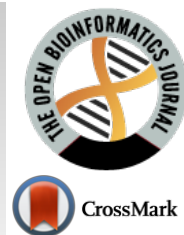




# The Open Bioinformatics Journal

Content list available at: <https://openbioinformaticsjournal.com>



## RESEARCH ARTICLE

### Electronic Health Record (EHR) System Development for Study on EHR Data-based Early Prediction of Diabetes Using Machine Learning Algorithms

Jagadamba G<sup>1</sup>, Shashidhar R<sup>2</sup>, Gururaj H L<sup>3,\*</sup>, Vinayakumar Ravi<sup>4,\*</sup>, Meshari Almeshari<sup>5</sup> and Yasser Alzamil<sup>5</sup>

<sup>1</sup>Department of Information Science and Engineering, Siddaganaga Institute of Technology, Tumakuru, Karnataka- 57210, India

<sup>2</sup>Department of Electronics and Communication Engineering, JSS Science and Technology University, Mysuru, Karnataka 570006, India

<sup>3</sup>Department of Information Technology, Manipal Institute of Technology, Bengaluru, Manipal Academy of Higher Education, Manipal, 576104, India

<sup>4</sup>Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

<sup>5</sup>Department of Diagnostic Radiology, College of Applied Medical Sciences, University of Ha'il, Ha'il, Saudi Arabia

#### Abstract:

##### Aims:

This research work aims to develop an interoperable electronic health record (EHR) system to aid the early detection of diabetes by the use of Machine Learning (ML) algorithms. A decision support system developed using many ML algorithms results in optimizing the decision in preventive care in the health information system.

##### Methods:

The proposed system consisted of two models. The first model included interoperable EHR system development using a precise database structure. The second module comprised of data extraction from the EHR system, data cleaning, and data processing and prediction. For testing and training, about 1080 patients' health record was considered. Among 1080, 1000 records were from the Kaggle dataset, and 80 records were demographic information from patients who visited our health center of Siddaganaga organization for a regular checkup or during emergencies. The demographic information was collected from the proposed EHR system.

##### Results:

The proposed system was tested for the interoperability nature of the EHR system and accuracy in diabetic disease prediction using the proposed decision support system. The proposed EHR system development was tested for interoperability by random updations from various systems maintained in the laboratory. Each system acted like the admin system of different hospitals. The EHR system was tested for handling the load and interoperability by considering user view status, system matching with the real world, consistency in data updations, security *etc.* However, in the prediction phase, diabetes prediction was concentrated. The features considered were not randomly chosen; however, the features were those prescribed by a doctor who insisted that the features were sufficient for initial prediction. The reports collected from the doctors revealed several features they considered before giving the test details. The proposed system dataset was split into test and train datasets with eight proper features taken as input and one set as a target variable where the result was present. After this, the model was imported using standard "sklearn" libraries, and it fit with the required number of estimators, that is, the number of decision trees. The features included pregnancies, glucose level, blood pressure, skin thickness, insulin level, bone marrow index, diabetic pedigree function, age, weight, *etc.* At the outset, the research work concentrated on developing an interoperable EHR system, identifying the expectation of diabetic and non-diabetic conditions and demonstrating the accuracy of the system.

##### Conclusion:

In this study, the first aim was to design an interoperable EHR system that could help in accumulating, storing, and sharing patients' timely health records over a lifetime. The second aim was to use EHR data for early prediction of diabetes in the user. To confirm the accuracy of the system, the system was tested regarding interoperability to support early prediction through a decision support system.

**Keywords:** Electronic health record (EHR), Electronic medical record (EMR), Machine learning, Artificial intelligence, Random forest algorithm, Healthcare, Classification, Diabetic prediction.

#### Article History

Received: April 06, 2023

Revised: June 19, 2023

Accepted: July 07, 2023

## 1. INTRODUCTION

Many levels of care centers in India follow different approaches for managing the demographics related to a patient. Traditionally, in India, many hospitals manage patient health information in the paper format, including Primary Health Care Centre (PHC), Secondary Health Centre (SHC), and Tertiary Healthcare Centre (THC). However, the paper format system is an unorganized way of keeping records. As a result, the providers cannot provide better care because of the unavailability of patient health information. But presently, most of the SHC and THC have started digitalizing health records to control medical information in a better way. Hence, the Electronic Medical Record (EMR) plays a significant role in healthcare management [1].

EMR is an electronic version of a patient's health information collected from a single healthcare organization, *i.e.*, a centralized electronic repository of the particular care center. This system assists healthcare professionals in obtaining and sending patient data within the hospital while also controlling the flow of information at the time of treatment [2]. Each care center organization has an EMR system for collecting patient data in an electronic format. Still, complete health information will not be available to care center professionals during care time, and vendors use different standards for clinical data exchange. In many healthcare centers, the user is forced to use different EMR applications, and the same procedures are repetitive. In these situations, there will be no uniform access to get the data or necessary details at the point of care, and patient records will be distributed in various EMR programs to get complete health information about the patient. In some situations, healthcare centers must obtain patient records in various applications by logging in multiple times to get the user's full details or carry a paper document in a file. Even though EMR is a digital record, many challenges are identified when utilized in the healthcare industries, and few are discussed [3]:

- EMRs are designed in a way that will not be allowed to be shared outside the individual practice.
- Patient medical history from various healthcare institutes will not be available or updated when new changes are available.
- As EMR is maintained in a centralized way, repeated testing in different healthcare often adds to the patient's time and money.
- While sharing the EMR documents within the organization, it will often be in paper form, thus moving back towards the traditional phase.

To overcome the challenges discussed, information systems, such as EMR, Patient Health Records (PHR), and Electronic Health Records (EHR) were evolved [4]. As a digital copy of a patient's health history, EHR is a solution to support healthcare facilities to improve patient care by enabling functions that other types of applications cannot deliver. As it makes the data movement easier and safer, EHR integrating all

levels of the health system helps to battle the problem of the movement of patient data during the point of care [5]. EHRs help in assisting healthcare professionals in making decisions that would, in turn, result in better care for the patient.

The Office of Health Information Technology [6] defines an EHR system as one that is "built to go beyond standard clinical data collected in a provider's office and can be inclusive of a border view of a patient's care". Apart from the benefits discussed, EHRs can do beyond those, and they are as follows:

- Contains patient's comprehensive medical history like diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results.
- It allows access to evidence-based tools that providers can use to make decisions about a patient's care.
- Workflow is streamlined and automated.

Implementing EHRs involves numerous potential challenges, and the following is the list of major hurdles found in the process of EHR implementation:

- [a] Cost of implementation: Access to financial resources for EHR implementation is one of the major hurdles, especially for small institutions. It will take away a more significant budget allocation in setting up the hardware, software, implementation assistance, training for the staff, ongoing network fees, and maintenance. In addition, there could be unplanned costs during the implementation.
- [b] Staff resistance: Not all medical professionals are open to using technology in a healthcare setting. In addition, some healthcare professionals question the performance of electronic health records. As a result, they may show reluctance to stop the writing process, or in some cases, employees do not have information about current technological advances and their full benefits. Consequently, it leads to delays in the implementation of EHR.
- [c] Training is time-consuming: Before submitting an EHR system, employees need complete training on new workflows. In addition, doctors and medical teams should spend some time and effort understanding the new system. It is a time-consuming process and a problem for both employees and managers. Small and medium-sized organizations fear business losses during the training phase. Also, employees may, at times, view it as unnecessary.
- [d] Data privacy: Another significant challenge for the use of EHR is the concern for the privacy of patient and provider community data. The risk of data leakage due to natural disasters or cyber-attacks often affects participants. Federal law sets out a national policy to protect the privacy of personal health data. In the context of a security breach, the organization may be embroiled in a legal dispute and will have to spend millions of rupees to resolve the dispute. Therefore, it becomes a major responsibility of the

\* Address correspondence to these authors at the Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia; E-mails: vinayakumarr77@gmail.com, gururaj.hl@manipal.edu

designer/implementer to ensure the privacy of EHR data.

- [e] Data migration: It is challenging for staff to export paper documents to digital records. There will be plenty of literature about the medical history of hundreds of patients, and data entry can be a daunting and time-consuming task for staff. Data migration challenges are doubled due to inappropriate formats in the existing/previous system.
- [f] Limitation of technical resources: This is one of the challenges of EHR implementation often faced by small healthcare facilities and private healthcare professionals when they have an internal technology team. In addition, they may not have the hardware required to equip an EHR solution. Moreover, building an in-house technical team and purchasing hardware is a considerable expense.
- [g] Interoperability: Interoperability refers to the ability of different EHR systems or software to exchange information so that different providers can use it. In EHR, interoperability is a necessity to get a complete picture of the patient's health. It remains a huge challenge for healthcare providers to build an interoperable system that enables the transfer of information among multiple providers. The possible challenge that developers can face in developing and implementing interoperability is having consistent standards since abundant technology standards are available to help bring uniformity to interoperable models. However, customizing and implementing these standards causes variations in use and negatively affects the seamless exchange of information through different systems. Moreover, patient records will be stored in various EHRs across care providers. Therefore, the challenge of locating and accessing records securely needs to be handled efficiently.
- [h] Decision support: The implementation of an EHR system can also help in decision support for continuous healthcare monitoring. The fine-tested support system is an added advantage to any EHR system. The integration of the decision support system to EHR is possible when plugins are provided with the EHR system. According to our knowledge, this type of module integration was not found in any of the works done while doing a survey.

