN, RBM, and GAN. The papers are then categorized according to the BS data modalities. Each segment contains information about the clinical applications, the deep-learning algorithm, the dataset, and the results. The second section discusses deep-learning strategies such as taxonomy centered toward deep-learning architectures such as RNN, AE, DBN, GAN, CNN, and U-Net [2]. A comprehensive systematic study is conducted on one-dimensional BS (base station) data, with a specific emphasis on developing a taxonomy. This study collected 71 articles, mostly ECG publications from 2010 to 2017. The survey's primary objective was to analyze a variety of DLM for the BS study. It then classified DLMs according to their data source, application purpose, input BS class, dataset volume, and neural network training aspect. The author of [6] listed a few biomedical domain factors in deep-learning intervention studies for healthcare complex challenges. It defined the use of DLM in medicine by categorizing it as biological systems, e-health records, medical images, and BS. It concluded by presenting research directions for optimizing health management through the use of BS applications.

3 Deep-Learning Techniques and Training Architecture

3.1 Deep-Learning Techniques

Deep learning concerns research on the extraction of information, predictions, sensible decision, or the recognition of complex patterns employing a collection of data. DNNs are far more adaptable compared to traditional learning approaches since higher accuracy is ideally achieved by increasing the network's size or the training set's scale. For certain modern implementations, shallow teaching strategies such as random forest and support vector machines (SVMs) are insufficient, requiring a large number of generalization experiments and significant manual skilled effort to specify a prior pattern to model [6, 7]. Several DLMs, such as the feed forward network (FFN), convolutional neural network (CNN), recurrent neural network (RNN) and restricted Boltzmann machine (RBM), and vision transformer, have been proposed recently to enhance the accuracy of various learning tasks. The contrastive learning and vision transformer network has recently been applied in learning theory which tremendously improves the DLM [8, 9]. There are also assemble learning technique which combines multiple neural networks [10] and recent approach like transfer learning which transfers the knowledge from one task to another task.

The diversity of training architecture was studied according to the range of data modality. The first type of architecture exploits the traditional MLM as a feature extractor and DLM as a classifier. For instance, mean absolute value divided the raw EMG sample into N levels then, feed into CNN to discriminate sample. The feature extractor and classifier design is shown in Fig. 3.

In contrast, the second architectural type utilizes the DLM as an extractor and the traditional MLM as a classifier. The deep-learning extraction process uses unlabeled data to train the raw sample. The training architecture depicted in Fig. 4 employs deep learning as a feature extractor and conventional machine learning as a classifier. For instance, SAE is used to divide the raw EEG sample into N levels, and then SVM classifies the emotion state based on the featured data.

The third architecture type uses only a DLM to train the raw sample and achieve the final output. The training architecture depicted in Fig. 5 demonstrates the use of only deep-learning approaches to acquire raw data, perform classification, and produce the outcome. For instance, the ECG sample is fed into LSTM to discriminate the patient's cardiac condition.

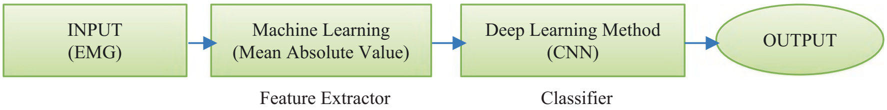


Fig. 3 Training architecture with machine learning acts as feature extractor and deep learning acts as a classifier

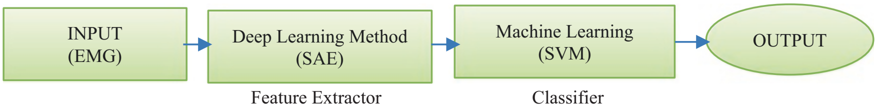


Fig. 4 Deep learning is used to derive features and machine learning is used to classify

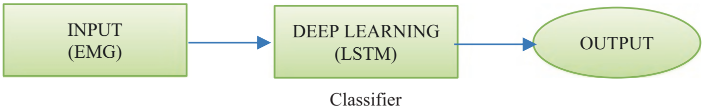


Fig. 5 End-to-end deep-learning training architecture

4 Biomedical Signal Analysis

BS analysis is research that estimates a physical phenomenon for human health. In order to record biomedical parameters, three strategies are practiced: reports (RP), reading (RD), and behavior (BH) [11]. The RP is a manual evaluation of the subject by the specialist. The RD consists of data that are collected by a computer for reading the condition of the patient body, including muscle strength, pulse, etc. The measuring of BH records a range of behavioral patterns like human eye response. This article emphasizes only the RD and BH measurement technique in which the outcomes of the response are given in an ECG, EMG, signal modality.

The EMG signal's muscle tension pattern gives recognition to the movement of the muscles. The rhythm of the heart or pulse version includes a range of cardiovascular diseases, sleep, sentimentality, and gender. This section includes categorization of BS modality in relation to various deep learning models. We compare the DLM quantitatively and qualitatively. The list of DLM used in medical application is depicted for quantitative comparison. Since the performance metric is not given universally for qualitative comparison, we presume an accuracy value as a base criterion for an overall performance comparative analysis.

4.1 Deep Learning with Electrocardiogram (ECG)

As shown in Tables 1, 2, and 3, we identified 56 research papers that used DLM to evaluate ECG signal datasets from the public, private, and hybrid dataset sources. Their important contributions include recognition of cardiac diseases, classification of heartbeat signals, classification of sleep stages, gender-age estimation, and emotion recognition. In a public dataset, the CNN model achieves an overall accuracy of more than 83% for heart disease classification. The LSTM model gives an average accuracy of greater than 90%. The CNN + LSTM model achieves an average accuracy of more than 98%. In the private dataset's heart disease identification, the CNN model achieves an average accuracy of greater than 97%, while the CNN + LSTM model achieves an accuracy of greater than 83%. Therefore, the CNN model outperforms the CNN + LSTM model. As a result, the CNN + LSTM model outperforms the others. For heartbeat signal classification in a publicly available dataset, the CNN model achieves an overall accuracy of greater than 95%. The LSTM model achieves an overall accuracy of greater than 98%, and the hybrid of CNN and LSTM achieves an overall accuracy of greater than 87%. In heartbeat signal classification of the private dataset, only the CNN model is employed and the model performs with an overall

Table 1 Medical task in ECG analysis using public dataset sources

Medical task	No. of subjects	Dataset	DLM	Results
Anomaly class identification [12]	43	MIT-BIH Arrhythmia	LSTM + MLP, LSTM + SVM	LSTM + MLP = 50.0% LSTM + SVM = 42.86%.
Atrial fibrillation detection [13]	23	MIT-BIH Arrhythmia	STFT + CNN	Acc. = 98.29%, Sens. = 98.34%, Spec. = 98.24%
ECG signal detection [14]	47	PhysioNet	LSTM + CNN	Acc. = 99.86%
Heart failure detection [15]	15	BIDMC-CHF	LSTM	Acc. = 99.23%
	18	MIT-BIH NSR		Acc. = 98.84%
	40	Fantasia		Acc. = 98.93%
Dofetilide plasma prediction [16]	42	PhysioNet	CNN	Correlation = 0.85
ECG characteristic detection [17]	23	QT database (ST-T + MIT-BIH Arrhythmia + other)	CNN + RA	QRS-on = −0.70 ± 10.9 QRS-off = −04.80 ± 13.10, T-peak = −0.30 ± 10.50 T-off = −0.30 ± 18.5 P-on = 0.40 ± 14.4, P-peak = −0.40 ± 10.10 P-off = −02.00 ± 12.70
ECG signal compression [18]	48	MIT-BIH Arrhythmia	AE	RMS = 8.00%, comp ratio = 106.50
Electrocardiogram diagnosis [19]	19,000	Chinese Cardiovascular Disease	CNN + RNN	Acc. = 87.69%
Heartbeat classification [20]	N/A	MIT-BIH Arrhythmia	LSTM	F1 = 95.5%, acc. = 99.2% Sens. = 93.0%, spec. = 99.8%
Heartbeat classification [21]	48	MIT-BIH Arrhythmia	CNN	F1 = 90%, acc. = 96%
Heartbeat classification [22]	47	MIT-BIH Arrhythmia	CNN + RBM	AUC = 0.999
Heartbeat classification [23]	48	MIT-BIH Arrhythmia	CNN	Acc. = 98.6%

Medical task	No. of subjects	Dataset	DLM	Results
Multilead ECG classification [24]	48	MIT-BIH Arrhythmia	DLCCA Net	Acc. = 95.3%
	78	INCART Database	TLCCA Net	Acc. = 95.5%
Classification of premature ventricular contraction [25]	119	MIT-BIH Arrhythmia	EBR	Acc. = 100%, prec. and recall = 100%
Ventricular/supraventricular heartbeat identification [26]	44	MIT-BIH Arrhythmia	RBM + DBM	Acc. = 95.57% (supraventricular) Acc. = 93.63% (ventricular)
Arrhythmia classification [33]	47	MIT-BIH Arrhythmia	AE + LSTM	Acc. = 99.0%
Arrhythmia diagnosis [34]	47	MIT-BIH Arrhythmia	CNN + LSTM	Spec. = 98.70%, sens. = 97.50%, acc. = 98.10%
Arrhythmias detection [35]	48	MIT-BIH Arrhythmia	CNN	Acc. = 99.30%
Atrial fibrillation prediction [36]	139	MIT-BIH	CNN	Spec. = 98.7%, sens. = 98.6%, acc. = 98.7%
Beat-wise arrhythmia detection [37]	74	MIT-BIH NSRDB + MIT- BIH AFDB + PAFDB	U-Net + AE	Sens. = 98.7%, spec. = 98.6%, acc. = 98.7%
Cardiac arrhythmia classification [13]	208	PhysioBank	MLP, CNN	Acc. = 88.6%
Heart failure detection [44]	128	Heart Failure Database	CNN	AUC = 84.01%
	208	Kaggle		Acc. = 83.50%
Cardiac arrhythmia classification [38]	45	MIT-BIH Arrhythmia	1D CNN	Acc. = 91.30%
Cardiologist-level arrhythmia classification [39]	53,877	Ziomonitor (iRhythm Technologies)	CNN	F1 = 0.837, AUC = 0.97%
Detection of myocardial ischemia [40]	N/A	PhysioNet	CNN	F1-score = 89.2%, sens. = 84.4%, spec. = 84.9%, AUC = 89.6%
Heart disease classification [41]	47	MIT-BIH	Faster RCNN	Acc. = 99.21%

(continued)

Table 1 (continued)

Medical task	No. of subjects	Dataset	DLM	Results
Heart disease classification [42]	N/A	PhysioNet	LSTM	Acc. = 98.4%
Cardiac arrest detection [43]	35 + 22	Malignant Ventricular Arrhythmia + Creighton University Ventricular Tachyarrhythmia	CNN	Sens. = 97.07%, spec. = 99.44%, acc. = 99.26%
Heart failure detection [44]	128	Heart Failure Database	CNN	AUC = 84.01%
Mental stress recognition [45]	18	Zephyr BioHarness 3	CNN + LSTM	F1 = 0.81%, AUC = 0.92%, acc. = 83.9%,
Apnea detection [46]	35	PhysioNet	CNN	Sens. = 93.0%, spec. = 94.9%, acc. = 94.4%
Sleep position classification with signal quality [47]	12	MIT-BIH Arrhythmia	CNN	Recall and prec. = 0.99%
Sleep apnea detection [48]	70 + 25	University College Dublin + PhysioNet Apnea	CNN	AUC = 0.950, sens. = 83.1%, spec. = 90.3%, acc. = 87.6%
Sleep apnea detection [49]	86	SA Dataset	1D-CNN, GRU	Recall = 99.0%, acc. = 99.0%
	9	KU Leuven University, Belgium		Acc. = 73.95%
Stressful state classification [50]	13	Kwangwoon University, Korea	RNN + CNN	Acc. = 87.38%
Gender-age prediction [50]	275,056	Mayo Clinic Digital Data Vault	CNN	AUC = 0.97%, acc. = 90.4%
Arrhythmia triage in the ED [51]	142,040	University Medical Center Utrecht	Residual CNN	ROC (concordance statistic) = 93.
Atrial fibrillation [52]	75,778	UK Biobank	CNN	F1 = 87.22%, Sens. = 88.55%, Spec.= 85.95%

Medical task	No. of subjects	Dataset	DLM	Results
Atrial fibrillation [53]	11,994	Chinese PLA General Hospital, wearable ECGs, CPSC2018	CNN	Acc. = 99.35 ± 0.26%
Left ventricular hypertrophy [54]	21,286	Sejong General Hospital, Mediplex Sejong Hospital, Korea	CNN	AUC = 0.880
Racial bias [55]	97,829	Mayo Clinic	CNN	AUC = 0.930
Mortality [56]	422,311	Geisinger Hospital System	CNN	AUC = 0.88
Mitral regurgitation [57]	38,241	Sejong General Hospital, Mediplex Sejong Hospital, Korea	CNN + RNN	Area under the ROC = 0.859
Aortic stenosis [58]	39,371	Sejong General Hospital, Mediplex Sejong Hospital, Korea	CNN	Areas under the ROC = 0.884

Table 2 Medical task in ECG analysis using a private dataset sources

Medical task	No. of subjects	Dataset	DLM	Results
ECG anomaly detection [27]	155,8415	Telehealth Network of Minas Gerais	CNN	Spec. = 99%, F1-score = 80%
Veritas detection and cardiologs [28]	1500	ECGs of HCMC	CNN	Acc. = 92.2%, spec. = 94.0%, sens. = 88.7%
Detection of dysfunction of left ventricular systolic [29]	16,056	Mayo Clinic ECG	CNN	Acc. = 86.5%, spec. = 86.8%, sens. = 82.5%
Noise detection with screening model [30]	165, 142, 920	Trauma intensive-care unit	CNN	F1 = 0.80%, AUC = 0.93%, sens. = 0.88%, spec. = 0.89%
ECG classification (scalogram) [31]	290	Physikalisch-Technische Bundesanstalt	ResNet (CNN)	Acc. = 0.730%
	100	Chosun University-ECG		Acc. = 0.940%

Table 3 Medical task in ECG analysis using hybrid dataset sources

Medical task	No. of subjects	Dataset	DLM	Results
Ventricular fibrillation detection [32]	N/A	Creighton Uni. Ventricular Tachyarrhythmia PhysioNet MIT-BIHMVA	1D-CNN + LSTM	Spec. = 98.9%, sens. = 99.7%, BAC = 99.3%
	N/A	OHCA Sujects		Spec. = 96.7%, sens. = 99.2%, BAC = 98.0%

accuracy above 78%. Therefore, the LSTM model outperforms CNN and a hybrid of CNN and LSTM models for heartbeat signal classification. Only the CNN model is used to detect and classify sleep stages in the public dataset, with an accuracy level greater than 87%. The private dataset's sleep-stage classification was performed with 99% accuracy using the CNN-GRU model. We only found gender-age prediction and emotion analysis from the private dataset source. In gender-age prediction, the CNN model performs with an accuracy of more than 90%. CNNs along with RNN models provide an accuracy of more than 73% in classifying emotions. The CNN model performs well in terms of gender-age estimation, with an accuracy of more than 90%.

Deep Learning with Electromyogram

Electromyographic signals (EMG) are biomedical signals that depict the electric charge produced by skeletal muscle fibers. EMG detectors acquire signals from multiple motor units simultaneously, resulting in signal interaction. Since distinct muscle knowledge is characterized by distinct movement, it is capable of discriminating between patterns of action such as an open or closed hand. To classify certain motion patterns based on EMG signal knowledge, 26 research works were performed using DLM, as shown in Table 2. There are two kinds of significant contributions within these study works. One is devoted to hand gesture detection, and the other is devoted to muscle movement recognition in general. In publicly available datasets for hand motion recognition, CNN and CNN + RNN models are the widely used model. The CNN model achieves an average accuracy of more than 68%. However, the CNN + RNN model outperforms the CNN model with a score of more than 82%. In private dataset, the DBN model outperforms other model with an average 88% accuracy. For muscle activity recognition in public dataset, the CNN is most favored. The CNN model has NMSE of 0.034 ± 0.017, while RNN-LSTM model has NMSE of 0.096 ± 0.014. As a result, the CNN is better option than LSTM model in this medical challenge as shown in Table 4.

Table 4 Medical task in EMG analysis using a public dataset, private dataset, and hybrid dataset sources

Medical task	No. of subjects	Dataset	Sources	DLM	Results
Gesture recognition [59]	27	NinaProDB1	Public	CNN + RNN	Acc. = 87.0%
	40	NinaProDB2			Acc. = 82.2%
	17	BioPatRec			Acc. = 94.1%
	18	CapgMyo			Acc. = 99.7%
	5	CSL-HDEMG			Acc. = 94.51%
Gesture recognition [60]	128	NinaPro		CNN	Acc. = 85.78%
	53	BioPatRec			Acc. = 94.0%
Gesture signal classification [61]	17	MYO		CNN	Acc. = 98.31%
	10	NinaPro			Acc. = 68.9%
Hand gesture classification [62]	10	NinaPro		GFM	Acc. = 63.8 ± 05.12%
Hand movement classification [63]	78	Ninapro		CNN + RNN	Acc. = 87.31 ± 04.9%
Sign language recognition [64]	8	6D inertial sensor	Private	DBM	Acc. = 95.1%
Hand-grasping classification [65]	15	MYO		SAE	Acc. = 95%
Hand motion classification [66]	7	MYO		CNN	MCE ± SD = 09.79 ± 04.61
Limb movement estimation [67]	8	NCC Medical Co., LTD		CNN + RNN	Mean_R2 = 90.30 ± 04.50%

(continued)

Table 4 (continued)

Medical task	No. of subjects	Dataset	Sources	DLM	Results
Movement multi-labeled info. extraction [68]	14	ELSCH064NM3		CNN	Mean matchRate = 78.7%.
Muscle activity detection [69]	N/A	Vastus Lateralis		RNN	SNR < 5
Musculoskeletal force prediction [70]	156	TrignoWireless EMG system		CNN	Std. = 0.13, RMSE = 0.25
Prosthetic limb control [71]	2	Grapevine NIP system		CNN	NMSE = 0.032 ± 0.018
Wave form identification [72]	83	Tokushima University Hospital		CNN	Acc. = 86%
Gesture recognition [73]	137 182	NinaPro	Public	CNN	Acc. = 98.15%
Gesture recognition [74]	18+5	CapgMyo (DB-a) + CSL-HDEMG		3D CNN	Acc. = 90.7%
Sleep staging classifier [75]	5213	Sleep Heart Health Study		CNN + LSTM	F1 = 0.87, K = 0.82
Neuromuscular disorder detection [76]	25	EMGLAB database		CNN + kNN	Mean acc. = 100%
Multi-stroke handwriting character [77]	3	MYO		CNN + LSTM	Acc. = 94.85%

5 Discussion

All the performance of the studied architecture is investigated by using seven popular metrics. They are accuracy (acc.), sensitivity (sen.), specificity (spe.), precision (pre.), false positive rate (FPR), F1 score, and Cohen's kappa index.