Among the challenges discussed so far, we took the challenge of interoperability and decision support system, which is addressed in the proposed system. A detailed introduction to these will be discussed in the below section before moving to the literature survey.

### 1.1. Interoperability in EHR

The requirement of interoperability enables easy sharing and exchange of healthcare data between the various levels. Interoperability results in increased transparency, portability, and accessibility. Health records can be securely shared with patients and other clinicians for better and timelier decision-making, particularly in critical situations, and lead to better health outcomes. Accurate information, with interoperability

among the different stakeholders, such as physicians, practitioners, hospitals, pharmacies, urgent care centers, and specialized medical centers, allows different systems and applications to communicate. Patient medical records follow throughout the healthcare system, conserving the time and efforts of the staff. Moreover, there are several technical and administrative advantages involved in integrating an EHR solution with other systems [7]:

- Assists in safeguarding patient data
- Helps Researchers predict the disease and enables them to share the data more readily with researchers.
- Staff productivity can be improved in healthcare
- Data exchange with other systems can be improved
- Patient experience will reduce paperwork for medical practitioners and the associated staff to fill out the repetitive forms again.
- Patient do not need to explain their situation repeatedly, as the right person can see it in the follow-up or consultation in the EHR system.

Despite the advantages of EHR, there are various technical and administrative related challenges involved in integrating EHR with other systems such as [7]:

- [a] **Unique Patient Identifier:** The absence of a consistent way of uniquely identifying a patient leads to misidentification. However, calculating a unique patient identifier for each patient can resolve this issue.
- [b] **Absence of Common Standards:** Difficulty in copying and sharing information from one EHR to another due to different healthcare technology provided by vendors creates mismatched formats and data fields that will require manipulation and then isolation before it is fed into another system. Hence, a lack of common standards for sending, gathering, communicating, storing, and managing health data remains a greater hurdle to interoperability.
- [c] **Vendors Blocking Information:** Lack of data blocking and sharing data among healthcare is a major hurdle for interoperability from the vendor's side. However, mandatory regulation from the government will somewhat shut down the blocking of data.
- [d] **Lack of Analysis and Measurement:** It is difficult to measure the costs of error rates that occur when healthcare systems do not communicate with each other in a useful way. Due to the lack of ability to trace outcomes, health systems and other healthcare providers can boost their important processes, resulting in the lack of initiative to improve interoperability.

### 1.2. Decision Support System using EHR

Any decision support system helps in decision-making. The decision-making system in treating the patient will act like a secondary opinion system for experts and an advisory for young doctors. The DSS attached to the EHR system has comprehensive information about a patient. The DSS with EHR system assists juniors, trainees, physicians, and researchers in providing better health care, diagnosis, and

defining the appropriate health status. The EHR data from patients, laboratories, photography centers, and others can be accessible to practitioners anywhere, anytime, through any device. The whole process of identifying and analyzing information through the EHR system is required for any decision support system rather than going with the traditional method.

Data analysis within EHRs uses computer-based programs to provide instructions and reminders to assist in timely health care. However, this is possible when healthcare providers access complete and accurate information. EHRs can improve the ability to predict, diagnose and reduce or prevent medical errors, improving patient outcomes with decision support systems. The comprehensive picture from the decision support system helps providers diagnose patients' problems more quickly. A trained EHR not only maintains a record of the patient's medication or allergies but can also automatically help to diagnose complications whenever a new drug is prescribed and informs the clinician of potential conflicts that assist in monitoring the patient's health [8]. Apart from this, the EHR improves disaster management by providing clinical warnings and reminders, improving the integration, analysis, and communication of patient information, making it easier to consider all aspects of a patient's condition, providing built-in protections against prescribing adverse events, improving research and monitoring to improve clinical quality [9].

The Decision Support Systems (DSS) are often classified as informative or non-knowledgeable [10, 11]. In knowledge-based systems, rules (if-then statements) are created, and the system returns data to test the law and generates action or output [10]. Regulations can be made using documentary, performance-based, or patient-based evidence [12]. Non-knowledgeable DSS still requires a data source, but the decision improves Artificial Intelligence (AI), Machine Learning (ML), or mathematical pattern recognition rather than being designed to follow professional medical information [10].

In non-knowledgeable DSS uses, the rapidly growing AI includes many problems related to understanding the concept of how it uses AI to produce recommendations (black boxes) and problems with data availability [13]. AI, along with machine learning, allows computers to find patterns in clinical data or allow computers to learn from past experiences. This enables the elimination of input from the systems. Machine learning (ML) mainly relies on data-driven rules (inputs), which are taken from larger datasets rather than a specified programmer (human). ML applications could guide decisions, make predictions, and improve the experience as the computers learn on the basis of the training data. ML serves as an upfront aid to decision-making and supports healthcare professionals in object classification, detection, segmentation, and robotic-assisted surgery by matching individual patients' characteristics with computerized clinical knowledge. Pattern recognition is an analysis method that automatically uses machine learning to recognize patterns and regulations in data. When it comes to healthcare, pattern recognition can be applied for risk prediction, progress in disease deciphering, subtyping the patient *etc.*

DSS has yet to reach widespread implementation in all the categories. The scope of functions provided by DSS is vast, including diagnostics, alarm systems, disease management, prescription, drug control, and much more [14]. In addition, DSS can manifest as computerized alerts and reminders, guidelines, order sets, patient data reports, documentation templates, and clinical workflow tools [15].

DSS provides various functions and advantages:

- Patient safety: Reduces errors during medications or prescriptions and thus prevents adverse effects [16, 17].
- Clinical management: Compliance with the clinical guidelines [18] following the reminders and treatment notifications [19].
- Cost containment: Decreases the cost of duplicating tests, recommends cheaper treatment methods, and reduces the workload for providers as increased in the automated work practice [10, 20].
- Administrative function/automation: Choosing a diagnostic code is similar to prioritizing patients and ordering treatments and testing. The designed algorithms can help physicians to choose the appropriate one from the list of suggested diagnostic codes. Automated documentation and note auto-fill [20, 22].
- Patient-focused decision support: DSS is a great step toward patient-focused care, and DSS-supported EHR is the ideal tool to implement the decision-making between the individual and provider, especially because DSS can remove a 'lack of information as a barrier to a patient's participation in their own care [23]. DSS acts as a dashboard for the patients to view their portal and contact their physicians [24].

To conclude, most of the time, the technical challenges motivate the researcher and developer. The proposed EHR system for disease prediction is a second attempt to overcome the technical challenges of interoperability and disease prediction models. Previous to this work, a blockchain-based EHR system was proposed [25]. The system concentrated on adaptive EHR Management and secure distribution using Blockchain. The secured EHR management [25] was adopted in the proposed system and secure data management.

## 2. LITERATURE SURVEY

### 2.1 EHR-based Healthcare Information Systems

The advancement in technology and massive use of computers in healthcare sectors led to an organized way of storing health records electronically. The individual departments in hospitals started using computers for electronic storage, which led to the use of the Health Information System (HIS). EMR is one such HIS where health information transfer occurs between health service providers of the same organization for quick and easy sharing [16].

A major change from EMR to EHR happened in 2009 and was named Health Information Technology from the Economic and Clinical Health Act (HITECH). The purpose of this

HITECH was to motivate the implementation of EHR and to support IT in health for better care in the United States [24].

In 2001, Canada set up certain objectives to standardize procurement practices for a sustainable environment in order to achieve modernization in Communication and Information Technology. The field accelerated innovation scale across the digital health marketplace [25, 18]. The EHR benefited Canadians, and the majority of doctors and healthcare professionals started using the EHR system in their workplace for recording medical records. This resulted in the seamless transfer and exchange of patient data, thereby increasing patient safety and reducing errors, which increased the quality of medical care.

In Germany, the government created the health network system [27] to provide secure transfer of sensitive medical reports across the country. The setup works for patients' information to connect to this health network when needed. The patient can allow or deny access in Germany, except in an emergency. During an emergency, the doctors can access the portal directly with a security code given by the authority of the health organization of Germany [28]. We attempted a similar kind of giving access during an emergency for Indian infrastructure, which was successful [25]. By this, the government can hold command over the private sector to focus on the quality of service with reliable cost and make way for better care management.

In Korea [29], a Health Information Exchange (HIE) platform is built, where the tertiary hospitals exchange patient data using EHR. As this platform is connected to programming interfaces, they have various features, including registration to document, repository, and an index for the patient for identification.

From the discussion conducted by Bahga and Madiseti [30], the United States' original health record, Veterans Health Information Systems and Technology Architecture (VistA), is not a single application but contains 168 packages integrated at the database level by the VHA Office of Information (OI). This OI provides nationwide technical support, resources, and coordination, enabling VistA to operate as a unified system. VistA acts as a standalone system, allowing only retention and self-contained management of healthcare data within an institution. They also discussed the limitations of this application, depicting that this was using an unstructured way to record patient data, leading to a bulky database with many tables, as each report and diagnosis and demographics were maintained in a table separately for each entry. They also projected that this system uses only a national standard dictionary with its own common data word, programming conventions, and database standards, which caused an issue of interoperability in common method while sharing data with other hospitals, resulting in conflicts in traditional EHR about the problem related to integration and interoperability due to usage of various different standards and regulations. They also mentioned different technology languages led to the problem of transferring patient reports to each hospital. Hence, cloud models were made in the United Nations, where they integrated health services to overcome this.

Indian Healthcare also uses IT applications but are limited to only a few particular private hospitals like Max Hospital [31], Apollo Hospital [32], Sankara Nethralaya [33], *etc.* In India, electronic records are available only at the All India Institute of Medical Sciences (AIIMS) and the Postgraduate Institute of Medical Education and Research (PGIMER). However, all this healthcare has only EMR to transfer patient data within their hospitals and groups [34]. One of the limitations of the EMR is that the patient can only get a print copy of medical data for further treatment at other hospitals. So, to attain the need of the hour, EHR can be implemented and integrated at all levels of health systems, making the patient data flow easier and safer during treatment. In this process, secure sharing and management of interoperable records are much needed in India.

## 2.2. EHR-based Disease Detection System

Many modern technological devices are used to monitor patient health, focus on medication, and maintain health support. Along the same lines, EHR outside India has altered patient-doctor interactions through visual consultation and telemedicine [35]. However, the EHR will be connected to various machines used in patient monitoring and capture data using visual interface applications [36]. From the collected data, a certain level of health issues can be drawn with the help of Clinical Decision Support (CDS). These programs have been initiated as an integrated development to tone and accelerate the CDS [37].

The data from EHR is used for various research and to monitor the patient and the quality of medical care [38]. The large amount of information calculated from EHRs generates extra value if compiled and stored in a business repository [39]. This information is not only used for managing strategic decisions but also for advanced data analysis and health care.

An effective automatic diagnosis study using machine learning models [40] demonstrated an effective automatic diagnosis using machine learning models. The work discussed coronavirus, diabetes, and heart disease from the data entered into an Android application. The data were analyzed and performed on a real-time website using a pre-trained model. Similar to this work, machine learning was used in the diagnosis of human diseases [41] on the data tools used in the medical field. In addition, it also provides information about each method depicted with their advantages and disadvantages to assist the professionals and developers in future implementations for a real-life situation. Finally, it shows the relationship between all strategies and all supervised and unregulated learning algorithms, stating non-labeled and no notion of output along the learning process. Some of the unsupervised learning tasks are found to be like K-means clustering, Density-based Spatial Clustering of Application with Noise (DBSCAN), Self-Organized Maps (SOMS), Similarity Network fusion (SNF), Perturbation Clustering for Data Integration and disease Subtyping (PINS) algorithms. Supervised learning is stated as a labeled set of training data used to calculate or map the desired output. The same type was found [3] for Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Naïve Bayes, Decision trees, Random

Forest, linear regression, and logistic regression. One of the works [42] showed that KNN for 10 neighbors provided an accuracy of about 86% for the training dataset and about 77% for testing data. The work also confirmed that the highest accuracy for the decision tree was about 98% for diabetic prediction for 8 features (age, pedigree function, glucose, BMI, insulin, skin thickness, blood pressure, and pregnancies).

The research work on diabetes prediction was done using 3 ML models and SVM, Naive Bayes (NB), and Random Forest (RF) algorithms. The dataset for research was from a PIMA hospital. It was found that the prediction was based on the patient's preconditioning. The work discussed the application of RF, NB, and SVM models. From the article, it was learned that RF achieved an accuracy of 80.7%, and NB showed 77% and 77.3% for SVM. Apart from the model adaptability, the article showed a way for data balancing techniques like SMOTE, under-sampling, oversampling, *etc.* Another similar work was found [43] for the early prediction of diabetes for improved treatment. In that research paper, significant attribute selection was done *via* the principal component analysis method. In the work, an Artificial neural network (ANN), random forest (RF), and K-means clustering techniques were implemented for the prediction of diabetes, and it was found that the ANN technique provided the best accuracy of about 75.7%. A work on prediction for type 2 diabetes and hypertension based on individual risk factors data [42] was also studied. The study showed that the isolation forest (iForest) based outlier detection would remove outlier data, and the synthetic minority oversampling technique Tomek link (SMOTETomek) would balance data distribution and ensemble approach [44] to predict the diseases. In the work, they considered four datasets to build the model and extracted the most significant risk factors.

A work adopted Logistic regression (LR) to identify the risk factors for diabetes disease based on p-value and odds ratio

(OR). In this work, the four classifiers like, naïve Bayes (NB), decision tree (DT), Adaboost (AB), and random forest (RF), were used to predict the diabetic patients. A repeated trial of 20 was executed on three types of partition protocols (K2, K5, and K10). However, they claimed that the combination of LR-based feature selection and RF-based classifier gives 94.25% accuracy and 0.95% of the area under the curve (ACC) for the K10 protocol.

To conclude the survey, it was learned that machine learning algorithms are well suited for any disease prediction. The prediction systems can also help in decision-making systems in many healthcare information systems. From the study, we are motivated to design a web application for data collection from the EHR system and train the proposed system for detecting diabetic disease from the health data collected from the EHR system.

### 3. MATERIALS AND METHODS

#### 3.1. Proposed System

All users of health information systems are gathered under the proposed EHR system to collect and share health data. The main aim is to make patient care more efficient and consistent. The proposed system enables all levels of healthcare professionals to contribute to the content of EHRs. Different care centers, such as primary, secondary, and tertiary care levels, can share their resources. The data collected from each and every healthcare center is stored and processed in the EHR provider database called the global database. At the moment of care, healthcare providers can access patients' data securely through the EHR web application designed for the proposed system. To implement an integrated system, health-related data from several modules, such as physician, administrative, nursing, radiology, laboratory, and pharmacy are integrated securely. The different components of the proposed EHR modules are shown in Fig. (1).

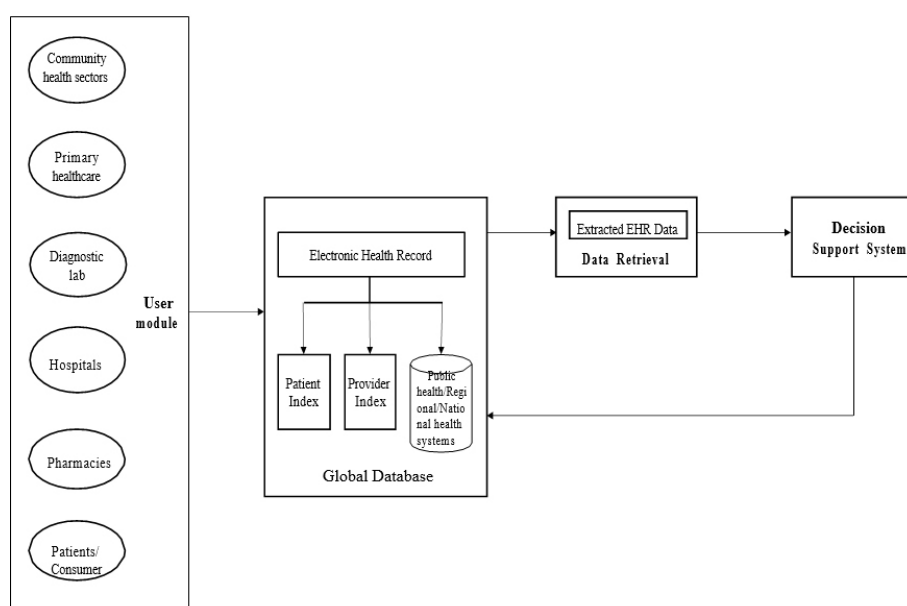


Fig. (1). Block diagram of Proposed EHR System for disease prediction.

Fig. (1) shows the block diagram of the proposed EHR system. This proposed system consists of four modules: User module, Storage module, Data retrieval, and Decision Support System (DSS) module. The user module consists of data entry from various health sectors like community health workers, primary healthcare, diagnostic labs, hospital pharmacies, and patients or consumers. In the EHR storage module, the data is

stored in the database in which the data is recorded, and from EHR, the necessary data is extracted in the Data extraction module and pre-processed. The data which is pre-processed is sent to the decision support system. The decision support system in the proposed EHR system is shown in Fig. (2). The decision support system is designed and developed for the early detection of diseases.

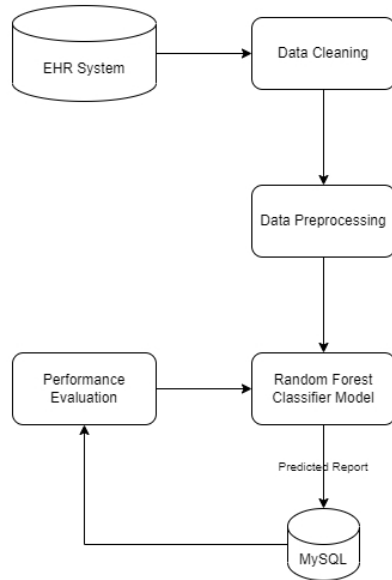


Fig. (2). Structure of decision support system for her.