For the deep-learning task in medical application, we observed that the majority of the contributions were made for the detection, feature retrieval, and data compression tasks. The detection task emphasizes whether or not the instance exists. Arrhythmia diagnosis [35], for example, determines whether a cardiovascular pattern is natural or arrhythmic. The classification activity often concentrates on a grouping or even leveling the instance types. For instance, the classification of emotions analyzes emotion into the depressed, pleasant, normal, or fear categories. The objective of the extraction of features [41] focuses on improving the training dataset to escape an added stress of manually labeled data, using unsupervised learning techniques. Data compression [28] minimizes data size while ensuring high data quality for processing and transfer.

The studied paper has three basic training architecture from using traditional MLM as feature extractor and classifying using DLM to DLM as feature extractor and traditional MLM as the classifier. However, this type of architecture consumes more time and needs extra human labor in selecting the suitable parameter or model. The third architecture takes direct raw input and only needs a DLM to reach the final output. This architecture's goal is to strengthen the algorithm of the DLM without focusing on the input type, which eases the implementation stage. Our survey concludes that each of them has its limitation and benefit, but the third architecture has a much better future scope as dataset size increases by time, and so computation power.

6 Conclusion

This survey conveyed an analysis of deep-learning systems implemented in biomedical 1D signal over the past 3 years. We found 87 papers using the DLM in EMG and ECG biosignal. By examining these works, we present to identify critical parameters used to predict age, gender, sleep analysis, emotion, heart signal analysis, and hand motion.

Additionally, we exhibit that the CNNs outperform the BS at the ultramodern level. We also have determined that there is still no well-defined standardized hyperparameter setting. These nonuniform parameters make it problematic to compare actual performance. The comparison we made should acquaint other researchers to make quick decisions on choosing input data, deep-learning architecture for achieving their desired MA, and obtaining more reliable outcomes. As a lesson learned from these reviews, our discussion can also help fellow researchers make suitable decisions for future work. The parameter we studied has a significant weight on the system performance. In conclusion, a DLM has confirmed promising for bringing those modern contributions to the up to the minute BS analysis for medical treatment.

References

1. Ganapathy, N., Swaminathan, R., & Deserno, T. M. (2018). Deep learning on 1-D biosignals: A taxonomy-based survey. *Yearbook of Medical Informatics, 27*, 98–109.
2. Mu, R., & Zeng, X. (2019). A review of deep learning research. *TIISs, 13*, 1738–1764.
3. Bashar, S. K., Noh, Y. A., Walkey, J., McManus, D. D., & Chon, K. H. (2019). VERB: VFCDM- based electrocardiogram reconstruction and beat detection algorithm. *IEEE Access, 7*, 13856–13866. https://doi.org/10.1109/ACCESS.2019.2894092
4. Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine, 161*, 1–13.
5. Tobore, I., Li, J., Yuhang, L., Al-Handarish, Y., Abhishek, K., Zedong, N., & Lei, W. (2019). Deep learning intervention for health care challenges: Some biomedical domain considerations. *JMIR mHealth and uHealth, 7*, e11966.

6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
7. Loni, M., Sinaei, S., Zoljodi, A., Daneshtalab, M., & Sjödin, M. (2020). Deepmaker: A multi-objective optimization framework for deep neural networks in embedded systems. *Microprocessors and Microsystems, 102989*.
8. Gharehbaghi, A., Dutoit, T., Sepehri, A., Kocharian, A., & Lindén, M. (2015). A novel method for screening children with isolated bicuspid aortic valve. *Cardiovascular Engineering and Technology, 6*(4), 546–556.
9. Hinton, G. E., Sejnowski, T. J., et al. (1986). Learning and relearning in Boltzmann machines. In *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 282–317).
10. Wahengbam, K., Singh, M. P., Nongmeikapam, K., & Singh, A. D. (2021). A group decision optimization analogy-based deep learning architecture for multiclass pathology classification in a voice signal. *IEEE Sensors Journal, 21*(6), 8100–8116.
11. Zhang, D., Yao, L., Chen, K., Wang, S., Chang, X., & Liu, Y. (2019). Making sense of spatio- temporal preserving representations for EEG-based human intention recognition. *IEEE Transactions on Cybernetics*, 1–12.
12. Keyvanrad, M. A., & Homayounpour, M. M. (2015). Deep belief network training improvement using elite samples minimizing free energy. *International Journal of Pattern Recognition and Artificial Intelligence, 29*(5), 1551006.
13. Liu, M., & Kim, Y. (2018). Classification of heart diseases based on ECG signals using long short-term memory. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society* (EMBC), pp. 2707–2710.
14. Wang, L., & Zhou, X. (2019). Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals. *Sensors, 19*, 1502.
15. Attia, Z. I., Sugrue, A., Asirvatham, S. J., Acherman, M. J., Kapa, S., Freidman, P. A., & Noseworthy, P. A. (2018). Noninvasive assessment of dofetilide plasma concentration using a deep learning (neural network) analysis of the surface electrocardiogram: A proof of concept study. *PLoS ONE, 13*.
16. Chen, M., Wang, G., Xie, P., Sang, Z., Lv, T., Zhang, P., & Yang, H. (2018). Region aggregation network: Improving convolutional neural network for ECG characteristic detection. In *Proceedings of the 40th annual international conference of the IEEE engineering in medicine and biology society* (EMBC), pp. 18–21.
17. Wang, F., Ma, Q., Liu, W., Chang, S., Wang, H., He, J., & Huang, Q. (2019). A novel ECG signal compression method using spindle convolutional auto-encoder. *Computer Methods and Programs in Biomedicine, 175*, 139–150.
18. Xu, S. S., Mak, M. W., & Cheung, C. C. (2018). Towards end-to-end ECG classification with raw signal extraction and deep neural networks. *IEEE Journal of Biomedical and Health Informatics, 23*, 1574–1584.
19. Saadatnejad, S., Oveisi, M., & Hashemi, M. (2019). LSTM-based ECG classification for continuous monitoring on personal wearable devices. *IEEE Journal of Biomedical and Health Informatics, 24*, 515–523. [CrossRef] [PubMed].
20. Brito, C., Machado, A., & Sousa, A. (2019). Electrocardiogram beat-classification based on a ResNet network. *Studies in Health Technology and Informatics, 264*, 55–59.
21. Yildirim, Ö. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Computers in Biology and Medicine, 96*, 189–202.
22. Zhao, W., Hu, J., Jia, D., Wang, H., et al. (2019). Deep learning based patient- Specific classification of arrhythmia on ECG signal. In *Proceedings of the 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE.
23. Mathews, S. M., Kambhamettu, C., & Barner, K. E. (2018). A novel application of deep learning for single-lead ECG classification. *Computers in Biology and Medicine, 99*, 53–62.
24. Ribeiro, A. L. P., Paixao, G. M. M., Gomes, P. R., Ribeiro, M. H., et al. (2019). Tele- electrocardiography and bigdata: The CODE (Clinical Outcomes in Digital Electrocardiography) study. *Journal of Electrocardiology, 57*, S75–S78.

25. Smith S.W, Walsh B, Grauer K, Wang K, Rapin, Li J, Fennell W & Taboulet P. (2019). A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. Journal of Electrocardiology, 52, (pp. 88– 95).
26. Attia, Z. I., Kapa, S., Yao, X., Lopez-Jimenez, F., Mohan, T. L., Pellikka, P. A., et al. (2019). Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *Journal of Cardiovascular Electrophysiology, 30*, 668–674.
27. Yoon, D., Lim, H., Jung, K., Kim, T., & Lee, S. (2019). Deep learning-based electrocardiogram signal noise detection and screening model. *Healthcare Informatics Research, 25*, 201–211.
28. Picon, A., Irusta, U., Alvarez-Gila, A., Aramendi, E., Alonso-Atienz, F., et al. (2019). Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia. *PLoS ONE*.
29. Yildirim, O., Baloglu, U. B., Tan, R., Ciaccio, E. J., & Acharya, U. R. (2019). A new approach for arrhythmia classification using deep coded features and LSTM networks. *Computer Methods and Programs in Biomedicine, 176*, 121–133.
30. Oh, S. L., Ng, E. Y. K., Tan, R. S., & Acharya, U. R. (2018). Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Computers in Biology and Medicine, 102*, 278–287.
31. He, Z., Zhang, X., Cao, Y., Liu, Z., Zhang, B., & Wang, X. (2018). LiteNet: Lightweight neural network for detecting arrhythmias at resource-constrained mobile devices. *Sensors, 18*, 1229.
32. Erdenebayar, U., Kim, H., Park, J., Kang, D., & Lee, K. (2019). Automatic prediction of atrial fibrillation based on convolutional neural network using a short-term normal electrocardiogram signal. *Journal of Korean Medical Science*, 34.
33. Oh, S. L., Ng, E. Y. K., Tan, R. S., & Acharya, U. R. (2019). Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types. *Computers in Biology and Medicine, 105*, 92–101.
34. Yıldırım, Ö., Plawiak, P., Tan, R. S., & Acharya, U. R. (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Computers in Biology and Medicine, 102*, 411–420.
35. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine, 25*, 65–69.
36. Yildirim, O., Talo, M., Ay, B., Baloglu, U. B., Aydin, G., & Acharya, U. R. (2019). Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals. *Computers in Biology and Medicine, 113*, 103387.
37. Ji, Y., Zhang, S., & Xiao, W. (2019). Electrocardiogram classification based on faster regions with convolutional neural network. *Sensors, 19*, 2558.
38. Sbrollini, A., Jongh, M. C., Haar, C., Man, S., Burattini, L., et al. (2019). Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: A deep-learning approach. *Biomedical Engineering Online, 18*, 15.
39. Seo, W., Kim, N., Lee, C., & Park, S. (2019). Deep ECG-respiration network (DeepER Net) for recognizing mental stress. *Sensors, 19*, 3021.
40. Wang, L., Lin, Y., & Wang, J. (2019). A RR interval based automated apnea detection approach using residual network. *Computer Methods and Programs in Biomedicine, 176*, 93–104.
41. Kido, K., Tamura, T., Ono, N., Altaf-Ul-Amin, M., Sekine, M., et al. (2019). Novel CNN-based framework for classification of signal quality and sleep position from a capacitive ECG measurement. *Sensors, 19*, 1731.
42. Wang, T., Lu, C., Shen, G., & Hong, F. (2019). Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network. *PeerJ, 7*.
43. Erdenebayar, U., Kim, Y. J., Park, J., Joo, E., & Lee, K. (2019). Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. *Computer Methods and Programs in Biomedicine, 180*, 105001.
44. Hwang, B., You, J., Vaessen, T., Myin-Germeys, I., Park, C., & Zhang, B. T. (2018). Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals. *Telemedicine e-Health, 24*, 753–772.

45. Attia, Z. I., Friedman, P. A., Noseworthy, P. A., Ladewig, D. J., Satam, G., Pellikka, P. A., et al. (2019). Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circulation: Arrhythmia and Electrophysiology*, 12.
46. Ivanovic, M. D., Vladimir, A., Alexei, S., Ljupco, H., & Maluckov, A. (2019). Deep learning approach for highly specific atrial fibrillation and flutter detection based on RR intervals. In *Conference proceedings: Annual international conference of the IEEE engineering in medicine and biology society*, pp. 1780–1783.
47. Van de Leur, R., Blom, L. J., Gavves, E., Hof, I. E., Van der Heijden, J. F., & Clappers, N. C. (2020). Automatic triage of 12-lead ECGs using deep convolutional neural networks. *Journal of the American Heart Association, 9*, e015138.
48. Oster, J., Hopewell, J. C., Ziberna, K., Wijesurendra, R., Camm, C. F., Casadei, B., et al. (2020). Identification of patients with atrial fibrillation: A big data exploratory analysis of the UK Biobank. *Physiological Measurement, 41*, 025001. https://doi.org/10.1088/1361-6579/ab6f9a
49. Wang, F., Kaushal, R., & Khullar, D. (2020). Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Annals of Internal Medicine, 172*, 59–60.
50. Chen, T.-M., Huang, C.-H., Shih, E. S., Hu, Y.-F., & Hwang, M.-J. (2020). Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model. *iScience, 23*, 100886. https://doi.org/10.1016/j.isci.2020.100886
51. Cai, W., Chen, Y., Guo, J., Han, B., Shi, Y., Ji, L., et al. (2020). Accurate detection of atrial fibrillation from 12-lead ECG using deep neural network. *Computers in Biology and Medicine, 116*, 103378.
52. Kwon, J.-M., Jeon, K.-H., Kim, H. M., Kim, M. J., Lim, S. M., Kim, K.-H., et al. (2020). Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *Europace, 22*, 412–419.
53. Lin, C.-S., Lin, C., Fang, W.-H., Hsu, C.-J., Chen, S.-J., Huang, K.-H., et al. (2020). A deep-learning algorithm (ECG12Net) for detecting hypokalemia and hyperkalemia by electrocardiography: algorithm development. *JMIR Medical Informatics, 8*, e15931.
54. Noseworthy, P. A., Attia, Z. I., Brewer, L. C., Hayes, S. N., Yao, X., Kapa, S., et al. (2020). Assessing and mitigating bias in medical artificial intelligence: The effects of race and ethnicity on a deep learning model for ECG analysis. *Circulation: Arrhythmia and Electrophysiology, 13*, e007988.
55. Kwon, J.-M., Lee, S. Y., Jeon, K.-H., Lee, Y., Kim, K.-H., Park, J., et al. (2020). Deep learning-based algorithm for detecting aortic stenosis using electrocardiography. *Journal of the American Heart Association, 9*, e014717.
56. Han, X., Hu, Y., Foschini, L., Chinitz, L., Jankelson, L., & Ranganath, R. (2020). Deep learning models for electrocardiograms are susceptible to adversarial attack. *Nature Medicine, 26*, 360–363.
57. Hu, Y., Wong, Y., Wei, W., Du, Y., Kankanhalli, M., & Geng, W. (2018). A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. *PLoS ONE, 13*, e0206049.
58. Wei, W., Dai, Q., Wong, Y., Hu, Y., Kankanhalli, M., & Geng, W. (2019). Surface electromyography-based gesture recognition by multi-view deep learning. *IEEE Transactions on Biomedical Engineering, 66*, 2964–2973.
59. Sun, W., Liu, H., Tang, R., Lang, Y., He, J., & Huang, Q. (2019). sEMG-based hand-gesture classification using a generative flow model. *Sensors, 19*, 1952.
60. Wang, W., Chen, B., Xia, P., Hu, J., & Peng, Y. (2018). Sensor fusion for myoelectric control based on deep learning with recurrent convolutional neural networks. *Artificial Organs, 42*, E272–E282.
61. Yu, Y., Chen, X., Cao, S., Zhang, X., & Chen, X. (2019). Exploration of Chinese sign language recognition using wearable sensors based on deep belief net. *IEEE Journal of Biomedical and Health Informatics, 24*, 1310.
62. Li, C., Ren, J., Huang, H., Wang, B., Zhu, Y., & Hu, H. (2018). PCA and deep learning based myoelectric grasping control of a prosthetic hand. *Biomedical Engineering Online, 17*, 107.

63. Rehman, M. Z., Waris, A., Gilani, S. O., Jochumsen, M., Niazi, I. K., Jamil, M., et al. (2018). Multiday EMG-based classification of hand motions with deep learning techniques. *Sensors, 2018*(18), 2497.
64. Xia, P., Hu, J., & Peng, Y. (2018). EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks. *Artificial Organs, 42*, E67–E77.
65. Olsson, A. E., Sager, P., Andersson, E., Bjorkman, A., Malesevic, N., & Antfolk, C. (2019). Extraction of multi-labelled movement information from the raw HD-sEMG image with time-domain depth. *Scientific Reports, 9*, 7244.
66. Khowailed, I. A., & Abotabl, A. (2019). Neural muscle activation detection: A deep learning approach using surface electromyography. *Journal of Biomechanics, 95*, 109322.
67. Dantas, H., Warren, D. J., Wendelken, S. M., Davis, T. S., Clark, G. A., & Mathews, V. J. (2020). Deep learning movement intent decoders trained with dataset aggregation for prosthetic limb control. *IEEE Transactions on Biomedical Engineering, 66*, 3192–3203.
68. Nodera, H., Osaki, Y., Yamazaki, H., Mori, A., Izumi, Y., & Kaji, R. (2019). Deep learning for waveform identification of resting needle electromyography signals. *Clinical Neurophysiology, 130*, 617–623.
69. Tam, S., Boukadoum, M., Campeau-Lecours, A., & Gosselin, B. (2020). A fully embedded adaptive real-time hand gesture classifier leveraging HD-sEMG and deep learning. *IEEE Transactions on Biomedical Circuits and Systems, 14*(2), 232–243. https://doi.org/10.1109/TBCAS.2019.2955641
70. Chen, J., Bi, S., Zhang, G., & Cao, G. (2020). High-density surface EMG-based gesture recognition using a 3D convolutional neural network. *Sensors (Basel), 20*(4), 1201. https://doi.org/10.3390/s20041201
71. Zhang, L., Fabbri, D., Upender, R., & Kent, D. (2019). Automated sleep stage scoring of the Sleep Heart Health Study using deep neural networks. *Sleep, 42*(11), zsz159. https://doi.org/10.1093/sleep/zsz159
72. Chakraborty, S., Aich, S., Joo, M. I., Sain, M., & Kim, H. C. (2019). A multichannel convolutional neural network architecture for the detection of the state of mind using physiological signals from wearable devices. *Journal of Healthcare Engineering, 5397814*. https://doi.org/10.1155/2019/5397814
73. Jabbari, M., Khushaba, R. N., & Nazarpour, K. (2020). EMG-based hand gesture classification with long short-term memory deep recurrent neural networks. In *Annual international conference on IEEE engineering in medicine and biology society* (pp. 3302–3305). https://doi.org/10.1109/EMBC44109.2020.9175279
74. Triwiyanto, T., Pawana, I. P. A., & Purnomo, M. H. (2020). An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28*(7), 1678–1688.
75. Jarchi, D., Andreu-Perez, J., Kiani, M., Vysata, O., Kuchynka, J., Prochazka, A., & Sanei, S. (2020). Recognition of patient groups with sleep related disorders using bio-signal processing and deep learning. *Sensors (Basel), 20*(9), 2594. https://doi.org/10.3390/s20092594
76. Samanta, K., Roy, S. S., Modak, S., Chatterjee, S., & Bose, R. (2020). Neuromuscular disease detection employing deep feature extraction from cross spectrum images of electromyography signals. In *Annual international conference on IEEE engineering in medicine and biology society* (pp. 694–697).
77. Zheng, W.L.; Liu, W.; Lu, Y.; Lu, B.L.; Cichocki, A. Emotionmeter: A multimodal framework for recognizing human emotions. IEEE Trans. Cybern. 2018, 49, 1110–1122.