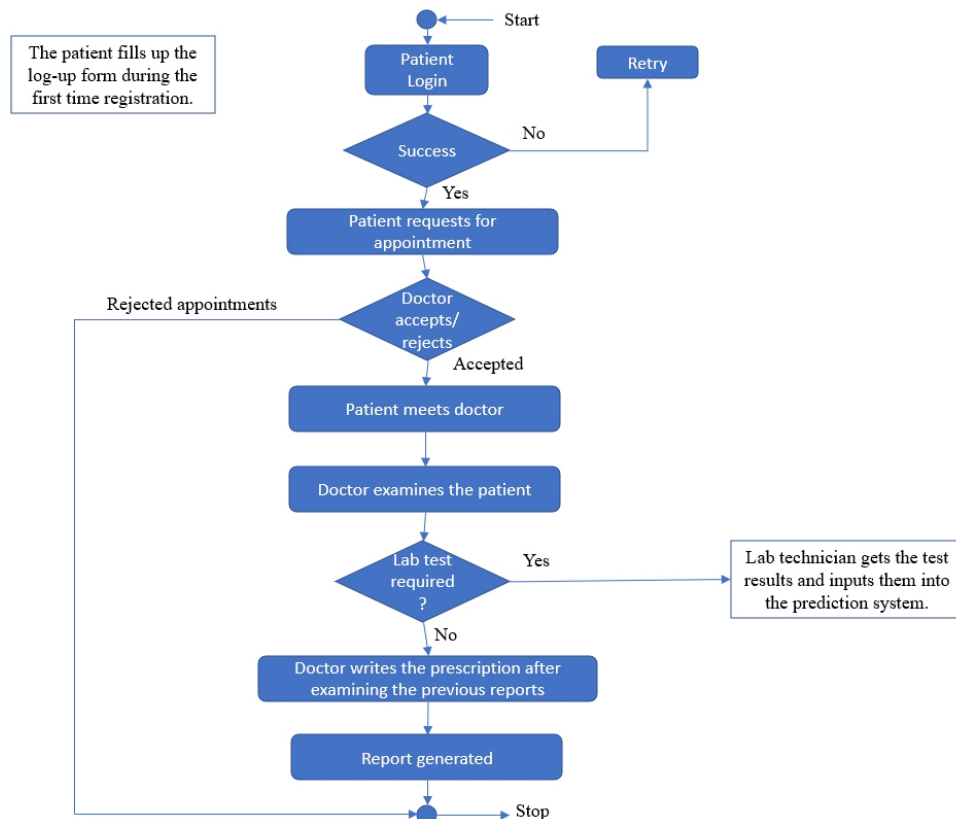


Fig. (3). Working model of the proposed EHR module with disease prediction.

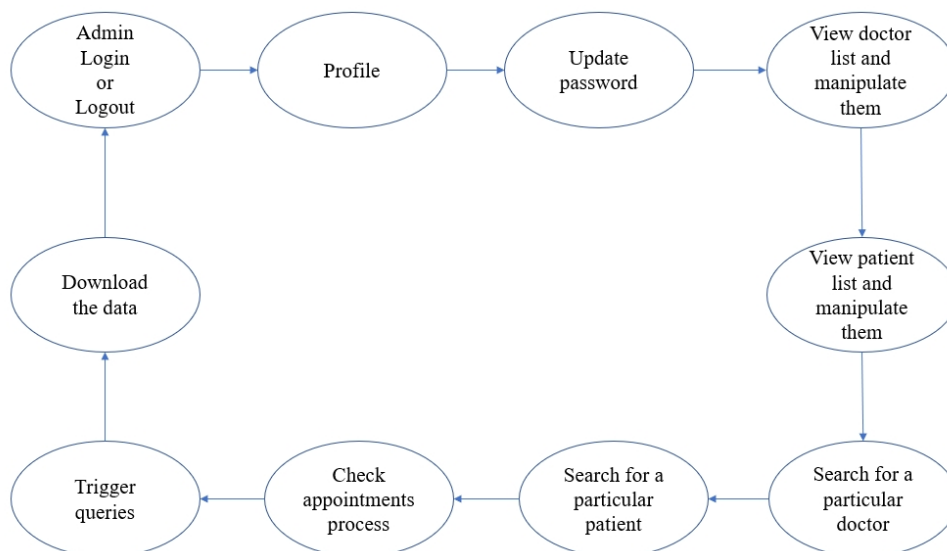


Fig. (4). Workflow of admin in EHR systems.

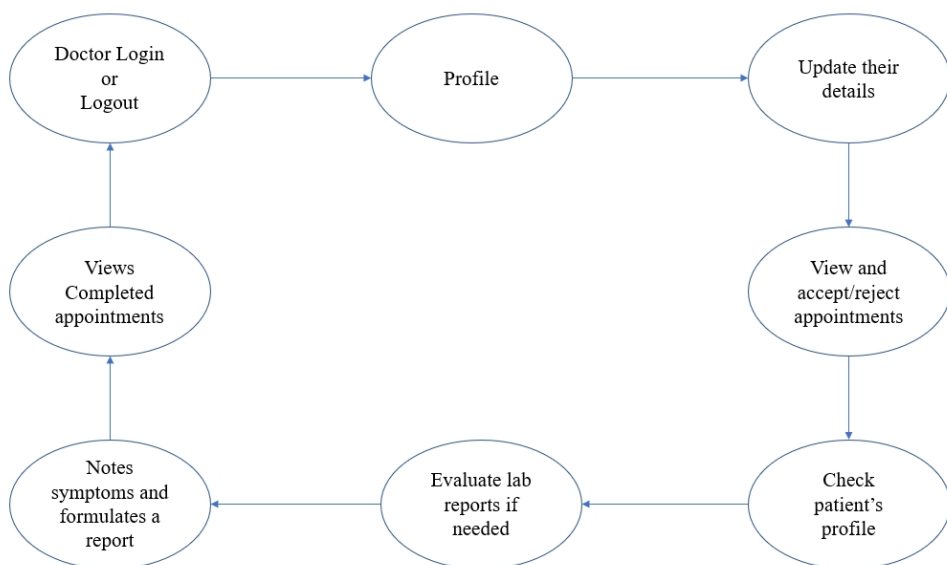


Fig. (5). Workflow of doctors in EHR systems.

3.1.1. System Workflow

The user module consists of three domain factors, such as administrator, doctor, and patient. Combining all three domains, the workflow of the proposed system is shown in Fig. (3). The workflow of each domain factor is identified and defined in Figs. (4 - 6). Fig. (4) shows the admin credentials and workflow for the EHR system. In a similar way, the workflow of the doctor and patient is designed as per Figs. (5 and 6), respectively.

3.1.2. Methodology

The proposed proof of concept is understood, and a functional product is designed. It is a fully functional product implemented with the help of a user interface (UI). This system is constructed in such a way that the admin can monitor and authorize any information, starting from adding/deleting of

patient/ doctor’s profile to canceling an appointment. The DSS shown in Fig. (3) is designed to predict the diseases through the patient’s profile when the required clinical data is obtained from the lab. In the implementation of the front end, PHP is used, and Python is used for the backend design. PHP scripts are embedded in HTML files to integrate into a dynamic front-end app. The major part of the implementation of the backend is achieved in PHP as it is dynamically utilized in the elements of PHP scripts. However, to use functions, Python libraries and frameworks are used with a microweb framework written in Python. No special tools or libraries are needed, but “Ngrok” is used to make a locally hosted web server look like it is hosted on a subdomain. Ngrok is a cross-platform application that allows publishing locally developed servers to the Internet with minimal effort, and there is no need for public Internet protocol.



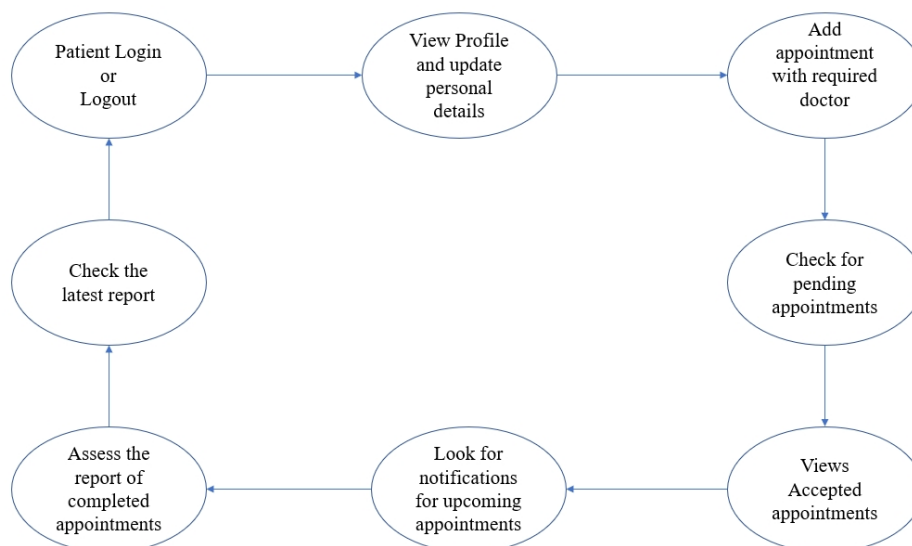


Fig. (6). Workflow of the patient in the EHR system.

MySQL is used to perform any kind of database operations and helps with application development within the database, support for triggers, stored procedures, functions, cursors, views, and ANSI standards. SQL is used to provide data security and strong transaction support to protect online transactions and improve user interaction. The interface that is built consists of around 13 tables, which come into play when integrated into the EHR system. In every table, the system takes several features as input and self-generated tables. Each table is normalized by the third normal form (3NF), which uses the principles of normalization for relational databases to ensure referential integrity, avoid data anomalies, simplify data management, and reduce data duplication.

3.1.3. Prediction System

Under the decision support system, the prediction of diseases has been the objective of the proposed system. The main motto is to predict diseases with the test results that lab technicians obtain for various tests that are written down by a doctor by assessing the patient’s health condition when he/she consults the doctor. Random Forest algorithm is utilized in the proposed system, and it uses decision trees for creating tree-like structures. A decision tree is chosen because it is fast, reliable, and easy to interpret, and very little data preparation is required. The Gini index is calculated for each attribute starting from the root. While designing the decision tree, the attributes possessing the lowest value of the Gini Index would be preferred, as stated in Eq. 1.

$$Gini\ index = 1 - \sum^n (P_i)^2 \tag{3.1}$$

i=1

Where  $P_i$  is the probability of an object being classified to a particular class.

Features that are taken to predict various diseases at early stages are chosen from standard test reports from certified hospitals. These features that are considered are mostly float and integer-based values, which are obtained once the test is

performed by a lab technician for a particular patient. Prediction is done multiple times for a particular patient, and every test report is reflected through the user interfaces of both patient and doctor profiles.

4. IMPLEMENTATION AND DESIGN OF THE PROPOSED SYSTEM

4.1. Design Issues of EHR Design

Before storing electronic health records, these records are to be precisely extracted, accumulated, and properly integrated. This is because when a doctor or an admin tries to access a patient’s or doctor’s profile, it needs to be stored well enough so that the system can fetch the required information within no time. In the process of EHR design, data extraction, data accumulation, and data integration play a major role. Before proceeding to the prediction system design, a brief discussion on the data extraction, data accumulation, and integration is done, as described in the following sections.

4.1.1. Data Extraction

The main responsibility of healthcare professionals is when a patient visits the hospital for the first time, they need to collect the demographic details of the patient accurately. To address the unique needs of patients, the hospital staff must collect demographic information consistently. Generally, the details are present in a patient demographic section, such as name, email, contact number, address, marital status, gender, occupation, previous diagnosis, location, emergency contact, and many more. The proposed EHR captures the patient’s past medical history and medications, such as his/her drug history, which the patient follows on a daily basis or occasionally before getting admission to the hospital. It also provides patient’s clinical encounters, which in turn enables them to focus on patient-centric medical care. Information about clinical encounters includes diagnosis, observations, physical examination, demographic details, type of encounter (interaction between the patient and doctors), lab investigation, and patient history.

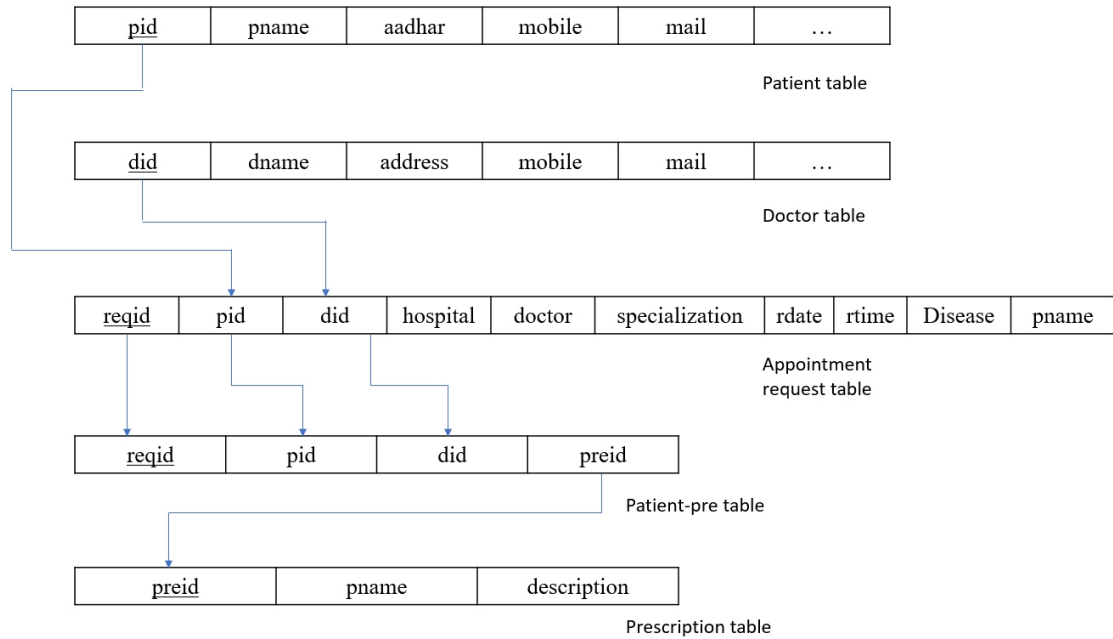


Fig. (7). Relational schema of the proposed EHR model.

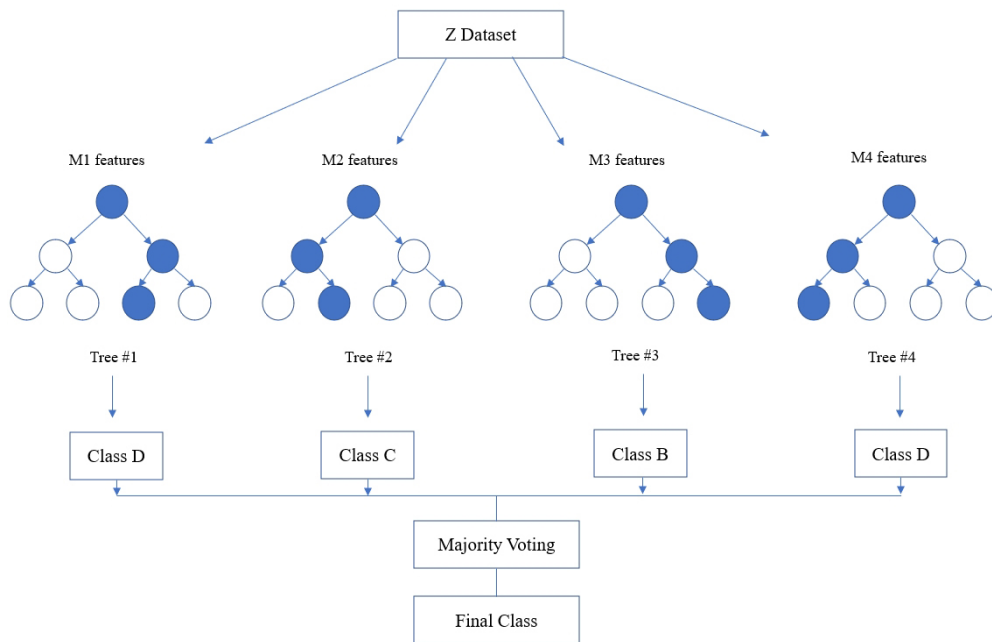


Fig. (8). Random forest classifier.

4.1.2. Data Accumulation

In the proposed system, normalization is used to accumulate the data in the database. These details establish relationships between those tables in accordance with the rules, which are in normal forms. The third normal form is achieved by making sure that the application can handle problems, such as redundancy and inconsistent dependencies, making the database flexible and secure. To avoid null values in the attributes, normal values are set to the corresponding attributes, which helps the decision support system to provide quality care without any major mistakes or confusion.

4.1.3. Data Integration

The main idea behind data integration is that without duplicating or moving the data, one can easily access the data and validate the data that a person is using. In the proposed system, data accessibility is quite simple. As the concept of the primary and foreign keys, the link between different tables becomes easier. Fig. (8) shows the relational schema between the tables for better understanding.

From Fig. (7), it is shown in the patient table that *pid* is the primary key, which is a foreign key in appointment requests and patient prescription tables. Coming to *did*, this is a primary

key to the doctor’s table and a foreign key to appointment request and patient prescription tables. There is less chance of generating null values because all the tables are normalized to the third normal form (3NF). This way, the data is integrated, and any information can be easily accessed with the help of foreign key references in other tables.

**4.2. Decision Support System using Ensemble Learning**

Random forest algorithm is used in the proposed system as the algorithm is strong enough to solve both classification and regression problems. Random Forest algorithm is also used due to the non-linear data nature and supervised learning method. This algorithm forms a tree forest by combining several trees. Similarly, decision trees are combined in the Random Forest classifier for greater accuracy for ensemble learning. A clear picture of the classifier working is shown in Fig. (8), and the outstanding performance of Random Forest is shown in Fig. (15).

In the random forest classifier, the correlation values are kept low to achieve superior accuracy rather than individual predictions. Even though a few tree predictions may go wrong, most of the predictions are correct and improve the overall accuracy of the model. Instead of relying on one single tree, this classifier accumulates predictions from each and every tree and predicts the final result by means of the majority vote that comes from predictions. Accuracy is found to be more when more decision trees are considered, and overfitting problems can be avoided.

In the prediction space, the system is trained and tested for diabetic prediction. The feature selections are basically from the prescriptions and test reports collected from the doctor. The accumulated and pre-processed diabetic data features for the proposed system are shown in Fig. (9).

During implementation, the dataset from Kaggle includes 1000 data and create a dataset of about 80 from the campus health center. In the process of prediction, the data in the prepared dataset is split into test and train datasets with eight proper features that are taken as input and one that is set as a target variable. After which, the model is imported using standard “sklearn” libraries, and it fits with the required number of estimators for the number of decision trees. As discussed earlier, each and every decision tree vote is considered before classifying the data based on the threshold value for each feature. The threshold is fixed from the precision-recall curve among the values of each feature. If more features do not cross the threshold value, the model

predicts that the patient’s condition is normal. If not, there is a majority of them whose values cross the threshold set. Then, the model predicts that the patient’s condition is abnormal. The implementation of the system using the Random Forest Algorithm is as follows:

1. Data pre-processing step: The identification of less contributing features is done through a comparison chart and updated as null. As this step removes negative and null values, the dataset, *df*, uses Eqs. (2 and 3), respectively.

$$[df < 0] = 0 \tag{2}$$

$$df.isnull(), df.fillna(), df.()... \tag{3}$$

2. Fitting of the algorithm to the training set: The Random Forest Classifier module is imported from the standard library “sklearn,” and the model is fitted using Eq. 4. From “sklearn”, an ensemble imported Random Forest Classifier model is designed.

$$model = RandomForestClassifier(n_estimators = 20) \tag{4}$$

$$model.(x_train, y_train)$$

3. Predict the test result by considering a test size of 20%. The selection of 20% is validated by adopting the k-fold method. Initially, the system is tested for 50% and then repeated *k* times by 5% incrementing.