Smart AI Bot for Healthcare Assistance

Yamna Ghoul and Naoufel Zitouni

Abstract The new coronavirus has become a worldwide epidemic. In fact, this is a huge challenge for medical staff and managing the flow of large numbers of affected cases. Therefore, when an epidemic occurs directly, it becomes more difficult to consult a medical expert. Indeed, during this pandemic, progress has been made in the use of health chatbots to combat the spread of the virus. These applications provide a form of instant dialog between humans and artificial intelligence (AI) programs by using voice- and/or text-based commands in natural human language. This research work proposes an intelligent chatbot for the diagnostic of nCOV-19. The chatbot developed is composed of three interdependent modules: an information module (IM), a data module (DM) which collects user information, and an action module (AM) which yields answers to the chatbot.

1 Introduction

The world is encountering a new health crisis caused by the coronavirus (COVID-19). So as to trace this epidemic, digital services are carried out for evaluation and triage to lessen pressure at the emergency responders. A chatbot is a particular AI technology that mimics humanlike conversation through speech or text chats on a messaging platform. This immediate messaging system has been extensively followed as one of the important technologies to speak and alternate information [1, 2].

Indeed, the majority of those platforms establish integrated support for the use of chatbots, which are machine-controlled informal agents able to interact with

---

Y. Ghoul (✉)
National Engineering School of Tunis, LR11ES20 Analysis, Conception and Control of Systems Laboratory, Tunis El Manar University, Tunis, Tunisia

N. Zitouni
Faculty of Sciences of Tunis, Application Laboratory of Energy Efficiency and Renewable Energies (UR-LAPER), Tunis El Manar University, Tunis, Tunisia

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_12

customers of the platform [3–5]. In this regard, chatbots are doubtlessly known as the maximum promising and superior shape of human-device interactions [6–9].

These software programs communicate by voice or text [10, 11]. Indeed, the developed chatbot is hooked up on smartphones [12]. There, a plethora of developed chatbot has been recognized. We will cite the most important ones, beginning with “Alexa” from Amazon. Another chatbot “Siri” is developed by Apple. A third significant chatbot “Cortana” has been created by Microsoft.

Currently, chatbots are performed for many related purposes as well as physical health, intimate accomplice violence, suicide, and different conversations [13–28].

For healthcare sector, these AI applications are becoming widespread. Indeed, chatbots help hospital administrators make appointments and answer habitual patient questions. These robots serve as coronavirus medical consultants.

In this research work, a smart digital chatbot, referred to as COVID-19 Assistant, for COVID-19 consultancy throughout and after quarantine is developed. This bot ensures live interactions with doctors and health specialists within the handling of difficult situations.

In the second part, we present the framework of the developed scheme. The architecture of the developed chatbot is introduced in the third part. In Part 4, we present some responses examples of the proposed bot. A discussion is then addressed in Part 5. In the last part, the major results are summarized.

2 Framework and Methods

The developed chatbot uses the artificial intelligence markup language (AIML) which is used to treat users demands. Indeed, with regard to user responses, the AIML is considered to evaluate the existing user’s medical conditions [29].

The designed chatbot is based on two ways:

(a) Request analysis
(b) Return response

Firstly, the developed chatbot uses a predefined questionnaire. Secondly, it collects data in order to use it in machine learning. Thirdly, responses are given by the COVID-19 Assistant (Fig. 1).

Knowledge Base
The knowledge base is performed using the user’s message response. The first step is to pick out the initial signs and symptoms of nCOV-1. In the second step, the bot engine begins to ask the person about symptomatic information or makes contact with health specialists.

Data Base
The developed chatbot is based on a backend logic for outcome generation. Indeed, it uses a questionnaire based on simple yes/no type questions as shown in Table 1.

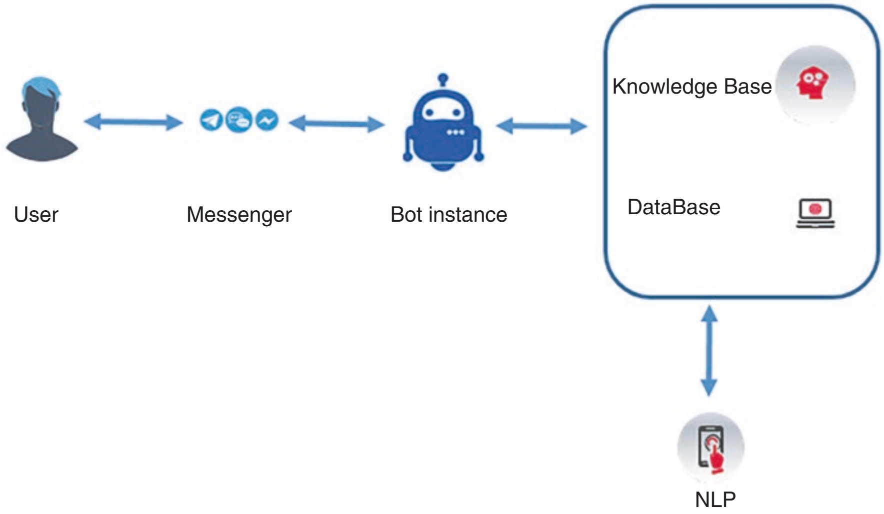


Fig. 1 The framework of COVID-19 Assistant

Table 1 Symptomatic COVID-19 inquiry dataset

No.	Question	Response's type
1	In the last few days, have you had a cough and difficulty breathing?	Yes/No
2	Is your tour records related to inflamed international locations within the remaining months?	Yes/No
3	Do you have interaction with other infected people?	Yes/No
4	Are you going through a surprising upward push in body temperature?	Yes/No
5	Do you cough a lot these days?	Yes/No

By answering with "yes," the bot gives the rating of 1. Contrariwise, the bot offers a rating of 0.

The steps of the developed chatbot are summarized in Fig. 2.

3 COVID-19 Assistant Architecture

The developed chatbot consists of three modules as shown in Fig. 3.

(a) Information module (IM)
(b) Data module (DM)
(c) Action module (AM)

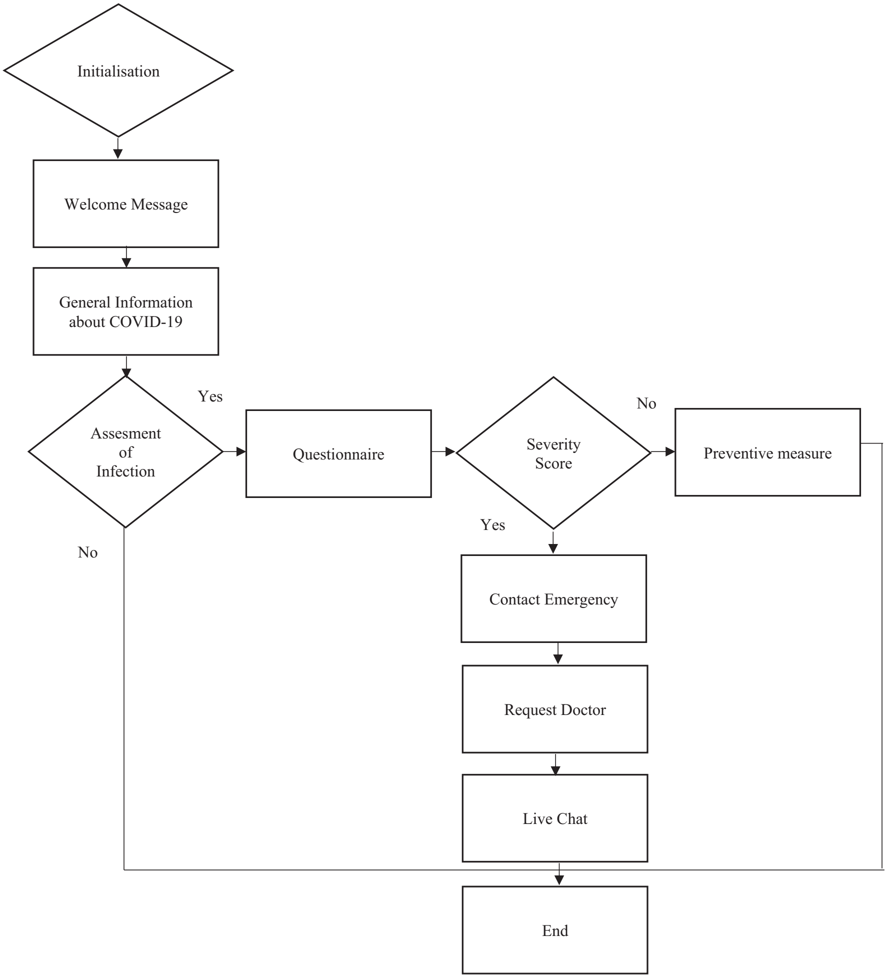


Fig. 2 Working functionality of the developed chatbot

Information Module (IM)

After identifying user's messages, the COVID-19 Assistant transforms unstructured text to a structured language.

Data Module (DM)

The data module's challenge is to acquire the user's nonexclusive facts and then create a dataset that incorporates those user's information: location, age, gender, symptoms (fever, coughs, etc.).

Action Generator Module (AM)

The developed bot returns an accurate and rapid action. The step is performed using the action module (AM).

The whole architecture of COVID-19 Assistant is depicted in Fig. 3.

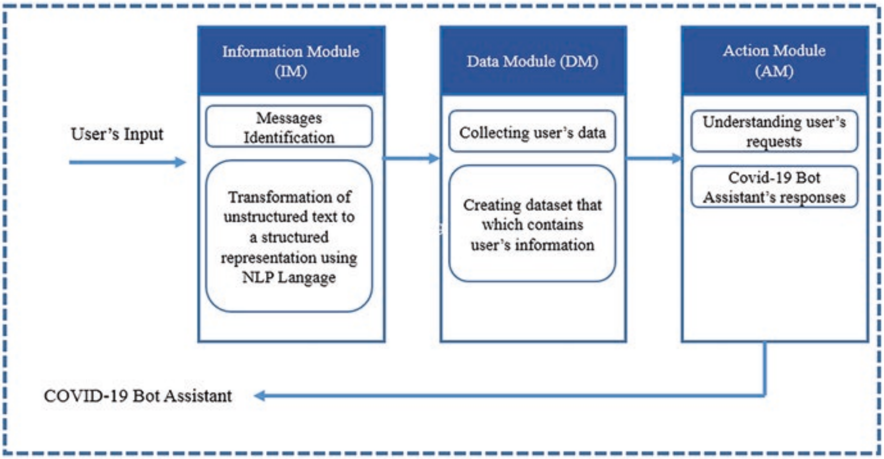


Fig. 3 Architecture of COVID-19 Assistant

4 Results

In this section, we present the developed COVID-19 Assistant.

4.1 *COVID-19 Assistant Responses*

Figure 4 presents a part of the proposed chatbot.

In summary, the presented bot seeks user's information in order to understand the situation and to collect data which is used for a further machine learning.

As soon as data are collected, bot's responses are performed.

The advanced bot named COVID-19 Assistant offers an extensive variety of information:

- COVID-19 symptoms.
- Test records.
- Local and countrywide records approximately the disease.
- Availability of offerings and the way to get entry to them.
- Unemployment benefits.
- Volunteering opportunities.
- Live chat with doctors and health specialists.

4.2 *Discussion*

Developed applications using artificial intelligence are based on building smart systems [30, 31].

Coronavirus (COVID-19) is an illness caused by a virus that can spread from person to person.
• The virus that causes COVID-19 is a new coronavirus that has spread throughout the world.
• COVID-19 symptoms can range from mild (or no symptoms) to severe illness.

• There is currently no vaccine to protect against COVID-19. The best way to protect yourself is to avoid being exposed to the virus that causes COVID-19.
• Stay home as much as possible and avoid close contact with others.
• Wear a mask that covers your nose and mouth in public settings.
• Clean and disinfect frequently touched surfaces.

• Stay home if you are sick, except to get medical care.
• Avoid public transportation, ride-sharing, or taxis.
• Separate yourself from other people and pets in your home.
• There is no specific treatment for COVID-19, but you can seek medical care to help relieve your symptoms.
• If you need medical attention, call ahead.

• Everyone is at risk of getting COVID-19.
• Older adults and people of any age who have serious underlying medical conditions may be at higher risk for more severe illness

Did Have you had a fever in the past few days? (Temperature> 37.8 ° C)

For more information, please contact

If you feel bad, call emergency:

Fig. 4 COVID-19 Assistant's responses examples

These main problems treated by these 6 applications include planning, diagnosis, automatic reasoning and automatic classification, robotics, and learning.

The developed chatbot is an AI-based strategy which uses a simple statistical questionnaire test. We can summarize the main features of the developed chatbot in the following:

- The developed chatbot is a digital application which provides a dynamic response system.
- It is an excellent informative, up-to-date, and well-organized tool.
- The chatbot information sources were approved by experts.
- The reliability is high.

- The developed chatbot protects user's privacy and security.
- It gives the possibility to speak with a doctor or a health specialist 24/7 with a live chat feature.

5 Conclusion

In this research work, a smart digital bot for COVID-19 Assistant has been proposed. The developed bot is based on three modules: information module (IM), data module (DM), and action module (AM). Moreover, it includes 24/7 accessibility.

The presented chatbot can be used for the following:

- Collecting user's data
- Providing reassuring information to take the necessary precaution actions
- Informing people for the precautions to take to protect themselves
- Ensuring live interactions with medical specialists in the case of difficult situations

Future work is performed to improve our developed chatbot with much larger datasets. In addition, we aim to enhance the bot's development by translating the contents to other languages. Moreover, the proposed scheme will be extended to the assistance to other epidemics or other services.

References

1. Nardi, B., Whittaker, S., & Bradner, E. (2000). Interaction and outeraction: Instant messaging in action. In *Proceedings of the 3rd SCW conference*, pp. 79–88.
2. Grinter, R., & Palen, L. (2002). Instant messaging in teen life. In *Proceedings of the 5th CSCW conference*, pp. 21–30.
3. Klopfenstein L, Delpriori S, Malatini S, Bogliolo A (2017). The rise of bots: A survey of conversational interfaces, patterns, and paradigms. Proc. Conf. Designing Interact. Syst. (DIS), p 555–565.
4. Catrin, S., Alsafi, Z., Niamh, O., Mehdi, K., Ahmed, K., Ahmed, A., Christos, I., & Riaz, A. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery, 76*, 71–76.
5. Goyal, K., Chauhan, P., Chhikara, K., Gupta, P., & Singh, P. (2020). Mini: Fear of COVID 2019: First suicidal case in India. *Asian Journal of Psychiatry, 49*.
6. Shao, C., Ciampaglia, G., Varol, O., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications, 9*(4787).
7. Eleonora, P., & Gabriele, P. (2020). Forecasting artificial intelligence on online customer assistance: Evidence from chatbot patents analysis. *Journal of Retailing and Consumer Services, 55*.
8. Ophir, G., Maayan, F., Hayit, G., & Patrick, D. (2020). Browning and Adam Bernheim and Eliot Siegel: Rapid AI development cycle for the coronavirus (COVID-19) Pandemic: initial results for automated detection and patient monitoring using deep learning CT image analysis.
9. Peng, M.,Yang, J., Shi, Q., Ying, L., Zhu, H., Zhu, G., & Ding. (2020). Artificial intelligence application in COVID-19 diagnosis and prediction.

10. Lucas, G., Gratch, J., King, A., & Morency, L. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94–100.
11. Miner, A. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine, 176*, 619–625.
12. Steinhubl, S., & Topol, E. (2018). Now we're talking: Bringing a voice to digital medicine. *Lancet*, 392–627.
13. Laranjo, L. (2018). Conversational agents in healthcare: A systematic review. *International Journal of the American Medical Informatics Association, 25*, 1248–1258.
14. Nobles, A. (2020). Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. *npj Digital Medicine, 3*(11).
15. Kocaballi, A. (2020). Responses of conversational agents to health and lifestyle prompts: Investigation of appropriateness and presentation structures. *Journal of Medical Internet Research, 22*, e15823.
16. Bickmore, T. (2018). Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and Google assistant. *Journal of Medical Internet Research, 20*, e11510.
17. WHO. (2020). *Director-general's opening remarks at the media briefing on COVID-19*. WHO. https://www.who.int/dg/speeches/detail/who-director-general-s-openingremarks-at-the-media-briefing-on-covid-19
18. Davis, C., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web*, pp. 273–274.
19. Ferrara, E. (2020). COVID-19 on Twitter: Bots. Conspiracies and social media activism.
20. Caldarelli, G., De Nicola, R., Del Vigna, F., Petrocchi, M., & Saracco, F. (2019). The role of bot squads in the political propaganda on twitter. *arXiv preprint arXiv*:1905.12687.
21. Yan, R. (2018). *Chitty-chitty-chat bot: Deep learning for conversational AI*. In Proceedings of the twenty-seventh international joint conference on artificial intelligence (IJCAI-18), Stockholm, Sweden, pp. 13–19.
22. Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science, 38*, 937.
23. Chung, K., & Park, R. C. (2019). Chatbot-based heathcare service with a knowledge base for cloud computing. *Cluster Computing, 22*, 1925–1937.
24. Dharwadkar, R., & Deshpande, N. A. (2018). A medical ChatBot. *International Journal of Computer Trends and Technology, 60*, 41–45.
25. Park, J. (2018). A development of ChatBot for emotional stress recognition and management using NLP. *67*. https://doi.org/10.5370/KIEE.2018.67.7.954
26. Chih-Wei, L., Yau-Shian, W., Tsung-Yuan, H., Kuan-Yu, C., Hung-Yi, L., & Lin-shan, L. (2018). Scalable sentiment for sequence-to-sequence Chatbot response with performance analysis. 1804.02504.
27. Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C., & Mctear, M. (2017). Towards a Chatbot for digital counselling, *24*, 1–7.
28. Flora, A., Stefano, M., Vincenzo, M., Gabriele, P., Antonio, P., & Carlo, S. (2017). Chatbots meet eHealth: Automatizing healthcare.
29. Satu, M., & Parvez, M. (2015). Review of integrated applications with AIML based chatbot. In *Proceedings of the 1st international conference on computer and information engineering, ICCIE 2015*, Rajshahi, Bangladesh, pp. 26–27.
30. Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism, 69*, 36–40.
31. Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*, 223–230.