4. Testing the accuracy of the result: To understand the accuracy of the model numerically, a Confusion matrix is created and studied.

5. Visualizing the test set result: For visualizing the test results, “matplotlib” is used. Using “matplotlib” library, graphs are plotted, and a clear representation of data is seen for analysis. The graphs for the same are shown in below section and discussed in the relative context.

**5. RESULTS AND DISCUSSION**

The proposed model consists of two modules: the EHR module and the decision support module. The testing was done to verify the working nature of both modules; however, while testing the operation of the proposed system, the real-time scenarios were also considered with respect to the patient mindset. In the first scenario, patients were assumed to visit the same hospital/physician/health caretaker; in the second scenario, the patient was assumed to change the hospital/physician/health caretaker. It is assumed that all the healthcare data is updated in the EHR when the user visits for the health check.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	2	138	62	35	0	33.6	0.127	47	1
1	0	84	82	31	125	38.2	0.233	23	0
2	0	145	0	0	0	44.2	0.630	31	1
3	0	135	68	42	250	42.3	0.365	24	1
4	1	139	62	41	480	40.7	0.536	21	0

Fig. (9). Features considered for diabetic prediction.

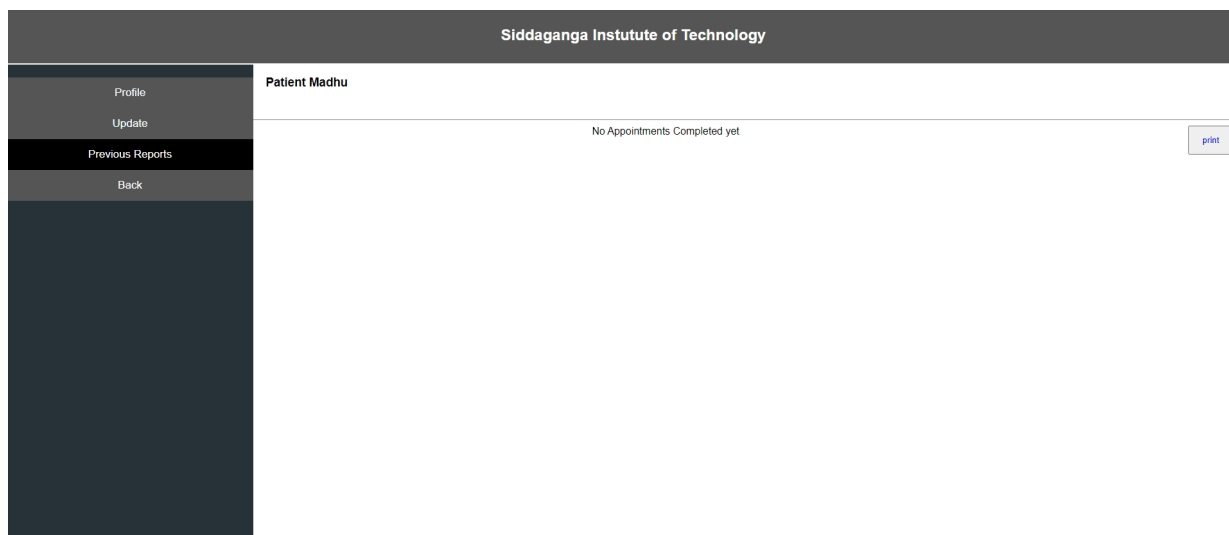


Fig. (10). EHR system deployment at the first hospital – Victoria hospital.

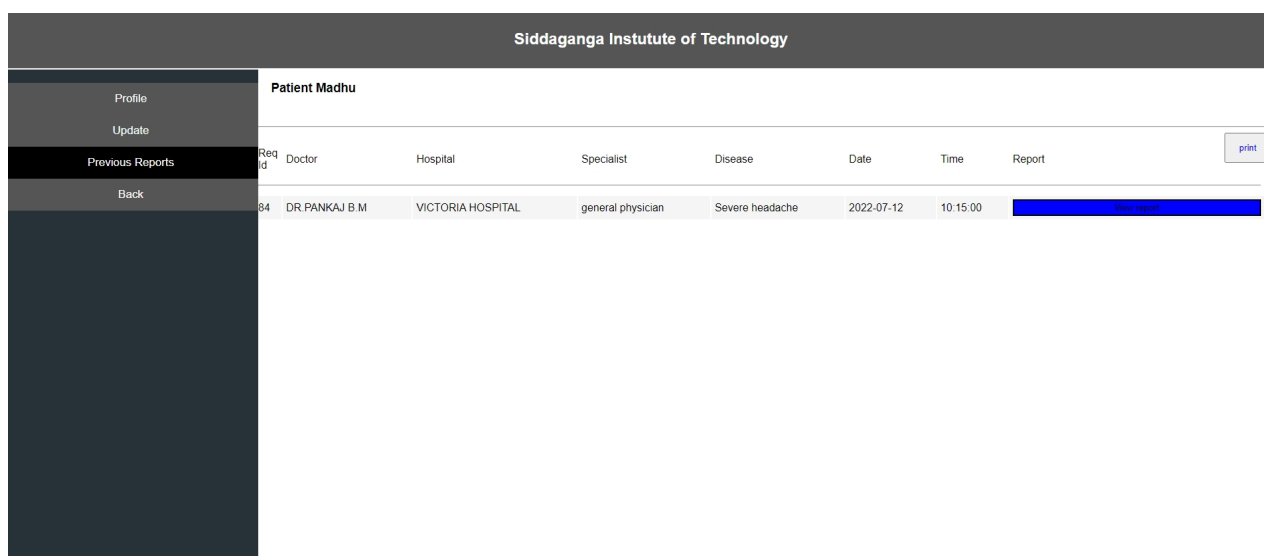


Fig. (11). EHR system deployment at the second hospital – Modern hospital.

### 5.1. Interoperability of the EHR System

Interoperability in healthcare is an exchange of clinical information and electronic data sharing among healthcare systems. To introduce interoperability context, let us assume a person goes to a hospital ‘X’ for diagnosis. A doctor assesses the patient and gives the report. Now, the same person goes to hospital ‘Y’ for a general health check-up. Here, the doctor of hospital ‘Y’ should be able to access the previous health reports of the person who had visited hospital ‘X’ before. In the last visit, the doctor is capable of extracting all the health-related data before proceeding further. The working of interoperability is confirmed when this scenario is successful. To achieve interoperability, a unique ID is considered to play an important role in accessing previous reports without any intermediary or human physical intervention. The

implementation of the interoperable nature of the proposed system can be seen in Figs. (10 - 12). The user view is updated without any errors and validated with real-world data.

From Fig. (12), it can be seen that the consistency of the data is maintained, and the authentication methodologies while sharing the data among many peer EHR system validates the security aspect of the proposed system. The EHR system's scalability and adaptability nature are verified by setting various instances from the previous work [42] with different operating systems and processors. Hence, the performance was as expected and is shown in Fig. (13). The lower processors took more time compared to higher processors. A similar type of experiment was conducted on the different operating systems. The behavior is shown in Fig. (14). The EHR system behaved similarly in Windows, Ubuntu and IOS machines.

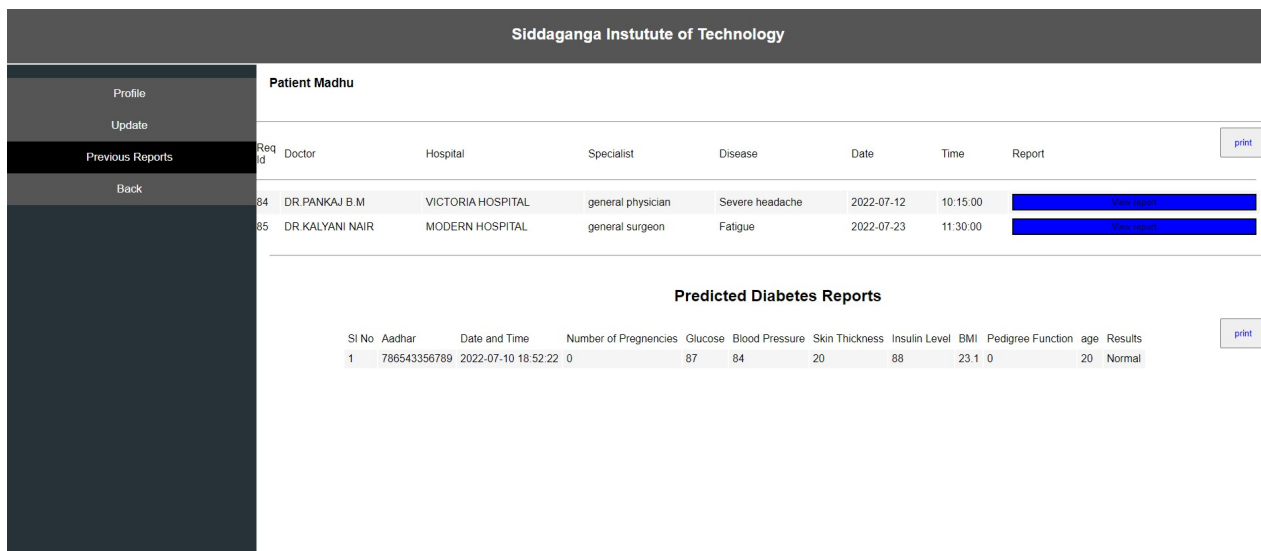


Fig. (12). EHR system deployment at the third hospital – Royal hospital.

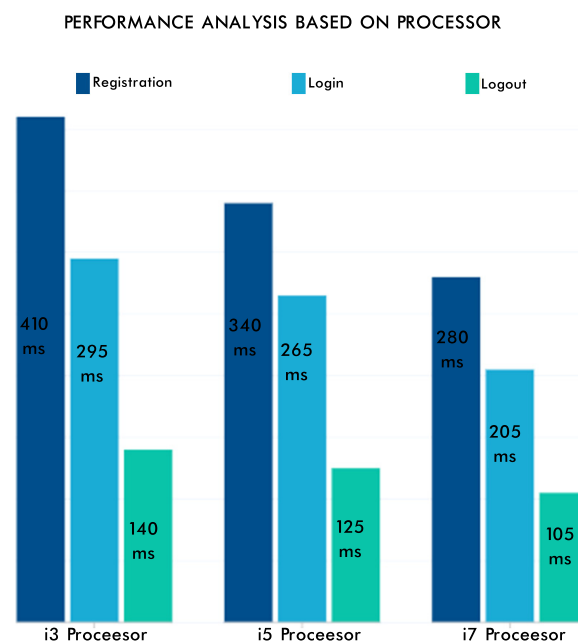


Fig. (13). EHR system performance based on different processors.

### 5. 2 Early Prediction of Disease using the Decision Support System

Once the interoperable EHR working was confirmed, the decision support system was substantiated with the Naive Bayes classifier and the KNN algorithm for the same dataset considered for training and testing. As discussed earlier, the data set originated from Kaggel [45]. After the implementation of Naïve and KNN, an accuracy level of about 76% was achieved through Naïve Bayes and 77% with the KNN algorithm.

However, the main focus was on the Random Forest Algorithm, and the experiment was repeated for the same dataset. After experimentation, an approximate accuracy of 97% was seen in predicting the patient’s condition. A comparative analysis of the same is shown in Fig. (15). After

the model was trained, the efficiency was maintained, and the model was reused by increasing the number of estimators. It was observed that as the number of decision trees is increased, the model gets more accurate prediction/decision [46 - 48].

For interpreting the performance measure and properly contracting an idea about how well the system can predict a particular measure with the features fed as the input, the confusion matrix was the one solution. The confusion matrix is a table that is frequently used to explain how well a classification model performs on a set of test data for which the true values are known. Table 1 shows the confusion matrix that looks for the proposed classification algorithm. In Table 1, the anticipated observations are displayed in green (*i.e.*, true positives and true negatives), and the false positives and false negatives are shown in red, and this needs to be reduced for greater accuracy scores.

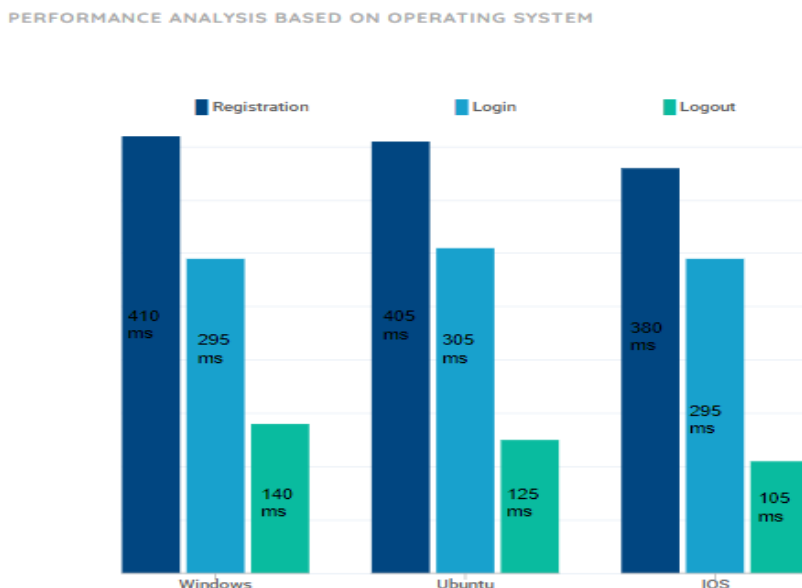


Fig. (14). EHR system deployment performance based on different operating systems.

### Comparative analysis of ML algorithms

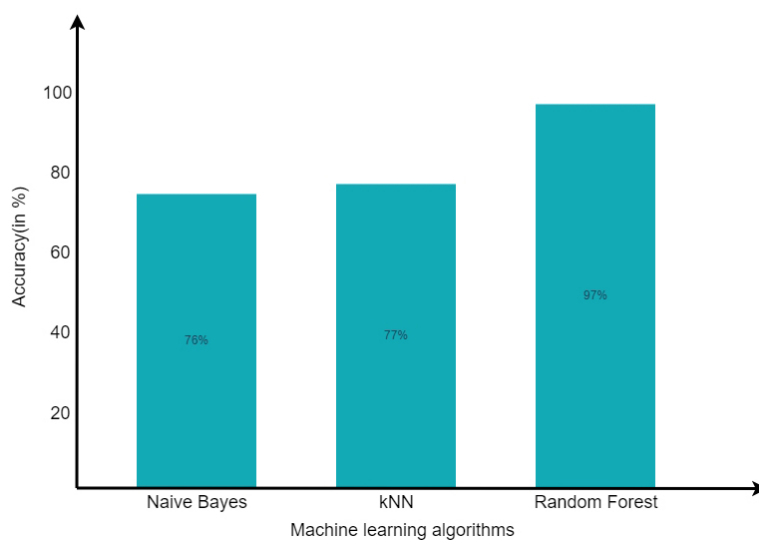


Fig. (15). Comparison between 3 ML algorithms.

Table 1. A confusion matrix.

Actual class	Predicted Class		
	-	Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
Class = No	False Negative	True Positive	

Here, the True Positive (TP) is the successfully predicted positive value, implying that both the actual and expected class values are true. The True Negatives (TN) are the accurately predicted negative values, indicating that neither the actual class value nor the projected class value is positive. The False Positives (FP) are the values when the anticipated class is true, but the actual class is false. The False negatives (FN) are the

values when the anticipated class is false, but the actual class is true. From these parameters, the accuracy, precision, recall, and F1 score for the proposed system were evaluated. When the test across each parameter on the split datasets was carried out, varied levels/scores in each of the parameters were found. The reason for this may be due to a number of elements present in different age groups. Hence, after the confirmation of the well-behaving of the algorithm, an analysis was done with respect to the age groups present in the standard dataset. The dataset had entries from the age of 21 to 70 years. However, for accurate analysis, an age group of 21 to 30, 31 to 40, 41 to 50, 51 to 60, and 61 to 70 was performed.

$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \tag{5.1}$$

Eq. 5.1 is used to evaluate the accuracy score, where it is the most instinctive performance measure, and it is the ratio of truly predicted observations to the total observations. A bar graph plotted in Fig. (15) shows the accuracy levels with respect to the different age groups. From Fig. (16), it is obvious that the accuracy level was almost similar to all the age groups. This shows that the prediction was accurate and performed similarly in all age groups. However, more accurate predictions were found for the age group of 51 to 60.

Precision is the proportion of accurately anticipated positive observations to all of the predicted positive observations. The precision score is evaluated using Eq. (5. 2). The precision scores with respect to different age groups are

shown in Fig. (17). From the observations, it was found that the age group of 51 to 60 precision score is zero. This score is due to the more false positives (FP) numbers than the number of true negatives (TN) and false negatives (FN) obtained at the time of analysis.

$$Precision = TP/(TP+FP) \tag{5.2}$$

A recall is defined as the proportion of accurately predicted positive observations to all of the actual class observations. The recall is evaluated from Eq. 5.3. This parameter is similar to the precision score with respect to the values obtained for each group. The recall parameter analysis is well understood from Fig. (18).

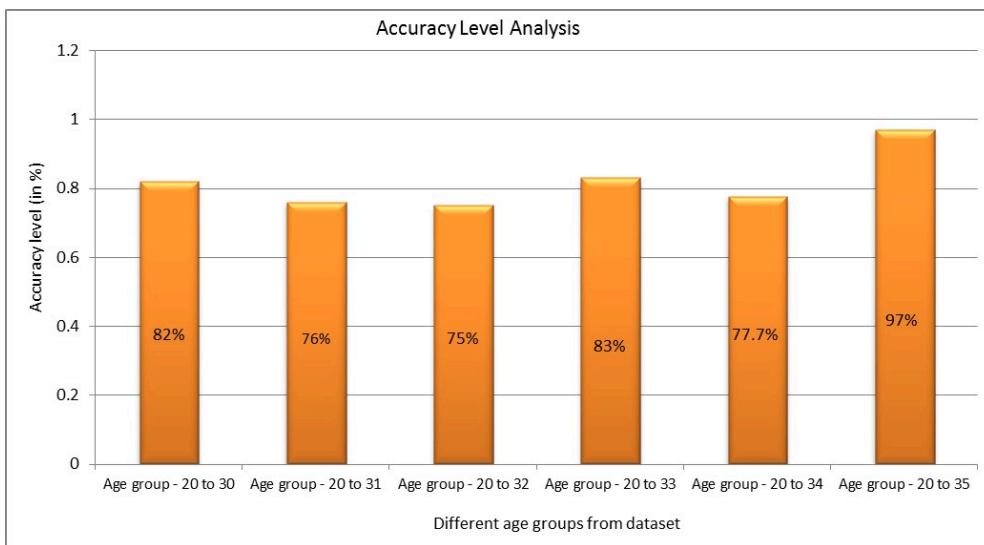


Fig. (16). A graph depicting the accuracy levels.