AI-Driven Hospital Readmission Predictor for Diabetic Patients

Rishab Arya, Vyom Dutt Sharma, Gopal Gupta, and Adwitiya Sinha

Abstract In smart healthcare delivery, predictive methods have been proven to be revolutionary in present time, especially amidst several critical health issues and pandemics. Healthcare is the most dynamic field wherein the demand of prediction is very high as it can offer lifesaving solutions and precautionary measures to patients who are particularly suffering from chronic ailments. This chapter deals with predicting hospital readmission rate of diabetic patients. Hospital readmission is the procedure in which a former patient is readmitted after discharge because of a new diagnosis or due to negative changes in the old diagnosis, based on historical medical record. The analysis would cover and benefit patients whose symptoms seem to be indicating toward a possibility of diabetes in the future. Our work would also benefit the medical world by pointing out the medical factors which may have vital impact on diabetes. The results would also support the initiative by the Centers for Medicare and Medicaid Services and assist in hospital readmission reduction.

1 Introduction

The menace of diabetes is elevated globally, especially in low- and middle-economic nations. It is mainly caused by overweight problems, unhealthy meals, and stressful lifestyles. Around 422 million are affected by diabetes worldwide with about 1.5 million deaths being directly accountable to diabetes annually. As per a survey conducted in 2019 in India, the number of diabetic cases estimates to 77 million of which 57% of cases remain deprived of diagnosis. Studies confirmed that it is even expected to upsurge above 134 million by the end of 2045 [1]. This chronic disease has also affected 26 million US citizens, which includes 18.8 million citizens whose disease has been diagnosed and 7 million citizens who remain undiagnosed [2].

R. Arya · V. D. Sharma · G. Gupta · A. Sinha (✉)
Department of Computer Science and Engineering & Information Technology,
Jaypee Institute of Information Technology, Noida, India

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_13

Diabetes is a common disease in which the glucose or sugar levels in our blood cross the normal level and reach to an alarming or threatening point. Broadly, it is of two types:

- Type 1: Body is incapable to make insulin.
- Type 2: Body is incapable to make or use insulin.

Patients affected by this disease face several health issues, including kidney failure, blindness, lower limb paralysis, trauma, and cardiovascular disease. Yearly millions of dollars are spent on diabetes, and it keeps increasing rapidly – \$327 billion in 2017 from \$245 billion in 2012. The above statistical data tells us how diabetes is a burden to the society and why we need to tackle this problem efficiently. It is always fruitful to know the possible future using experience and knowledge so that we can act on the negative outcomes and give meaning to the quote “Prevention is better than cure.”

Prediction is turning out to be an important method in the medical world as it is enabling medical bodies to act on factors which were found out to be most affecting or a deciding factor in diagnosis of a disease, rather than evaluating all factors. In our chapter, we aim to analyze the disease by analyzing vital factors like age, H1Ac result, metformin level, etc. to predict possibility of readmission of diabetic patients or readmission of patients who were in early stages of diabetes when last admitted. For the experimentation, we have employed various concepts of data mining to analyze, organize, and pre-process our dataset. Our approach attempts to provide the best possible results in terms of model accuracy, so that it could be further used and help in saving millions of people suffering from diabetes.

Research Objectives

- To understand and predict hospital readmission patterns of diabetic patients using different factors listed in our dataset
- To analyze and conclude which of the factors greatly affect the readmission rates in hospitals
- To analyze and conclude that readmission rates are independent of which factors
- To visualize using the vast number of data visualization tools the trends and attributes which affect readmission of a patient

2 Existing Literatures and Background Study

This section discusses and showcases the relevance of our research through analysis of previous studies and aims at intriguing others toward the disease. Huge amounts of data are available when it comes to medical healthcare, and thus the modeling becomes complex [2]. Thus, there is a need to construct representative variables in accordance with the problem given, after which modeling and testing could be done. A proposed algorithm tries to help fit the real data for this modeling and testing purpose by looking at the variables. This method could turn out to be very useful to find the high risk of diabetes among the patients and evaluate the variable efficacy. Our analysis of the clinical data is more efficient, and the percentage of accuracy increased significantly with the proposed method.

Data mining is essential for finding useful information from huge datasets with the help of machine learning and database systems [3]. After applying algorithms like KNN and Bayesian on diabetes patient's databases which contain various attributes for diabetes prediction, we would be capable or experienced enough to guide the medical world. Stepwise nature of prediction using data mining makes it really easy for analysts to interpret results. *In another* study conducted in [4], the authors discuss how can a person's daily habits such as his/her sleeping routine, eating habits, and physical work along with other essential indicators like BMI, waist circumference, etc. can develop the risk of diabetes [4]. Chi-squared test of independence along with CART algorithms and cross validation of the resulting data to remove the biasness is used. Taking in day-to-day activities in analysis of predicting diabetes makes it more accurate as it monitors the unhealthy lifestyle followed by some people.

People are adopting to modern lifestyle at the cost of decreased physical activity and increased consumption of fast food [5]. These fast foods contain large amounts of fat and sugar which makes them vulnerable to diabetes. Blood tests for diabetes are not feasible economically and take time. We can help people by predicting if they are prone to diabetes or not and save their both time and money by using various ML (machine learning) algorithms. In this study, ensemble boosting algorithm with perceptron algorithm is used to improve the accuracy and performance of prediction. This method is tested on three publicly available datasets. Type 1 diabetes is very common and can be kept in control by the use of insulin injections [6]. The glucose level can be checked by CGM (continuous glucose monitoring) system. In this paper, the algorithm which used glucose prediction is kernel based adaptive filtering algorithm. Selection of a proper kernel function increases the accuracy of prediction. It is mainly used for short term blood glucose prediction. The results are interpreted, analyzed, and evaluated to ensure high accuracy level. In yet another study, Rapid Miner data tool is used to work on the dataset involving North American Indians (Pima), which collects the data of the people with diabetes and the ones moving toward it [7].

Modern lifestyles and advancements in the fields of medical healthcare have resulted in the increase of number of readmissions [8]. In this study, the algorithm used is C4.5 (Modified ID3 algorithm). Several data mining techniques are applied as well for preprocessing the data and for increasing the accuracy. A number of attributes were analyzed, and the ones highly affecting readmission were picked. The best result has come out to be 74.5%. India is home to 62 million diabetic patients [9]. The massive amounts of data available can be used for a good cause in the medical world. This would be done by analysis using artificial intelligence and computational analytics. Developing medical intelligence is very beneficial for the diabetic patients as it will help in reducing medical costs and hospital readmissions as well. This algorithm predicts diabetes in different environment and comes up various conclusions. Most prominent is 56% of women developing gestational diabetes in 24th week of diabetes.

The data mining techniques are quite popular for prediction that is why a number of papers including this one use it [10]. In this research paper, Bayes network is employed to predict the condition of patients who are affected with type 2 diabetes. The dataset used involves North American Indians (Pima), which gathers data of diabetic people who are already affected and those who are in prediabetic risk. The data mining tool used here is Weka. Using the Bayes network, accurate results have been obtained; hence, it is effective in predicting type 2 diabetes. Inclusion of conditional probability makes it convenient to understand the ongoing analysis. Type 2 diabetes is becoming common among teenagers and children and is increasing at a swift pace [11]. This paper targets a more real-life approach for prediction of type 2 diabetes. It is done by creating an iPhone application called HealthiManage which can be individually accessed. It monitors our physical activities carried out by us in our day-to-day life using an accelerometer which already exists in Apple devices. This application enables the people to self-manage the disease by alerting the user about changes in insulin level analyzed by it using measured and predicted readings. To encourage the patients, a reward component is also added which helps boosting patient confidence and faith toward fighting the disease and having control over it. This model provides an accurate feedback as it takes into control different activity recognition models and uses its predictive model and other useful elements and manages type 2 diabetes.

Data classification is considered a vital step before executing other data mining techniques [12]. If properly done, it can increase the prediction accuracy. This paper classifies the diabetes dataset using k-nearest neighborhood algorithm (KNN) and artificial neural network (ANN). The classification done by ANN gives better results as compared to the one done by KNN. In another research work, authors have studied to develop an analysis technique to evaluate how it makes our prediction accurate [13]. This technique finds the minimum number of attributes required for prediction algorithms and which attribute is important and which are not. This would ensure that we handle or foresee only the important attributes and discard the non-vital ones. This analysis is applied on two datasets: one of a heart disease and other of a diabetes dataset.

The following section highlights the dataset description, followed by the implementation and experimental outcomes of our proposed model design.

3 Data Collection and Consumption

This section summarizes the data source and files used, along with the programming platform and library tools. We have used R programming language for analyzing data. Our approach is based on developing machine learning model; hence, our requirement includes certain machine learning libraries of R environment.

3.1 Data Source

We have considered the Diabetes Dataset which is provided by UCI Machine Learning Repository [14]. According to the article [15], the following details related to Diabetes Dataset were gathered. The dataset constitutes data of 10 years (1999–2008). Our data has over 50 attributes and a total of 101,766 patients' information (both diabetic and to be diagnosed with diabetes). A number of features of our dataset which need to be listed before working on the data are as follows:

- The data in the file is an inpatient encounter (all attributes in the table are interaction confined in the medical facility).
- Length of stay is in the range of 1–14 days.
- All tests were performed in the medical lab only.
- Medications prescribed are done in hospital premises.
- Data includes both diabetic patients and patients whose test is pointing toward a future diagnosis of diabetes.

The data contains a number of attributes like weight of patient, age, HA1C result, number of days in hospital, gender, race, metformin, max glucose level, admission type, discharge source, and more to describe each patient's personal information.

3.2 Libraries Used

For accessing and working of our analysis chapter, a number of libraries have to be installed:

- Dplyr – for data cleaning purposes
- GGally – for plotting alongside ggplot2
- Corrplot – for building and plotting correlation matrices
- Caret – for linear regression
- Rpart – for linear regression as well as trees
- Nnet – for neural network models
- Naivebayes – for naïve Bayes model

3.3 Data Files

For accessing and executing our chapter, familiarity with following files is important:

- diabetic_data.csv

- Contains our diabetes dataset on which analysis using various data mining techniques is done
- Mapping.csv

- Contains data of mapped attributes
- AnalysisCode.r
- Contains code in R programming language through which analysis and interpretation of the diabetes dataset are done

4 Implementation Details

The section highlights the execution and implementation results, along with preliminary data cleaning followed by single and multivariate analysis.

4.1 Preliminary Data Cleaning

Our dataset is initially preprocessed by removing attributes and missing values. The following are the attributes with highest number of missing values:

- Weight (97% missing)
- Medical specialty (47% missing)
- Payer code (40% missing)

Subsequently, we have removed data of all patients whose discharge_disposition_id mapped to either death or discharge to a hospice. This step would ensure no biasing to occur in our analysis. Discharge to a hospice refers to the transfer of patient to a smaller medical facility (not related to the current hospital) due to monetary problems or a case of definite termination.

- 11→Expired
- 13→Hospice/home
- 14→Hospice/medical facility
- 19→Expired at home
- 20→Expired in a medical facility
- 21→Expired, place unknown

4.2 Data Cleansing

In this preprocessing step, a number of attributes would be removed from the dataset due to the following reasons:

- Attribute value is same for nearly all patients.
- Attribute is not directly related to the disease under analysis (various medical tests).

- Attribute is relevant for medical facility not for analysis.

On the basis of the above factors, the following attributes were removed: encounter_id, repaglinide, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, citoglipton, glipizide.metformin, nateglinide, chlorpropamide, glimepiride, glimepiride.pioglitazone, metformin.rosiglitazone, metformin.pioglitazone, glyburide.metformin, and glyburide.metformin.

4.3 Single Variable Analysis

The foremost step in every data analysis chapter is to systematically depict and order the data for better understanding. Here, we summarize the data and find patterns in single variables of the dataset. We can therefore analyze how much and how exactly our dataset varies.

Race is the division of people on the basis of similar physical features or characteristics or ancestral residence. The outcome in Fig. 1 is vital for inferring if physical features are a deciding factor for readmission or not. The graph in Fig. 2 depicts the count of people in different age groups admitted or readmitted to the 130

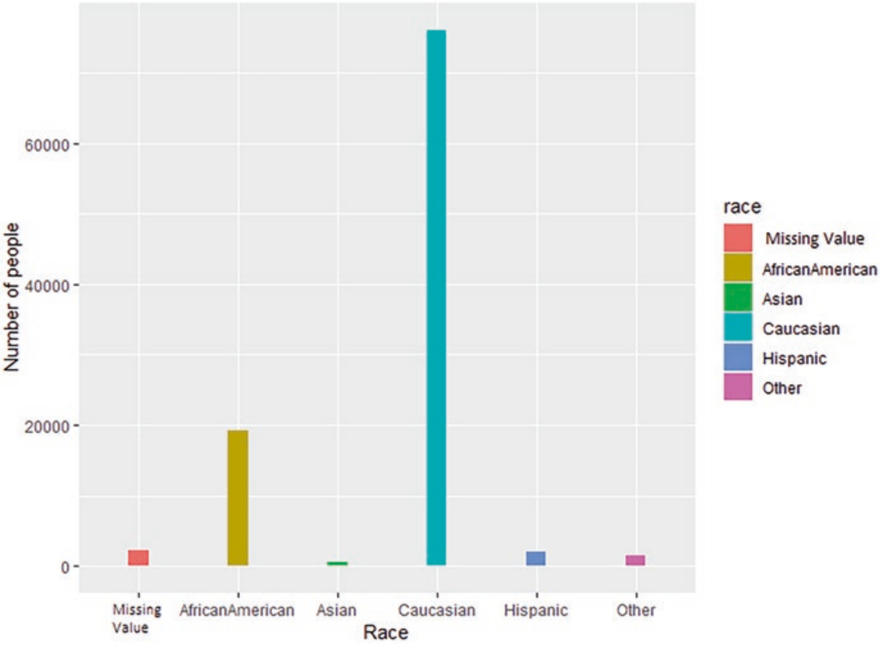


Fig. 1 Race wise distribution of data

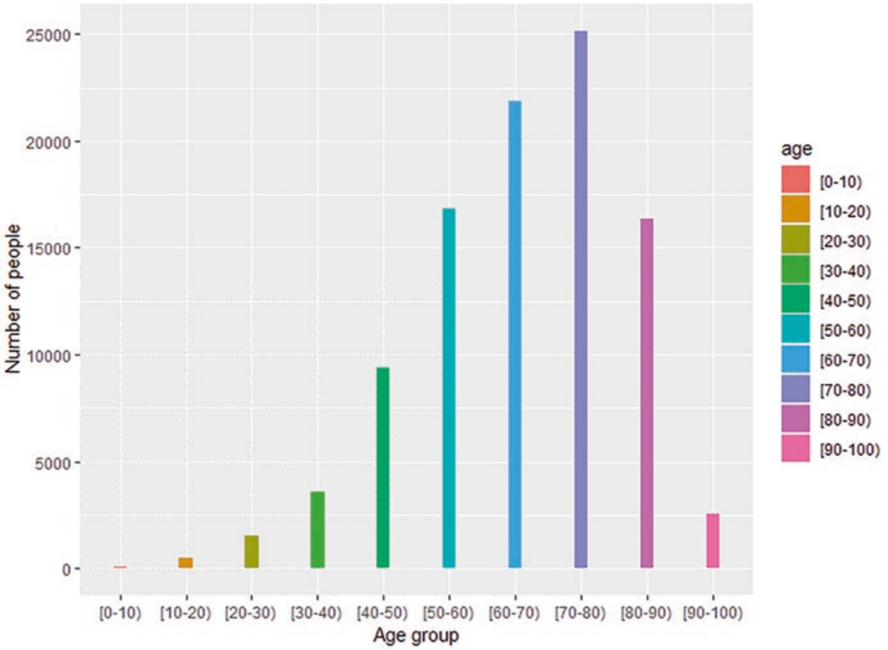


Fig. 2 Age wise distribution of data

hospitals whose data we are analyzing. This data is vital for inferring if age group is a deciding factor for readmission or not.

Figure 3 depicts division of patients on the basis of gender which would help by telling which gender is affected by the disease under analysis more than the other. Moreover, Fig. 4 highlights whether the patient in the dataset is already taking diabetes medication or not, i.e., this data is differentiating diagnosed and not or to be diagnosed patients. Further Fig. 5 highlights the distribution of data on the basis A1C test result.

Another prevalent test for diabetic patients is A1C test or hemoglobin A1C result (HA1C). It is also known as glycosylated hemoglobin test to monitor blood sugar control in people with diabetes. This medical test shows the sugar levels in the blood (average of the past 2–3 months). Biologically, it counts % of hemoglobin that have a layer of sugar on them. In other words, the A1C test provides a quantification of the content of glucose that is associated to hemoglobin. The presence of high glucose content can disrupt the functioning of hemoglobin and disturb the supply of oxygen throughout the body. Different divisions in the attribute are as follows:

- None: test not done on patient
- Norm: normal levels less than <5.7
- >7: greater than 7 but less than 8; diagnosed with diabetes but disease is under control
- >8: greater than 8; diagnosed with diabetes; not well controlled

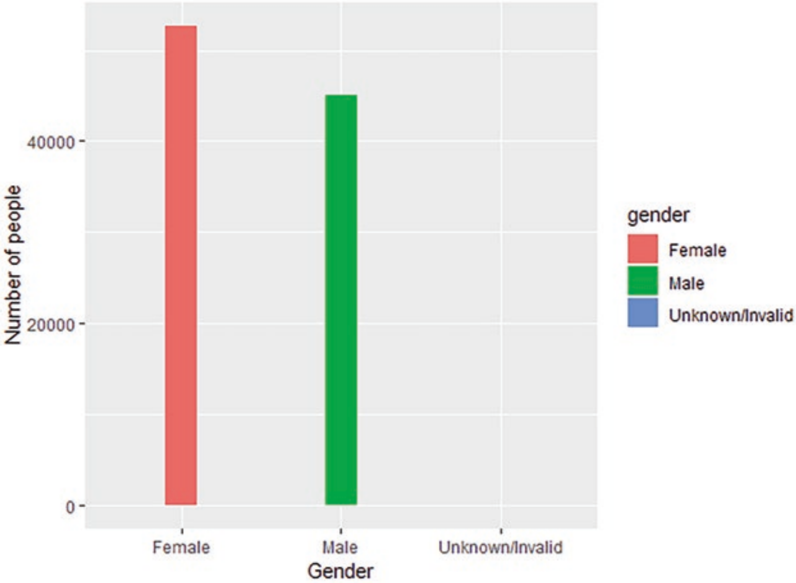


Fig. 3 Gender wise distribution of data

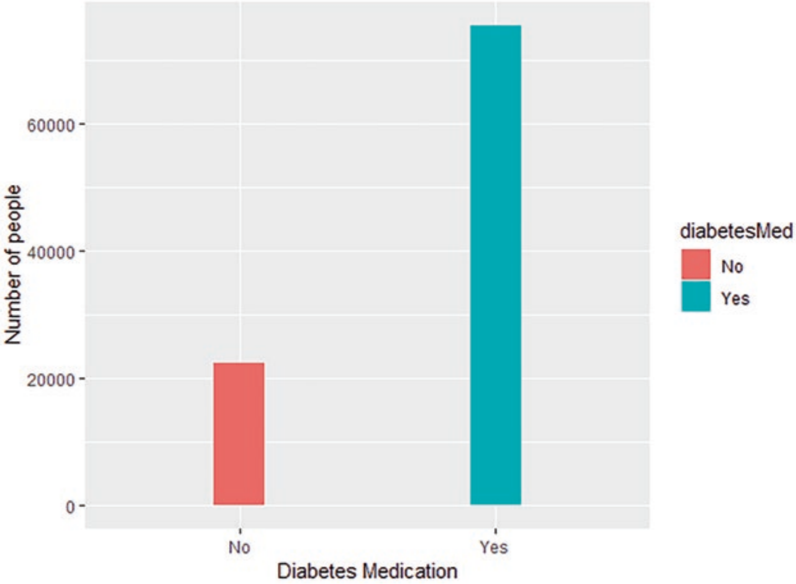


Fig. 4 Distribution of data on the basis of intake of diabetes medication

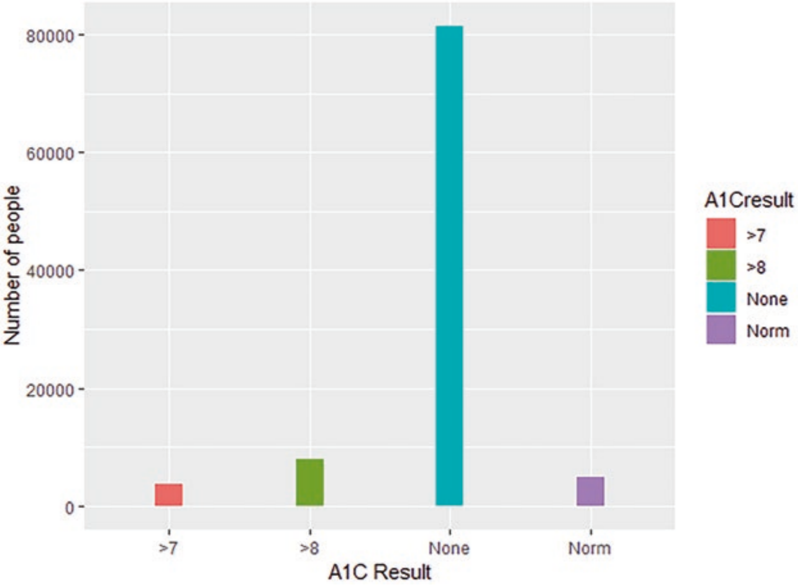


Fig. 5 Distribution of data on the basis of A1C test result

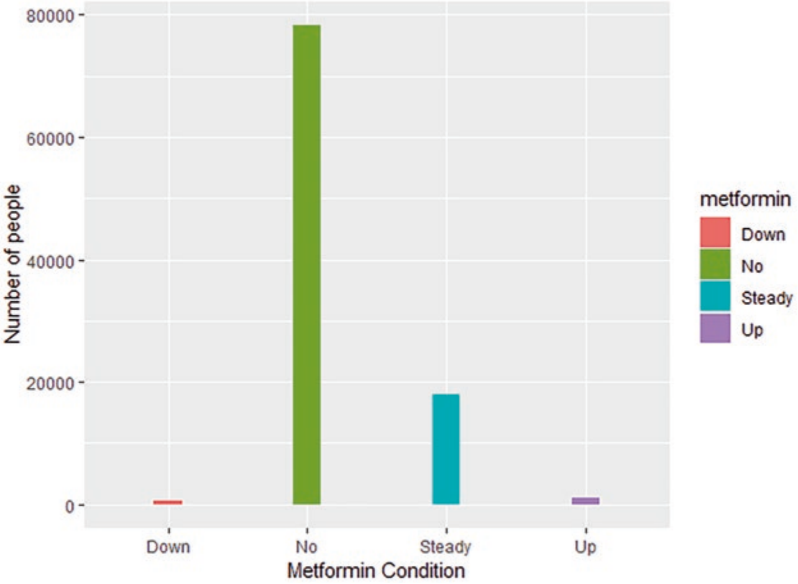


Fig. 6 Distribution on the basis of metformin levels in the blood

Figure 6 shows metformin levels in the blood of patients which indicates whether they were earlier on this medication or not. Metformin is a medication for type 2 diabetes patients which helps the body to use insulin and is generally prescribed to top obese patients. No division means the patient is not in metformin medication. Down, up, and steady are subparts of Yes indicating the medication levels in our blood.

In Fig. 7, the result analyzes the number of patient vs. the respective readmittance tenure. The divisions in attribute include the following:

- NO: not readmitted
- <30: within 30 days
- >30: after 30 days

4.4 Multivariate Analysis

This section deals with multivariable analysis helpful in comparing two attributes which are initially considered vital. This is an important step as it clearly divides dependent and independent attributes and tells us the level of difficulty in the prediction of result in the dataset. In our case, we check the "target" attribute, i.e., readmitted, and how it is related with some other attributes like age, gender, and A1C result (Fig. 8).

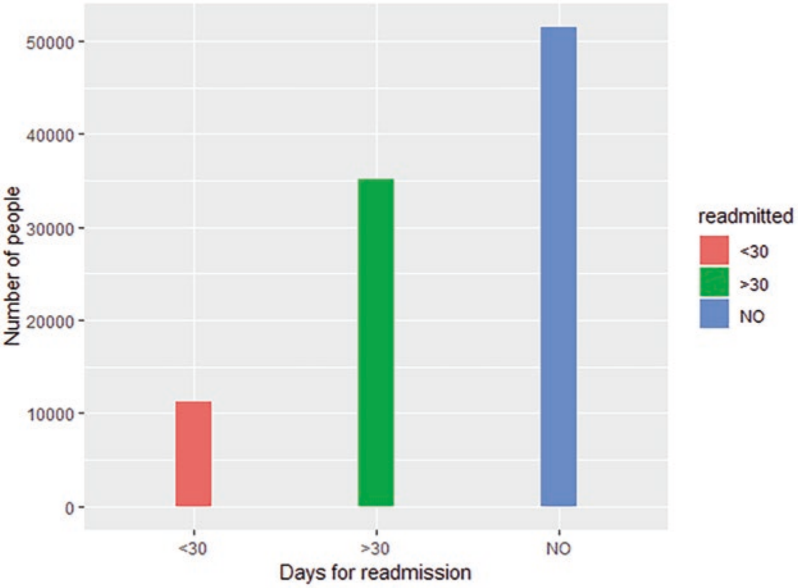


Fig. 7 Distribution of data on the basis of tenure after which they were readmitted

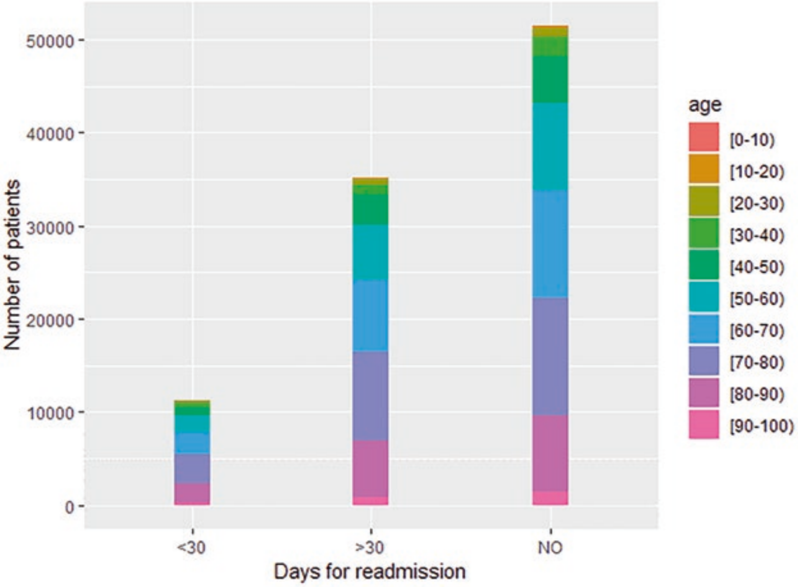


Fig. 8 Age wise analysis of readmission tenure

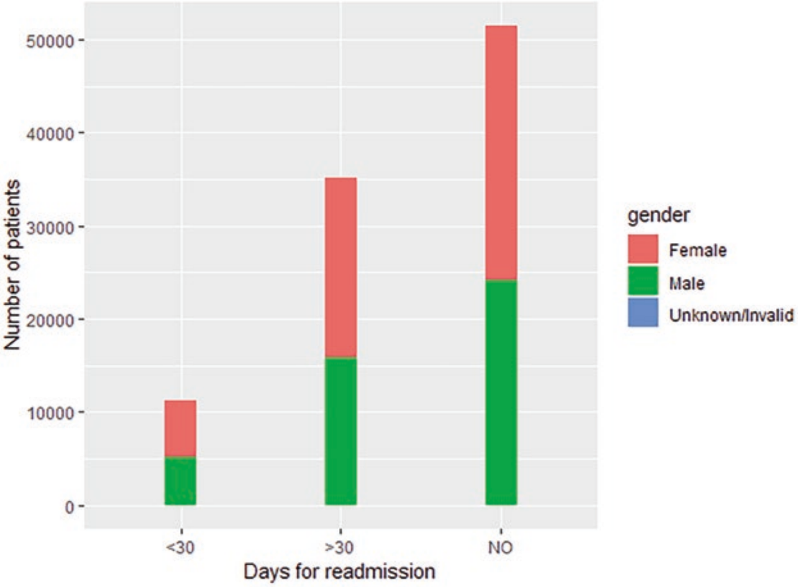


Fig. 9 Gender wise analysis of readmission tenure

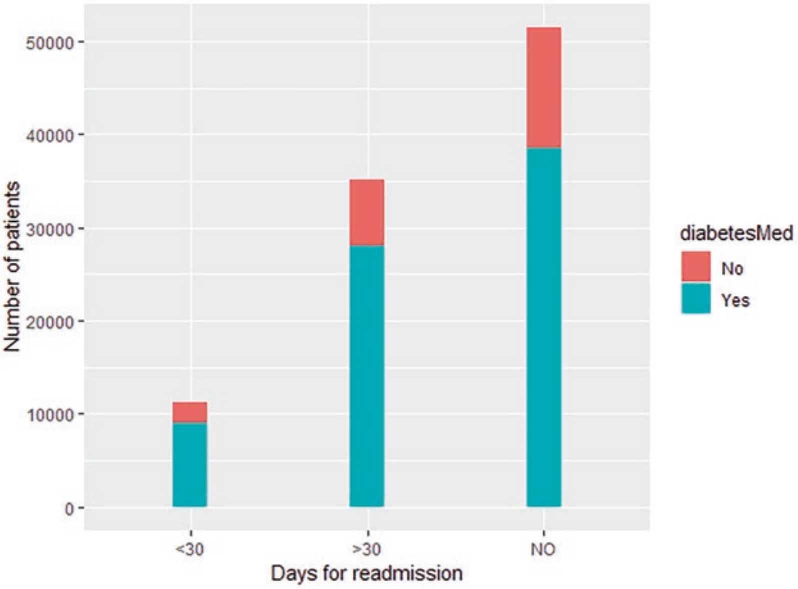


Fig. 10 Tenure of readmission under diabetes medication

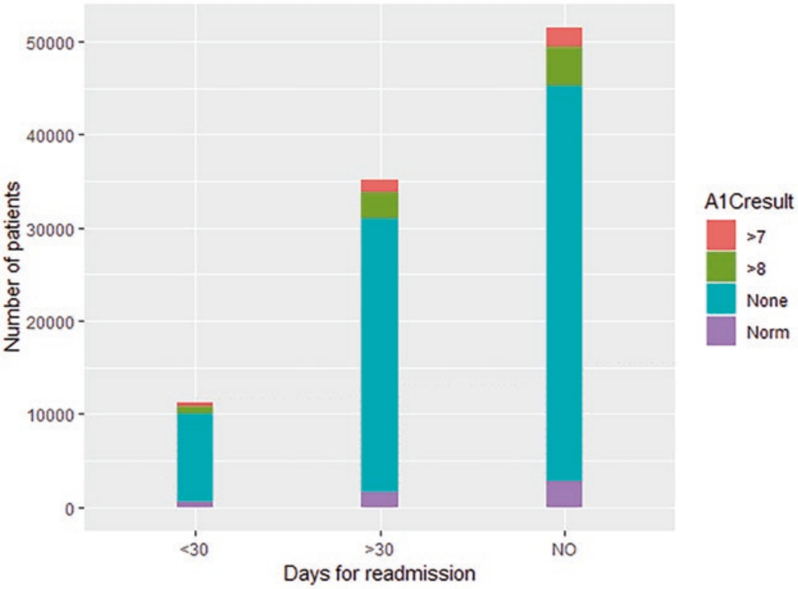


Fig. 11 Tenure of readmission on the basis of A1C result

Furthermore, the outcome in Fig. 9 can be used for systematic depiction of our pre-analysis to increase ease of understanding by directly comparing age and target variable, i.e., readmitted on a common axis. The data inferred here is the number of people in a particular age group who have been readmitted in the given tenure.

Further, we have examined the graph for systematic depiction of our pre-analysis by comparing gender (Fig. 10). This graph is for inferring the number of people who got readmitted given the fact that they were already taking some diabetes medication (Fig. 11). This graph is for systematic depiction of our pre-analysis to increase ease of understanding by directly comparing A1C result and target variable, i.e., readmitted on a common axis. The data inferred here is the number of people residing in a range of A1C result who have been readmitted in the given tenure.

4.5 Data Transformation

Data transformation is a fundamental process in data preprocessing, which involves converting or modifying raw data into a more appropriate format, as required for model building and evaluation. The aim of data transformation involves preparing the collected data for better modeling, handling missingness or redundancies, and eradicating outliers, along with performing feature scaling. This is an important step prior to modeling, as it can improve the accuracy and performance of machine learning models by addressing issues such as multicollinearity, overfitting, and bias.

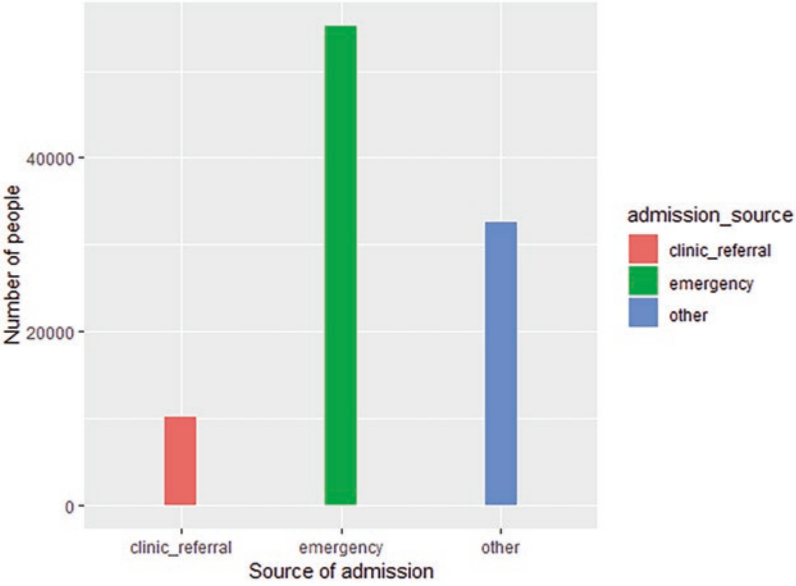


Fig. 12 Single variable analysis of admission source

In this section, we have transformed or mapped the attributes into divisions beneficial for understanding the dataset easily. The following is an attribute named admission source which has been transformed (Fig. 12). Clinic_referral is a superset of referral divisions; emergency was not altered; the rest of the divisions of admission sources like pregnancy, sick baby, etc. were not directly or currently related to our disease so they were added in other category.

4.6 Attribute Selection

This section examines the importance of attributes for our prediction and finds the strength of relation between them using correlation analysis, as shown in Fig. 13. The attribute selection phase is an important step in predicting readmission rates, because all features may not be relevant or informative for the given problem space.

Feature selection techniques, such as recursive feature elimination or principal component analysis, can help identify the most important features for the task. Also, a correlation matrix can be used to study the strength of relationship between sets of variables. We build a correlation matrix to get a general idea of what attribute is completely independent of other attributes in the data. This also tells us about the attributes that are more related to each other than other attributes. For instance, we

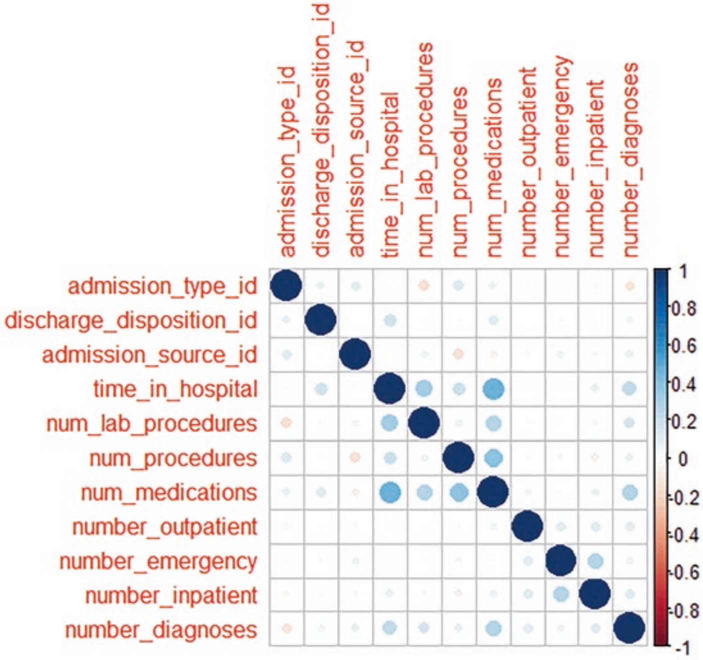


Fig. 13 Correlation matrix of features for model building

can see that “num_medications” attribute is highly related with “time_in_hospital” attribute which confirms the correctness of a data since it is a known fact that as the stay of patient in hospital increases, the number of meds prescribed also increases.

Principal component analysis (PCA) is considered as a significant technique for performing dimensionality reduction from high-dimensional to lower-dimensional feature space. The goal of PCA is to find the principal components, which are basically expressed as the linear combinations of the basic features, thereby capturing the variance in data. The principal components are orthogonal to each other, meaning they are uncorrelated. The first component in PCA captures the most amount of variance in the dataset, the second captures the second most variance, etc. The number of principal components to retain is typically chosen based on a threshold, such as retaining enough components to capture 90% of the total variance in the data. PCA can be used for various purposes, such as reducing the dimensionality of a dataset to visualize it in a lower-dimensional space, compressing data to reduce storage requirements, or improving the performance of machine learning models by reducing the noise and redundancy in the data. Figure 14 shows the principal component analysis process, which assists model building by selecting the most important attributes in a dataset, thereby further assisting in selecting those attributes that affect our target attribute, i.e., readmitted status.

4.7 Model Implementation

We have preprocessed our data and found out the vital attributes which affect the diabetes readmission prediction significantly. We have implemented different models on our dataset for binary and nonbinary cases. The dataset is partitioned into two

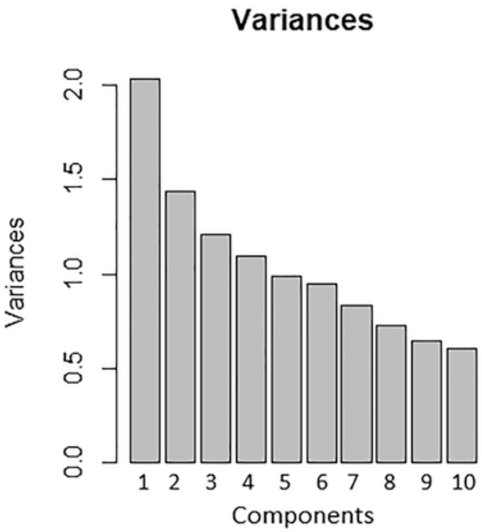
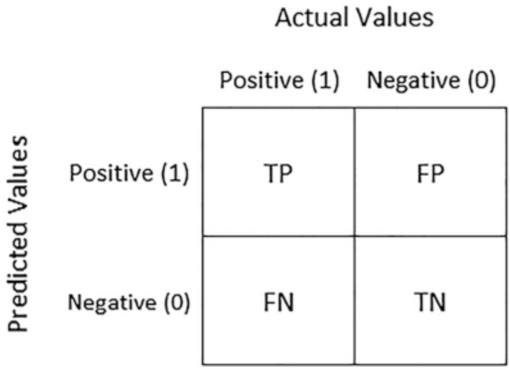


Fig. 14 PCA barograph



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Fig. 15 A general confusion matrix

parts, with 67% as training set and 33% as testing. The model accuracy would be found using confusion matrix which is a performance measure in tabular form applied on the testing set (Fig. 15).

The following are some acronyms to understand the above diagram and the formulas discussed in the subsequent section:

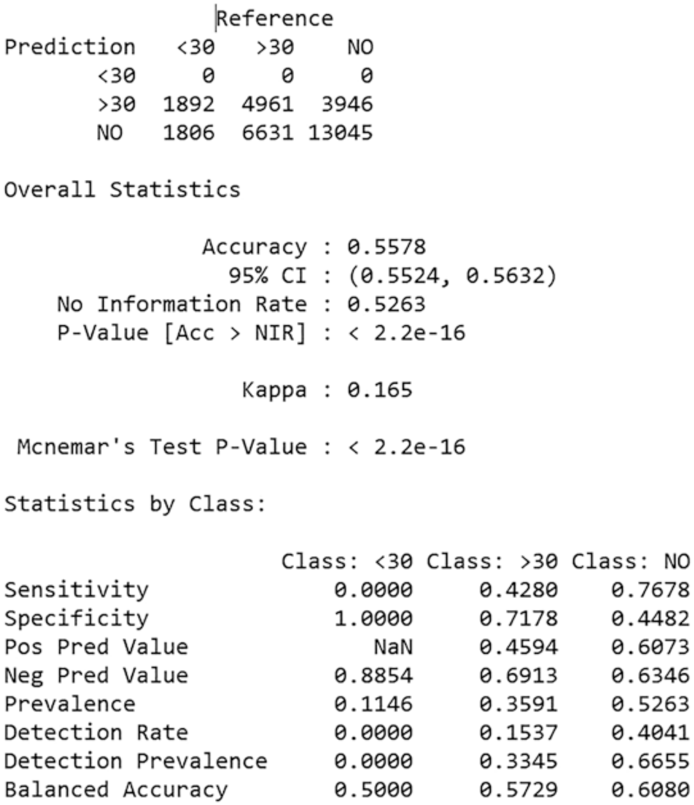
- TP → True positive
- FP → False positive
- TN → True negative
- FN → False negative
- N → All data

Before applying the models, we need to understand some terms which are necessary for model evaluation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{N}} \tag{1}$$

Accuracy can be described as the ratio of correct predictions to the sum total of predictions provided by the trained model. Moreover, sensitivity, also known as recall or true positive rate, can be referred as the ratio of appropriately predicted positive samples to the total number of actual positive samples. It measures the ability of a model to correctly identify positive samples. A high sensitivity indicates that the model can detect most of the positive samples. Also, specificity denotes the fraction of rightly predicted negative samples to sum total of actual negative samples. It measures the ability of a model to appropriately identify negative samples. The formulations are expressed in following equations (Eqs. 2 and 3):

$$\text{Sensitivity} = \frac{\text{TP}}{\left(\text{TP} + \text{FN}\right)} \tag{2}$$

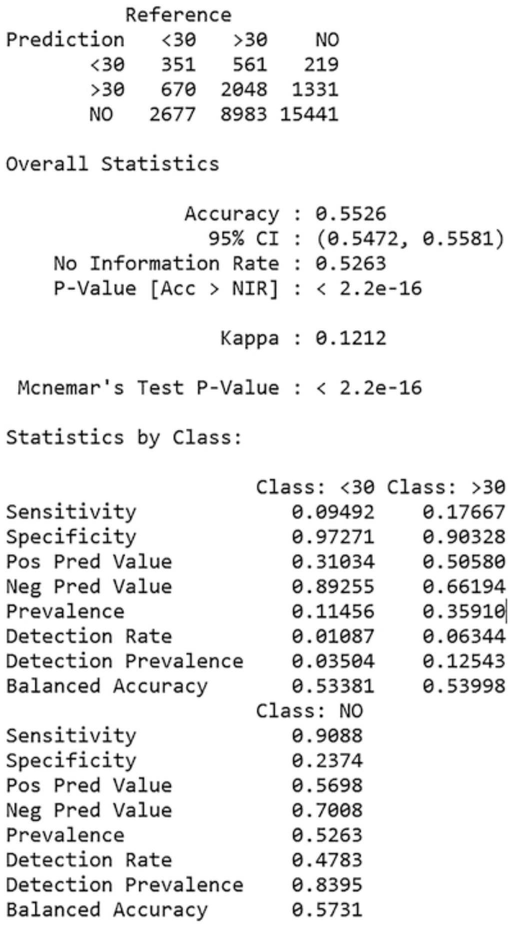


```
                 Reference
Prediction   <30   >30    NO
       <30     0     0     0
       >30  1892  4961  3946
       NO   1806  6631 13045

Overall Statistics

               Accuracy : 0.5578
                 95% CI : (0.5524, 0.5632)
    No Information Rate : 0.5263
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.165

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: <30 Class: >30 Class: NO
Sensitivity              0.0000     0.4280    0.7678
Specificity              1.0000     0.7178    0.4482
Pos Pred Value              NaN     0.4594    0.6073
Neg Pred Value           0.8854     0.6913    0.6346
Prevalence               0.1146     0.3591    0.5263
Detection Rate           0.0000     0.1537    0.4041
Detection Prevalence     0.0000     0.3345    0.6655
Balanced Accuracy        0.5000     0.5729    0.6080
```

Fig. 16 Confusion matrix and accuracy of decision tree model

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})} \tag{3}$$

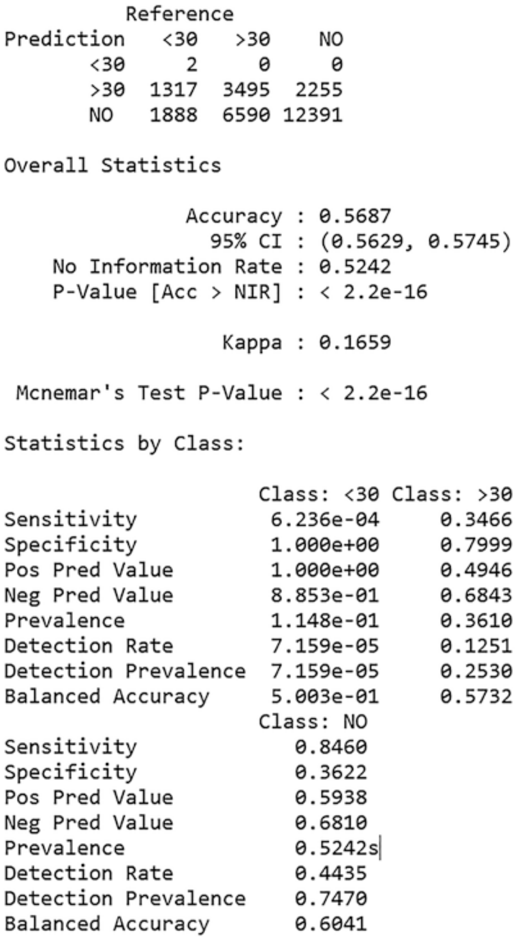
We have applied the several models to predict readmission of diabetic patients, including decision trees, naïve Bayes, artificial neural network, and logistic regression. The decision trees are mainly used in machine learning problems to build predictive models. They can be used for classification tasks, where the aim is to categorize data into distinct classes, or for regression tasks. Figure 16 showcases model evaluation of decision tree model using confusion matrix and several model parameters, like accuracy, sensitivity, specificity, balanced accuracy, etc. The balanced accuracy is the average of sensitivity and specificity. It is used when the dataset is imbalanced and the performance of the model needs to be evaluated on both positive and negative samples. It gives equal weightage to both sensitivity and specificity and is a good metric for evaluating the performance of a model on a balanced and imbalanced dataset.



```
          Reference
Prediction   <30   >30    NO
       <30   351   561   219
       >30   670  2048  1331
       NO   2677  8983 15441

Overall Statistics

               Accuracy : 0.5526
                 95% CI : (0.5472, 0.5581)
    No Information Rate : 0.5263
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.1212

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: <30 Class: >30
Sensitivity             0.09492    0.17667
Specificity             0.97271    0.90328
Pos Pred Value          0.31034    0.50580
Neg Pred Value          0.89255    0.66194
Prevalence              0.11456    0.35910
Detection Rate          0.01087    0.06344
Detection Prevalence    0.03504    0.12543
Balanced Accuracy       0.53381    0.53998
                     Class: NO
Sensitivity             0.9088
Specificity             0.2374
Pos Pred Value          0.5698
Neg Pred Value          0.7008
Prevalence              0.5263
Detection Rate          0.4783
Detection Prevalence    0.8395
Balanced Accuracy       0.5731
```

Fig. 17 Confusion matrix and accuracy of naïve Bayes model

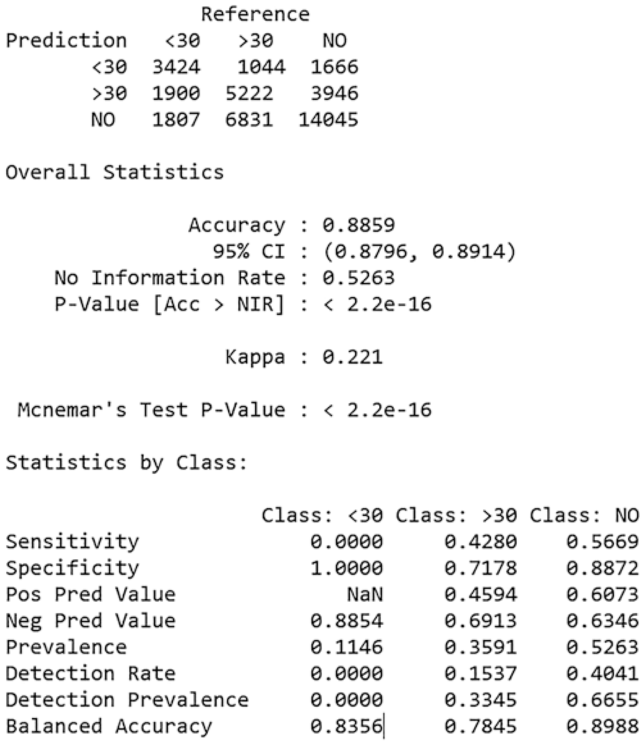
Naïve Bayes is a probabilistic norm of machine learning model used for classification tasks based on Bayes' theorem. While classifying new instances, naïve Bayes computes the posterior probability of each hypothesis (i.e., a class), using the attributes, and finally chooses the hypothesis having the maximum probability as the predicted class. The results in Fig. 17 showcase model evaluation of naïve Bayes model using confusion matrix which further gives details like accuracy, sensitivity, balanced accuracy, etc. Also, the constants, namely, kappa, and P-value are highlighted. The next model applied was artificial neural network (ANN). This allows feeding of the input data to the first layer of neurons, and then the outputs of each neuron in that layer are passed on as inputs to the next layer, and this continues until the final layer, which produces the output of the network. Figures 18 showcases model evaluation of neural network, for which accuracy and balanced accuracy were recorded to be 0.55 and 0.57, respectively. Hence, the model outcome is not recommendable in our case for hospital readmission prediction of diabetic patients.



```
          Reference
Prediction   <30   >30    NO
       <30     2     0     0
       >30  1317  3495  2255
       NO   1888  6590 12391

Overall Statistics

               Accuracy : 0.5687
                 95% CI : (0.5629, 0.5745)
    No Information Rate : 0.5242
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.1659

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: <30 Class: >30
Sensitivity           6.236e-04     0.3466
Specificity           1.000e+00     0.7999
Pos Pred Value        1.000e+00     0.4946
Neg Pred Value        8.853e-01     0.6843
Prevalence            1.148e-01     0.3610
Detection Rate        7.159e-05     0.1251
Detection Prevalence  7.159e-05     0.2530
Balanced Accuracy     5.003e-01     0.5732
                     Class: NO
Sensitivity             0.8460
Specificity             0.3622
Pos Pred Value          0.5938
Neg Pred Value          0.6810
Prevalence              0.5242s
Detection Rate          0.4435
Detection Prevalence    0.7470
Balanced Accuracy       0.6041
```

Fig. 18 Confusion matrix and accuracy of neural network model

Subsequently, another machine model, namely, logistic regression, was employed for binary and nonbinary classification problem, along with confusion matrix for various model parameters, as shown in Figs. 19 and 20. The logistic regression is a type of machine learning model used for binary as well as multi-class classification tasks. The logistic (sigmoid) function is used to transform a linear combination of the input features into a probability score, which can then be thresholded to make a binary prediction. The parameters of the logistic regression model are learned during training, using an optimization algorithm such as gradient descent, to maximize the likelihood of the training data. The likelihood is a measure of how well the model fits the observed data and assumes that the output is generated by a Bernoulli distribution with parameter equal to the probability score.



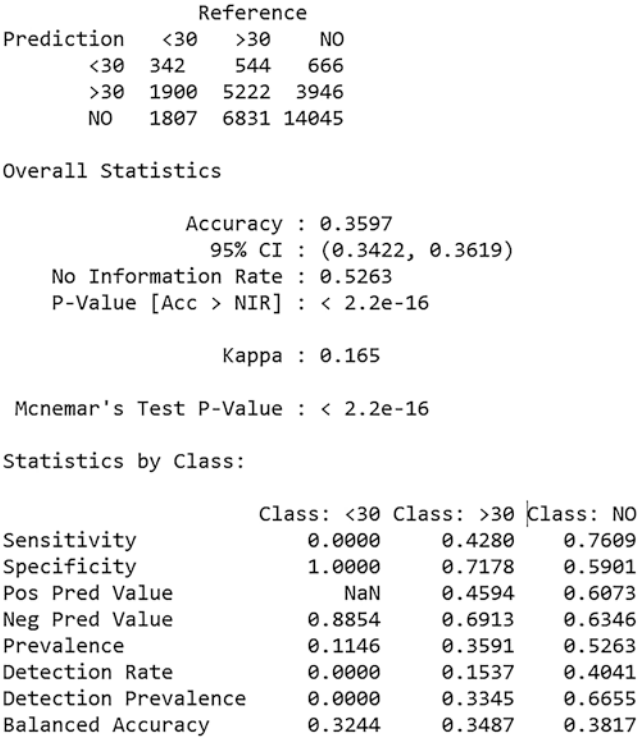
```
                    Reference
Prediction   <30   >30    NO
       <30  3424   1044  1666
       >30  1900  5222   3946
       NO   1807  6831  14045

Overall Statistics

               Accuracy : 0.8859
                 95% CI : (0.8796, 0.8914)
    No Information Rate : 0.5263
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.221

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: <30 Class: >30 Class: NO
Sensitivity              0.0000     0.4280    0.5669
Specificity              1.0000     0.7178    0.8872
Pos Pred Value              NaN     0.4594    0.6073
Neg Pred Value           0.8854     0.6913    0.6346
Prevalence               0.1146     0.3591    0.5263
Detection Rate           0.0000     0.1537    0.4041
Detection Prevalence     0.0000     0.3345    0.6655
Balanced Accuracy        0.8356     0.7845    0.8988
```

Fig. 19 Confusion matrix and accuracy of logistic regression model (binary)

When applied on our dataset, logistic regression proved to be robust to noise and outliers, hence better regularized to prevent overfitting. Our model implementation and evaluation of the diabetes data have led to some significant conclusions. It is apparent that the accuracy for binary target attribute was recorded to be 0.8859 and balanced accuracy score was 0.8988~0.90. Hence, Fig. 19 illustrates that maximum accuracy obtained was 90% using logistic regression model, which was maximum as compared to other existing models that were employed for our dataset.

5 Problem Significance and Discussion

Predicting hospital readmission rates of diabetic patients is an important task in healthcare analytics [16]. The goal is to find the critical patients who could be at an extreme risk of readmission within a certain time period after their initial discharge. Such prior knowledge can initiate appropriate interventions to be taken to prevent readmissions and improve patient outcomes, based on electronic health records [17]. For predicting readmission rates, machine learning algorithm can be used,



```
                Reference
Prediction   <30   >30    NO
       <30  342    544   666
       >30  1900  5222  3946
       NO   1807  6831 14045

Overall Statistics

               Accuracy : 0.3597
                 95% CI : (0.3422, 0.3619)
    No Information Rate : 0.5263
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.165

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: <30 Class: >30 |Class: NO
Sensitivity              0.0000     0.4280    0.7609
Specificity              1.0000     0.7178    0.5901
Pos Pred Value              NaN     0.4594    0.6073
Neg Pred Value           0.8854     0.6913    0.6346
Prevalence               0.1146     0.3591    0.5263
Detection Rate           0.0000     0.1537    0.4041
Detection Prevalence     0.0000     0.3345    0.6655
Balanced Accuracy        0.3244     0.3487    0.3817
```

Fig. 20 Confusion matrix and accuracy of logistic regression model (nonbinary)

which may include logistic regression, random forest, neural network, etc. trained with features extracted from patient data. Some possible features that could be used include the following:

- Patient demographics (age, gender, race, etc.)
- Clinical data (diagnosis, comorbidities, medications, lab results, etc.)
- Treatment data (length of stay, procedures performed, discharge disposition, etc.)
- Socioeconomic data (income, education, housing, etc.)

The data could be obtained from electronic health records, claims data, or other sources. After collecting the data and selecting the features, the next step is to train an artificial learning model to forecast the likelihood of readmission. The model could be evaluated using metrics such as model accuracy, recall, F1-score, precision, etc. [18]. Once a model is trained and validated, it can be used to identify critical diabetic patients. Moreover, certain healthcare interventions such as care management, medication adjustments, or lifestyle modifications can be targeted to such critical patients. By reducing readmission rates, the healthcare organizations can improve patient ailments and reduce healthcare costs, thereby improving the quality of care.

6 Conclusion and Future Work Directions

The hospital readmission rate prediction has proven to be a blessing for identifying the high-risk diabetic patients. This aims at expediating the proactive measures required to prevent readmission, such as adjusting medication regimens or providing additional education and awareness. For our experimentation, we have used dataset of the diabetic patients with some relevant features, including age, weight, gender, HA1C result, number of days under hospitalization, metformin, max glucose level, admission type, etc. We have applied several models, namely, decision trees, naïve Bayes, artificial neural network, and logistic regression, for prediction of hospital readmission for a patient. It was found that in our case, the logistic regression model provided approximately 90% balanced accuracy, which was maximum among its existing counterpart models. Our resultant outcomes and the trained model can be deployed for assisting healthcare service providers in reducing frequent hospital readmission rate, thereby promoting wellness and preventive care to diabetic patients.

References

1. Pradeepa, R., & Mohan, V. (2021). Epidemiology of type 2 diabetes in India. *Indian Journal of Ophthalmology, 69*(11), 2932.
2. Kaschel, H., Rocco, V., & Reinao, G. (2018, October). An open algorithm for systematic evaluation of readmission predictors on diabetic patients from data warehouses. In *2018 IEEE international conference on automation/XXIII Congress of the Chilean Association of Automatic Control* (ICA-ACCA) (pp. 1–6). IEEE.
3. Shetty, D., Rit, K., Shaikh, S., & Patil, N. (2017, March). Diabetes disease prediction using data mining. In *2017 international conference on innovations in information, embedded and communication systems (ICIIECS)* (pp. 1–5). IEEE.
4. Anand, A., & Shakti, D. (2015, September). Prediction of diabetes based on personal lifestyle indicators. In *2015 1st international conference on next generation computing technologies* (NGCT) (pp. 673–676). IEEE.
5. Mirshahvalad, R., & Zanjani, N. A. (2017, September). Diabetes prediction using ensemble perceptron algorithm. In *2017 9th international conference on computational intelligence and communication networks* (CICN) (pp. 190–194). IEEE.
6. Sun, X., Yu, X., Liu, J., & Wang, H. (2017, July). Glucose prediction for type 1 diabetes using KLMS algorithm. In *2017 36th Chinese control conference* (CCC) (pp. 1124–1128). IEEE.
7. Han, J., Rodriguez, J. C., & Beheshti, M. (2008, December). Diabetes data analysis and prediction model discovery using rapidminer. In *2008 second international conference on future generation communication and networking* (Vol. 3, pp. 96–99). IEEE.
8. Tamin, F., & Iswari, N. M. S. (2017, November). Implementation of C4. 5 algorithm to determine hospital readmission rate of diabetes patient. In *2017 4th international conference on new media studies* (CONMEDIA) (pp. 15–18). IEEE.
9. Prasad, S. T., Sangavi, S., Deepa, A., Sairabanu, F., & Ragasudha, R. (2017, February). Diabetic data analysis in big data with predictive method. In *2017 international conference on algorithms, methodology, models and applications in emerging technologies* (ICAMMAET) (pp. 1–4). IEEE.

10. Guo, Y., Bai, G., & Hu, Y. (2012, December). Using bayes network for prediction of type-2 diabetes. In *2012 international conference for internet technology and secured transactions* (pp. 471–472). IEEE.
11. Chemlal, S., Colberg, S., Satin-Smith, M., Gyuricsko, E., Hubbard, T., Scerbo, M. W., & McKenzie, F. D. (2011, April). Blood glucose individualized prediction for type 2 diabetes using iPhone application. In *2011 IEEE 37th annual Northeast bioengineering conference* (NEBEC) (pp. 1–2). IEEE.
12. Jasim, I. S., Duru, A. D., Shaker, K., Abed, B. M., & Saleh, H. M. (2017, August). Evaluation and measuring classifiers of diabetes diseases. In *2017 international conference on engineering and technology* (ICET) (pp. 1–4). IEEE.
13. Kanchan, B. D., & Kishor, M. M. (2016, December). Study of machine learning algorithms for special disease prediction using principal of component analysis. In *2016 international conference on global trends in signal processing, information computing and communication* (ICGTSPICC) (pp. 5–10). IEEE.
14. Dataset. archive.ics.uci.edu/ml/datasets/diabetes+130us+hospitals+for+years+1999-2008. Last accessed Jan 2021.
15. Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., & Clore, J. N. (2014). Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. *BioMed Research International*.
16. Singhal, S., Sinha, A., & Singh, B. (2022). Context awareness for healthcare service delivery with intelligent sensors. *Frontiers of Data and Knowledge Management for Convergence of ICT, Healthcare, and Telecommunication Services*, 61–83.
17. Miglani, P., Bhandari, J., Alreja, U., & Sinha, A. (2022). Electronic health record for sustainable eHealth. In *Advanced computational techniques for sustainable computing* (pp. 53–67). Chapman and Hall/CRC.
18. Chahal, E. S., Haritosh, A., Gupta, A., Gupta, K., & Sinha, A. (2019, October). Deep learning model for brain tumor segmentation & analysis. In *2019 IEEE 3rd international conference on recent developments in control, automation & power engineering* (RDCAPE) (pp. 378–383).

Gleason Grading System for Prostate Cancer Diagnosis

Aman Jatain, Manju, and Priyanka Vashisht

Abstract Artificial intelligence (AI) is the ability of a computer to think and behave like a human, without any human intervention. Through techniques like deep learning and neural networks where a machine is shown pictures of what it must classify, the model learns through examples like any other human would. The ability to mimic human intelligence is what makes these systems so accurate, and, when paired with a pathologist, these systems revolutionize cancer detection and patient prognosis. AI can allow pathologists to deal with complex relationships between different types of cancers and their features while enabling them to manage huge chunks of data in the form of high-resolution images. By making use of advanced algorithms, clinicians can remove subjectivity from decision-making which makes the process of diagnosis less biased. This research work is proposing an architecture to leverage the best out of state-of-the-art pretrained models and customize it to classify the type of prostate cancer without the need of a pathologist. This architecture is designed by using deep learning techniques such as CNN and EfficientNet; it is scalable on huge amounts of data, 300 GB tested in this work, and presented an effective way to combat the dataset size conundrum. With the use of such novel techniques, the aim of this work is to help pathologists all around the word to diagnose prostate cancer faster and better.

1 Introduction

Prostate cancer basically related to the prostate gland, and the reason behind this type of cancer is growing cells in an uncontrolled way. After certain period, cancer spreads even beyond the prostate which results in threat to the human life. In

---

A. Jatain · P. Vashisht
CSE Department, Amity University, Gurugram, Haryana, India

Manju (✉)
Department of Computer Science and Engineering & Information Technology,
Jaypee Institute of Information Technology, Noida, India

Manju et al. (eds.), *Artificial Intelligence-based Healthcare Systems*,
The Springer Series in Applied Machine Learning,
https://doi.org/10.1007/978-3-031-41925-6_14

general, there is no early sign of prostate cancer appearing in men. Globally, this type of gland cancer is one of the vital cancers happening to men. It is the sixth leading cause of cancer related deaths among men. In India, there is an increasing trend in the incidence of prostate cancer [1]. There is a survey done on men over 65, and it is found that 60% of cases are diagnosed, and the same study shows that diagnosis was less than 1% in men under 50. However, this type of cancer can be seen in men even in their 30s and 40s. Men with a family history of prostate cancer are more likely to develop prostate cancer than the general population [2]. In order to diagnose prostate cancer, a preliminary blood test is performed known as prostate specific antigen (PSA) test. If a higher level of PSA is found in test, doctors generally recommend another round of PSA test and a digital rectal exam (DRE) to check for abnormalities implying a tumor. When there is presence of any type of lump, then biopsy may be recommended. Once the biopsy process is over, the cancer is assigned a Gleason score which lies between the range of 2 and 10. Here, there is an observation that the lower score shows that there are less chances of spread. The moment the Gleason score is measured, it is converted into an ISUP grade on a 1–5 scale. The Gleason grading system is the most important prognostic marker for prostate cancer (PCa), and the ISUP grade has a crucial role when deciding how a patient should be treated. As shown in Fig. 1, grading process results in classifying cancer tissue in Gleason patterns (3, 4, or 5) based on the architectural growth patterns of the tumor. After the biopsy is assigned a Gleason score, it is converted into an ISUP grade on a 1–5 scale.

AI makes use of statistical models, otherwise known as artificial neural networks (ANN) which imitate human-like intelligence and are hence modeled on biological neural networks. They help capture the not-so-linear relationship between the features (pictures of cancer cells) with their corresponding outputs (detecting the presence of cancer in unmarked cells). Convolutional neural networks (CNNs) are neural networks (NN) which can be applied on each pixel in the image in a sliding manner as shown in Fig. 2.

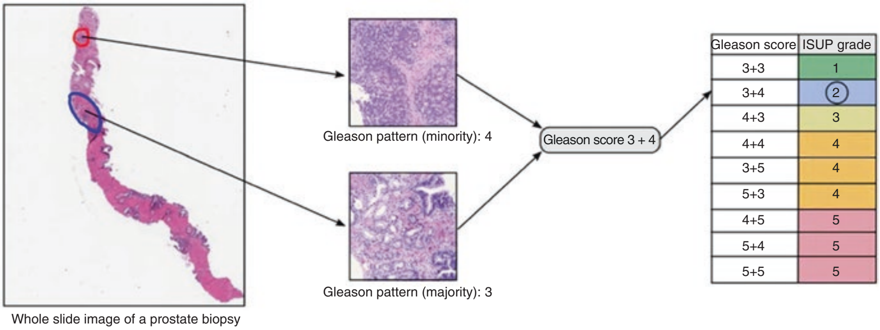


Fig. 1 An illustration of the Gleason grading process

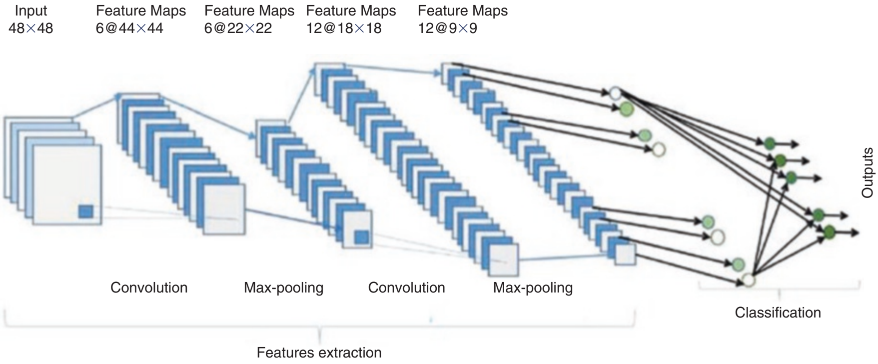


Fig. 2 Architecture of a standard convolutional neural network

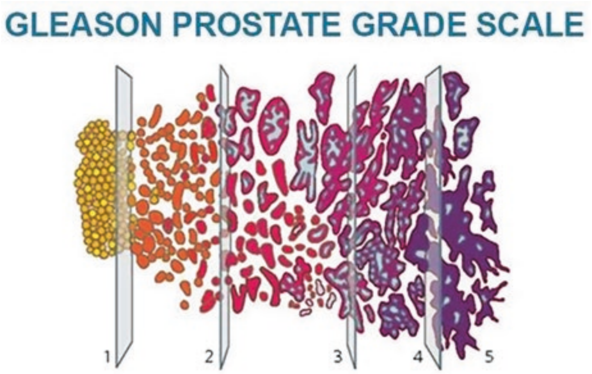


Fig. 3 Gleason grading system

With the use of techniques such as CNNs and pretrained models, a pathologist assisted by an AI tends to reach accuracies at detection of cancer which simply were not feasible before.

Historically, the Gleason score was used to categorize prostate cancer grades. When normal cells turn into cancer cells, they fall into five distinct patterns (Fig. 3). The scaling of 1–5 is used for grading of cells. Grade 1 prostate cells have a similar appearance to normal prostate tissue. This grade explains the extent of much a prostate cancer cells resemble a healthy prostate cell. This gives an idea to the doctor of how the cancer cell might behave during the treatment. To find the Gleason score or Gleason group, a pathologist looks at several biopsies from the patient's prostate.

2 Related Work

A comprehensive analysis is performed of the literature to assess how AI may be utilized to improve the detection and treatment of prostate cancer by accurately predicting the Gleason grade or score in WSI as well as through MRI.

2.1 Analysis of Prostate Biopsies as Whole Slide Images

Litjens et al. [3] in 2016 digitized 225 glass slides for evaluation, out of which 100 went into training the convolutional neural network (CNN), and 50 went into the validation set, while 75 digital slides were left for evaluation. The network consisted of four convolution layers, each followed by a max-pool layer. The last three layers of the network used by them were the classification layers which indicated whether an image patch is cancerous or not, based on the given features. The whole network was applied to each pixel of an image in a sliding manner. They obtained a 0.98 AUC score for the 90th percentile analysis in cancer diagnosis wherever it was present.

Ström et al. [4] were able to digitize around 6680 needle core biopsy slides from 976 individuals in their late 50s. The processed images were then fed to a deep neural network (DNN) that is a set of layers of neurons densely connected to each other. The model was tested by estimating the occurrence, magnitude, and malignant tissue's Gleason grade (GG). It attained an area under receiver operating characteristics curve (AUC-ROC) of 0.997 at 95% confidence level for being able to separate benign and malignant tissues. AUC of 0.96 for correctly identifying the presence of PCa and a mean pair-wise kappa of 0.62 for awarding GG were the interrelation between the model and a designated clinician (pathologist).

Campanella et al. [5] tested the structure of entire slide photos for PCa, achieving optimum outcomes on a PCa dataset with 44,732 samples with an AUC-ROC score of 0.989 at 20× magnification, while the automatic elimination of around 75% of images prompted no deprivation of sensitivity. By using an approach called multiple instance learning (MIL), they were able to develop a learning model that only uses the labels for training. Due to this approach, pixel-by-pixel manual annotations are thereby avoided, which are both costly and time-consuming. They used ResNet34 as the backbone of their network which had already been trained on the ImageNet dataset, thereby reducing the time required in training the model. A year prior to that, with 12,160 whole slide images, out of which 2424 were positive and 9736 were negative. With their best model on the VGG11-BN and ResNet34 of 0.977 and 0.976 AUC-ROC, respectively, Campanella et al. [6] obtained results using an AlexNet and a ResNet18 network which was trained on the full dataset and later pretrained on different image ImageNet models. They avoided image augmentation and went with a 15-15-70% split for validation, test, and train sets, respectively. A huge amount of computation was reduced for each image by adopting an approach called tiling, where the image was broken down into tiles of certain portions of the cell.

Lucal et al. [7] showed how retraining a state-of-the-art model, Inception V3 in their case, could yield great results. They studied a total of 96 biopsies from 38 individuals which were labeled on a very microscopic level. They focused on Gleason pattern (GP) of greater than 3 and 4, performing classification by retraining InceptionV3 CNN on such a small dataset. The CNN results are then translated into GP $\geq$ 3 and GP $\geq$ 4 probability maps. Further, the grade group of the whole biopsy is determined using the above probability maps. They showed that with proper training, the model could well differentiate between non-atypical and malignant with a sensitivity of 77%, accuracy of 91%, and specificity of 94%. Nagpal et al. [8] developed a deep learning (DL) model which was trained on a dataset of 112 million pathologist-annotated image patches from 1226 slides. The performance of the model was checked on a separate validation set of 331 slides. The model achieved an accuracy of 70% within a 98% confidence interval. The training of the model involves two steps, first being the ensemble of CNNs used to classify the patches and second being the collection of labels of GG assigned to each slide in the dataset.

Raciti et al. [9] recently reported findings on how an AI model such as Paige Prostate Alpha (PPA) can sway clinicians during PCa detection on needle biopsy samples. They used three AP board certified pathologists and assigned them the task of assessing prostate needle biopsies where they had to label each sample as benign or cancerous. The pathologists' sensitivity improved when diagnosing PCa with PPA, from the average sensitivity without PPA being 74% shooting up to 90% with the use of Paige Prostate Alpha. Bulten et al. [10] collected 5759 prostate cancer biopsies from 1243 individuals and created a DL model to identify the grade group and subsequently the Gleason score. The DL algorithm outperformed 10 pathologists out of the total of 15. The model achieved an AUC-ROC score of 0.990 on assessing the malignancy of a biopsy and an AUC-ROC score of 0.984 on an observer set.

Zeng et al. [11] used imaging modality like MRI and TRUS, and the imaging technique used was T2W and B-mode. The artificial intelligence CNN model was used, and the result was a depletion in the segmentation error and an aRVD depletion of 5–10%. Liu et al. [12] examined 50 confirmed patients with a Prostate Imaging-Reporting and Data System v2 score of 4 or 5. Imaging modality MRI was used, and imaging techniques like T1W, T2W, DW1, and DCE were employed. Various AI methods like SVM, RF, RFC, and KNN were used by Liu, and of all methods, KNN gave the best predicted efficacy and AUC of 0.88 with accuracy of 0.85. Tiny cancer patches isolated from benign tissue and PCa annotated regions were used to train a patch-based classifier in an observational analysis by Arvaniti et al. [13] on a training set of tissue microarrays from 641 individuals. The authors' deep learning patch model accomplished a far greater separation between the intermediate and low-risk classes of patients as compared to any pathologist. A kappa of 0.71 and 0.75 proved that the model showed a higher degree of interobserver consensus as compared to two pathologists with a kappa of 0.71.

According to Lenain et al. [14], pathology report carries data unavailable in the electronic health record format in an unstructured manner. They made use of three ML models, namely, decision trees (DT), support vector machines (SVM), and gradient boosting tree (GBT). They used words analyzation technique like natural language processing of 4470 patients. They managed to get an F1 score of 0.98 while classifying class N and an F1 score of 0.99 while classifying class M, where N and M belong to two different stage groups of PCa.

2.2 Diagnosis of Prostate Cancer Using MRI

Magnetic resonance imaging (MRI) is said to be an efficient method for diagnosing PCa and determining its aggressiveness [15]. Using tools such as apparent diffusion coefficient (ADC) and MR spectroscopic T2-weighted magnetic resonance imaging, professionals can very easily spot tumors which are malignant. On the other hand, determining the aggressiveness of the cancer was not so convenient. Hence, AI-facilitated diagnosis was proposed as a solution to this problem. Fehr et al. [16] proposed a new mechanism of utilizing ML-based classification by combining ADC and T2-weighted MRI-based features. This was highly proficient in separating high Gleason score from low ones with an accuracy of 93%. The model was also able to distinguish between two variations of GS7, which are (4 + 3) and (3 + 4) cancers. Aldoj et al. [17] analyzed 200 patients with a total of 218 lesions. The imaging modality that was used was MRI, and imaging techniques like T2W, ADC, and DWI were employed. CNN model of deep learning technique was used which gave an AUC score of 0.91 at 81.2% sensitivity compared to radiologist using PI-RADS v2. Ishioka et al. [18] reviewed data from 355 patients, and used imaging modality MRI images and imaging technique that was used was T2W. Deep learning model CNN was used, and an AUC of 0.645 was achieved in estimating the designated area containing cancer.

3 Gleason Grading Method

The Gleason grading of biopsies provides crucial prognostic data for individuals with prostate cancer and is a crucial component of therapy planning. The histological structure of the tumor tissue is used by pathologists to categorize cancers into several Gleason growth patterns. One of the most reliable ways to assess prostate cancer aggression was devised in 1967 and most recently revised in 2014, the Gleason grading system. Prostate adenocarcinoma growth patterns are described using Gleason scores, which are correlated with the severity of the condition. This method divides prostate tumors into five categories based on the differentiation patterns of the glands. The range is 1 (good outlook) to 5 (poor prognosis).

AI algorithms have shown promise for grading prostate cancer, specifically in prostatectomy samples and biopsies, and by assisting pathologists in the microscopic reviews and deep learning technology can make a substantial contribution to the automatic diagnosis of prostate cancer in tissues and the prognostication of cancer stage severity. The goal of the current study is to design and evaluate AI advancements in medical imaging in order to create and thoroughly validate the next algorithms for prostate cancer diagnoses and develop models for detecting PCa on images of prostate tissue samples using deep learning.

3.1 Method

The proposed framework consists of five steps: Data is loaded from Kaggle notebook, and exploratory data analysis (EDA) is performed on the dataset to check the data distribution of various data providers (Figs. 4 and 5). As the size of the dataset is too big so, all the non-informative parts from the images are removed using tiling of 128 * 128. Now the size of the dataset is reduced to 3 GB. Now packing the tiles so it gets converted to single image and then transfer learning is used to make the model more efficient (Figs. 6 and 7). Finally, model is tested on test dataset and gets isup_score for it (Fig. 8).

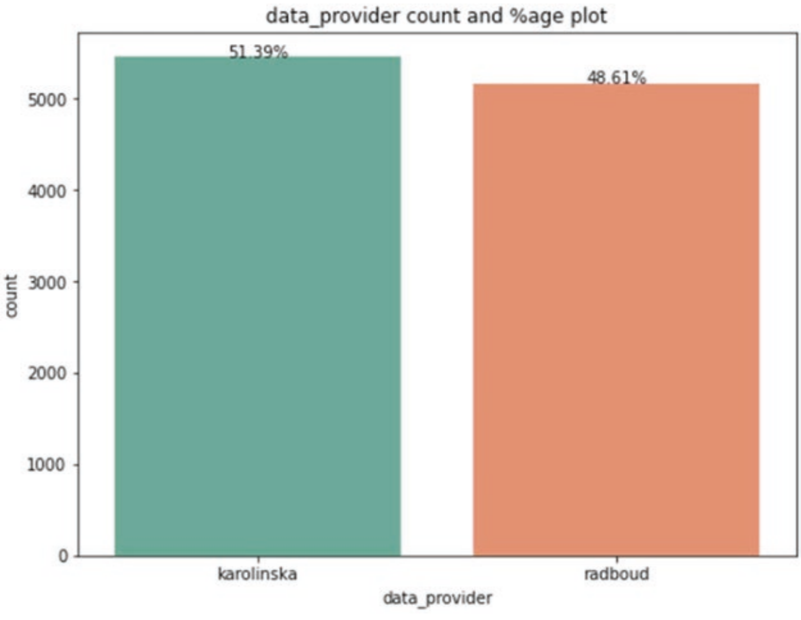


Fig. 4 Data distribution

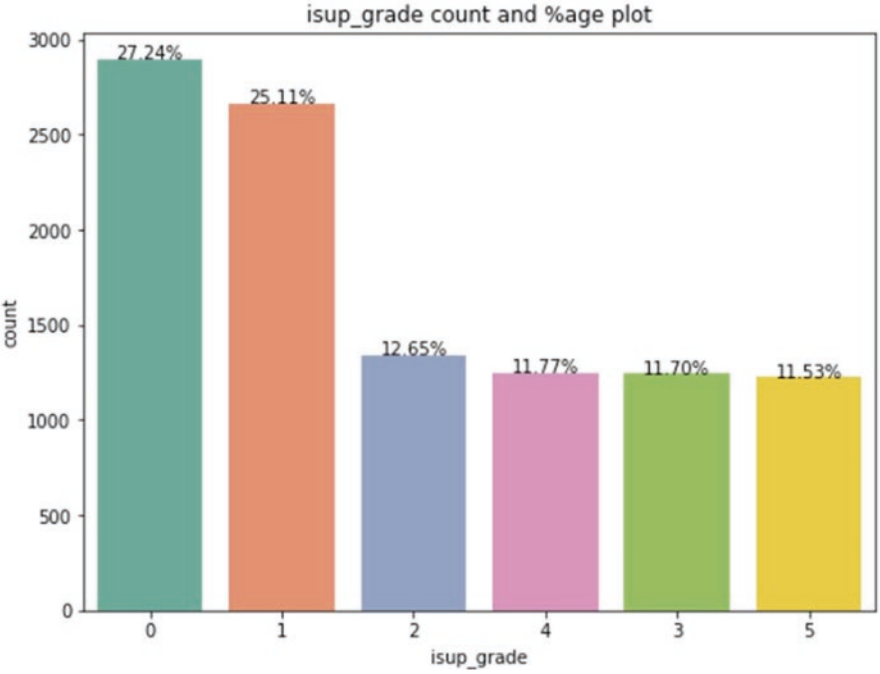


Fig. 5 isup grade distribution

Fig. 6 Unedited image

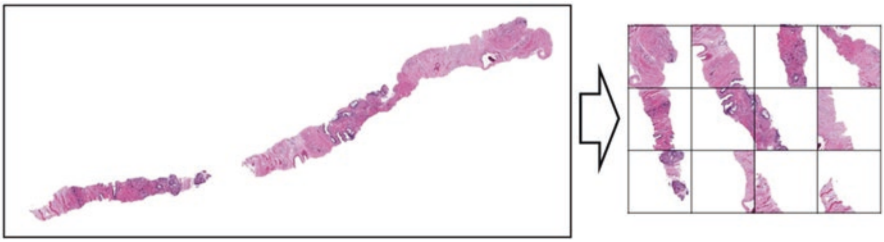


Fig. 7 Tiling

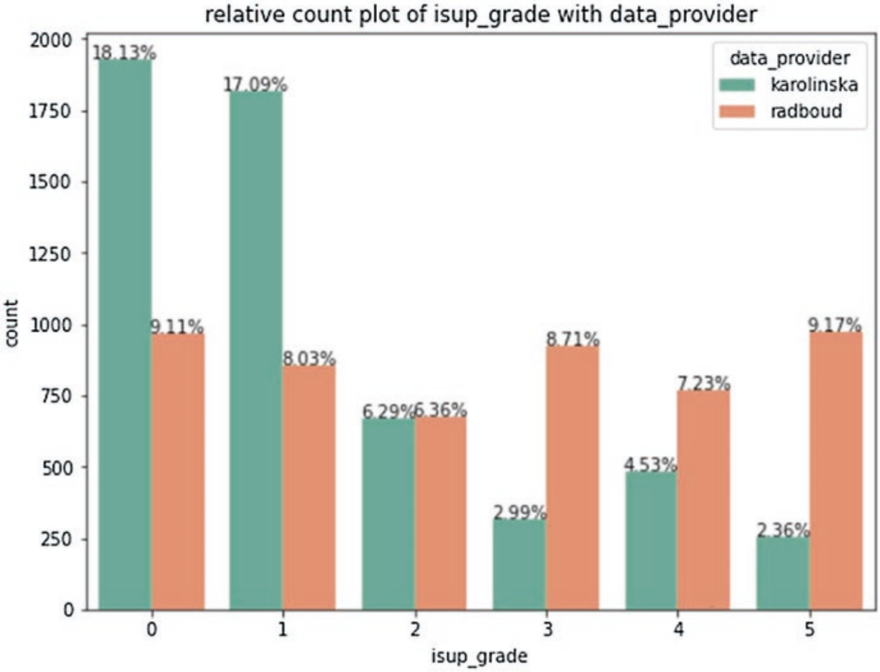


Fig. 8 Relative distribution of Gleason and isup

3.2 Characteristics of Dataset

The dataset was of a humongous size, 300GB, so it was virtually impossible to deal with the images at once. Even when leveraging Kaggle kernel, we only have a session time of 9 h. To avoid having to deal with such huge data, tiling techniques are used. The whole image is divided into 16 tiles of 128 × 128 containing pixels with value, i.e., the cell itself. However, by doing this, all the unnecessary white spaces cannot be omitted.

3.3 Transfer Learning Model: EfficientNet

EfficientNet is a convolutional neural network architecture (Fig. 9) and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient. The compound scaling method is justified by the intuition that if the input image is bigger, then the network needs more layers to increase the receptive field and more channels to capture more fine-grained patterns on the bigger image. EfficientNet-B0 network is based on the inverted bottleneck residual blocks of MobileNetV2, in addition to squeeze-and-excitation blocks. EfficientNet also transfers well and achieves state-of-the-art accuracy on CIFAR-100 (91.7%), flowers (98.8%), and three other transfer learning datasets, with an order of magnitude of fewer parameters.

3.4 Metric Used

The metric of choice for this problem was quadratic weighted kappa. The quadratic weighted kappa is calculated as follows. First, an $N \times N$ histogram matrix O is constructed, such that $O_{\mathrm{i,j}}$ corresponds to the number of true outputs i (actual) that received a predicted value j. An $N \times N$ matrix of weights, w, is calculated based on the difference between actual and predicted values:

$$w_{\mathrm{i,j}} = \frac{(i-j)^2}{(N-1)^2}$$

An $N \times N$ histogram matrix of expected outcomes, E, is calculated assuming that there is no correlation between values. This is calculated as the outer product between the actual histogram vector of outcomes and the predicted histogram vector, normalized such that E and O have the same sum. From these three matrices, the quadratic weighted kappa is calculated as:

$$K = 1 - \frac{\sum_{\mathrm{i,j}} w_{\mathrm{i,j}} O_{\mathrm{i,j}}}{\sum_{\mathrm{i,j}} w_{\mathrm{i,j}} E_{\mathrm{i,j}}}$$

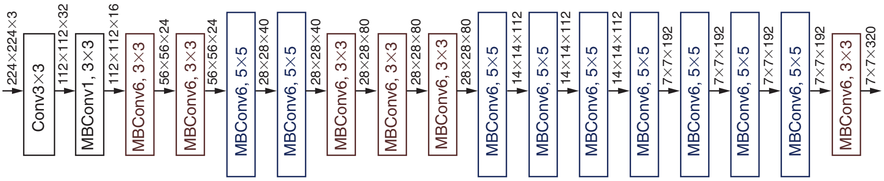


Fig. 9 EfficientNet architecture

Quadratic weighted kappa score is a ratio that can take a value between −1 and 1. A negative QWK score implies that the model is “worse than random.” A random model should give a score of close to 0. Lastly, perfect predictions will yield a score of 1.

3.5 Training and Evaluation

For training, a batch size of 2000 images is taken and the model ran for 40 epochs, which is the most that can run on Kaggle (9 h). A test set out of the images is made that is not used and sampled 100 images each time for testing. All the preprocessing steps are completed that we used during training for our test data and passed it onto the image loader. The same was passed onto the model for inference. The predictions were out in a stochastic manner, where we took the max and found the index of the prediction label as shown in Fig. 10.

Packing all tiles in a single image so all white spaces are removed as shown in Fig. 11.

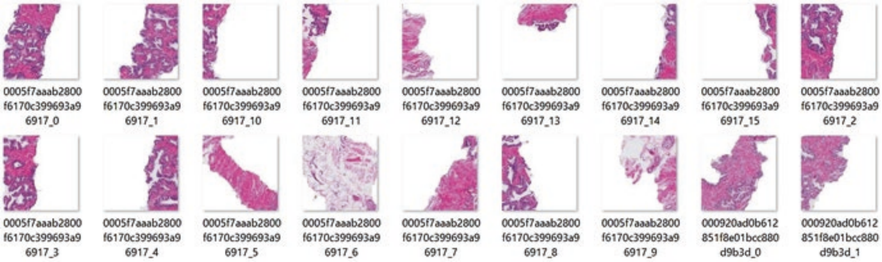


Fig. 10 Tiles after converting image into 16 tiles

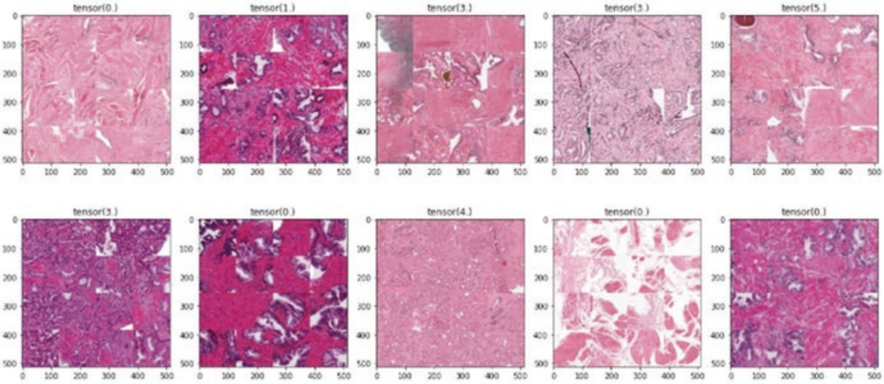


Fig. 11 Stacking tiles into an image

We trained the model for 40 epochs with a sample size of 2000 images out of the original 10,000 because of memory constraints. The model was trained on an Nvidia P-100 GPU for 9 h, which is the maximum allowed duration for a Kaggle kernel to stay connected. Post 9 h and 40 epochs, the model posted a quadratic weighted kappa of 70%, which is impressive considering the memory and GPU restrictions it was trained on. For reference, the winner of the contest containing the dataset trained his model on two GPUs in parallel, where each epoch took 10 h to complete.

4 Conclusion and Future Scope

The model obtained after 9 h of training is good enough to be deployed in production for a pathologist to help take a second look at their labeling, if not used for complete automation of identification. Considering the constraints we worked under, the model fared very well due to the use of a state-of-the-art pretrained model called EfficientNet which is developed by AI engineers at Google. The model is very flexible in terms of scaling depth wise as well as layer wise. It can adapt to very high-resolution images with no problem or bottlenecking in performance. All in all, a quadratic kappa score of 70% on commodity hardware is an achievement purely because of the choice of architecture and tuning of the model. This work can be improved by using the following approaches: increasing individual tile resolution, increasing number of tiles so as to capture more information, training on a larger batch size, sampling more images per epoch, training for much longer, and fine-tuning the architecture even more

References

1. Indian Prostate Cancer Foundation – IPCF. http://www.indianprostatecancerfoundation.com/#view2. Accessed 19 Nov 2020.
2. What Causes Prostate Cancer?. https://www.webmd.com/prostate-cancer/guide/what-causes-prostate-cancer. Accessed 19 Nov 2020.
3. Litjens, G., Sánchez, C. I., & Timofeeva, N. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports, 6*, 1–11.
4. Ström, P., Kartasalo, K., Olsson, H., et al. (2020). Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: A population-based, diagnostic study. *The Lancet Oncology, 21*, 222–232.
5. Campanella, G., Hanna, M. G., Geneslaw, L., et al. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine, 25*, 1301–1309.
6. Campanella, G., Krauss Silva, V. W., & Fuchs, T. J. (2018). Terabyte-scale deep multiple instances learning for classification and localization in pathology. *arXiv*.
7. Lucas, M., Jansen, I., Savci-Heijink, C. D., et al. (2019). Deep learning for automatic Gleason pattern classification for grade group determination of prostate biopsies. *Virchows Archiv, 475*, 77–83.

8. Nagpal, K., Foote, D., & Liu, Y. (2019). Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *npj Digital Medicine, 2*.
9. Raciti, P., Sue, J., Ceballos, R., et al. (2020). Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. *Modern Pathology, 33*, 2058–2066.
10. Bulten, W., Pinckaers, H., van Boven, H., et al. (2020). Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study. *The Lancet Oncology, 21*, 233–241.
11. Karimi, D., Zeng, Q., Mathur, P., et al. (2019). Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images. *Medical Image Analysis, 57*, 186–196. https://doi.org/10.1016/j.media.2019.07.005
12. Liu, B., Cheng, J., Guo, D. J., et al. (2019). Prediction of prostate cancer aggressiveness with a combination of radiomics and machine learning-based analysis of dynamic contrast- enhanced MRI. *Clinical Radiology, 74*, 896.
13. Arvaniti, E., Fricker, K. S., Moret, M., et al. (2018). Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific Reports, 8*, 1–11.
14. Lenain, R., Seneviratne, M. G., Bozkurt, S., et al. (2019). Machine learning approaches for extracting stage from pathology reports in prostate cancer. *Studies in Health Technology and Informatics, 264*, 1522–1523.
15. Moore, C. M., Petrides, N., & Emberton, M. (2014). Can MRI replace serial biopsies in men on active surveillance for prostate cancer? *Current Opinion in Urology, 24*, 280–287.
16. Fehr, D., Veeraraghavan, H., Wibmer, A., et al. (2015). Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America, 112*, E6265–E6273.
17. Aldoj, N., Lukas, S., Dewey, M., & Penzkofer, T. (2020). Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network. *European Radiology, 30*, 1243.
18. Ishioka, J., Matsuoka, Y., Uehara, S., et al. (2018). Computer-aided diagnosis of prostate cancer on magnetic resonance imaging using a convolutional neural network algorithm. *BJU International, 122*, 411–417.