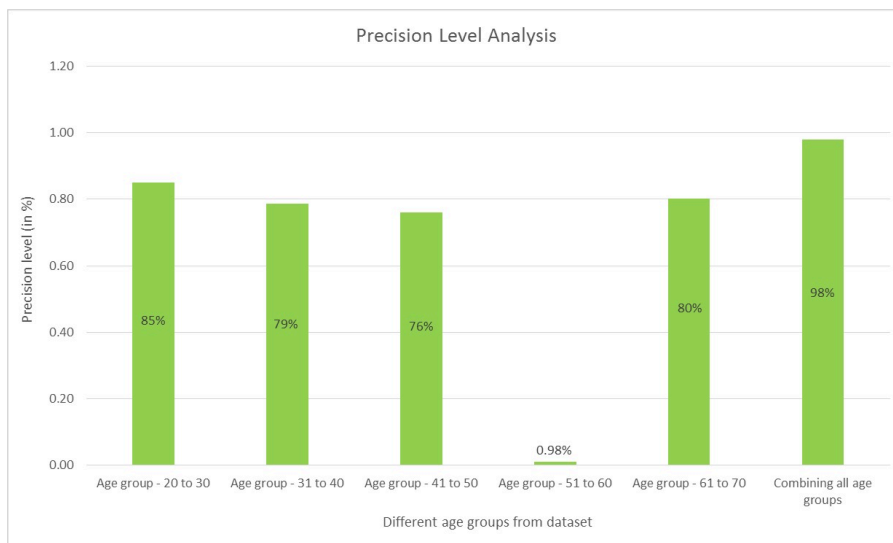


Fig. (17). Precision score analysis.

### Recall score analysis

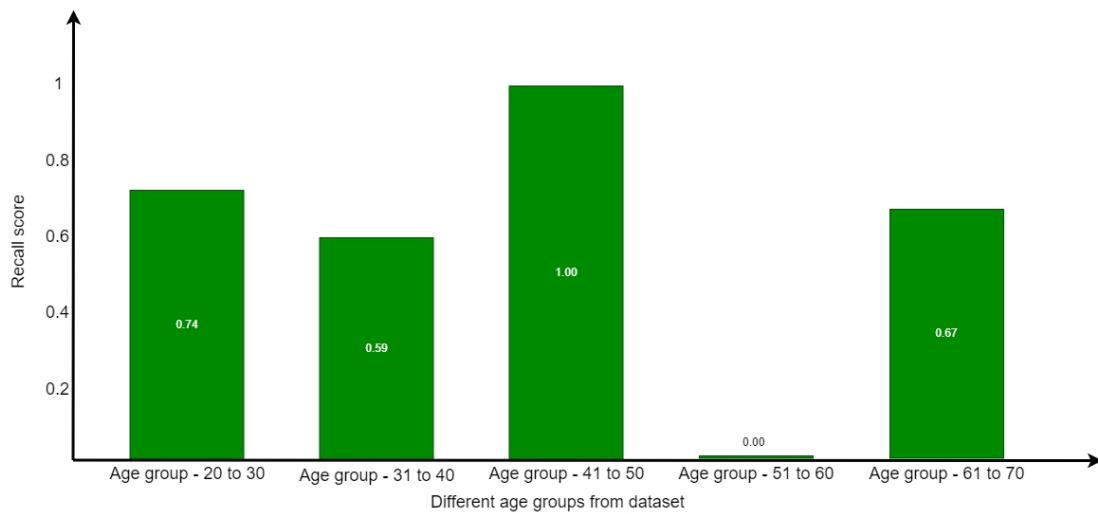


Fig. (18). Recall score analysis.

### F1-score analysis

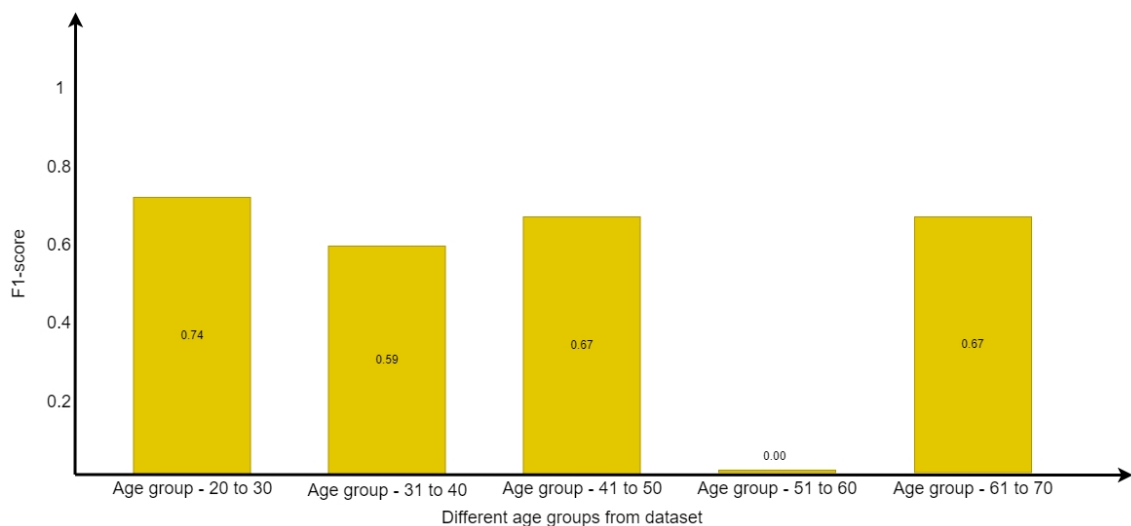


Fig. (19). F1-Score analysis.

$$Recall = (TP)/(TP+FN) \tag{5.3}$$

In a similar way, the F1 score was evaluated using Eq. (5.4), which is the weighted average of recall and precision. This parameter is useful when an uneven class distribution is present in the dataset for prediction. F1 score takes both false negatives and false positives into account, even when they are not of a similar cost. The graph in Fig. (19) depicts how the F1 score parameter results differ from one another. However, precision and recall scores are almost zero for the age group of 51 to 60 and move towards infinity. Since the value cannot be undefined, the value is considered as zero.

$$F1 = 2 * (Precision*Recall/Precision+Recall) \tag{5.4}$$

All the parameters considered for the analysis gave the

expected score. From the analysis, it is obvious that the prediction system developed gives the expected results. The prediction system developed through DSS also provided appropriate results for diabetic data. These results have shown a path to predict more diseases from the EHR data.

### CONCLUSION

A user-friendly interoperable EHR system with early prediction of diseases using the web application was designed and tested for its working. The EHR system behaved equivalently in all the different types of processors and operating systems. The usage of the system window provided upright sharing of the EHR data to all the valid users among doctors and patients. However, the developed EHR system did not provide only timely data for treatment; it also stood distinct with the adoption of the prediction system. The adoption of



Machine learning algorithms for early prediction of diabetes using Naïve- Bayes, KNN, and the random forest algorithm was acceptable. From the ensemble learning, the EHR system is innovative and efficient in predicting diseases using EHR data. The model reflects the result of the prediction in the EHR and is cross-verified according to the patient's index for future manual consultation. In the future, the system will be verified with the various disease predictions and its efficiency.

#### LIST OF ABBREVIATIONS

<b>ML</b>	=	Machine Learning
<b>EHR</b>	=	Electronic Health Record
<b>PHC</b>	=	Primary Health Care Centre
<b>SHC</b>	=	Secondary Health Centre
<b>THC</b>	=	Ternary Healthcare Centre
<b>EMR</b>	=	Electronic Medical Record
<b>DSS</b>	=	Decision Support Systems
<b>AI</b>	=	Artificial Intelligence
<b>HIS</b>	=	Health Information System

#### ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

#### HUMAN AND ANIMAL RIGHTS

No humans were used for studies that are the basis of this research.

#### CONSENT FOR PUBLICATION

Not applicable.

#### AVAILABILITY OF DATA AND MATERIALS

The data and supportive information are available within the article.

#### FUNDING

None.

#### CONFLICT OF INTEREST

Dr. Vinayakumar Ravi is the Associate Editorial Board Member of journal The Open Bioinformatics Journal.

#### ACKNOWLEDGEMENTS

Declared none.

#### REFERENCES

- [1] Rozenblum Ronen, Jang Yeona, Zimlichman Eyal, *et al.* A qualitative study of canada's experience with the implementation of electronic health information technology. *Can Med Assoc J* 2011; vol.183(5): E281-8.
- [2] Cabana M D, Rand C S, Powe N R, *et al.* Why don't physicians follow clinical practice guidelines? a framework for improvement. *JAMA* 1999; 282(15): 1458-65. [http://dx.doi.org/10.1203/00006450-199904020-00719]
- [3] James R. The American College of surgeons, minimum standards for hospitals, and the provision of high-quality laboratory services. *Arch Pathol Lab Med* 2017; 141(5): 704-17.
- [4] Gold M. Assessing hitech implementation and lessons: 5 years later. *J Milbank Q* 2016; 94(3): 654-87.
- [5] Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing paper-based with electronic patient records: Lessons learned during a study on diagnosis and procedure codes. *J Am Med Inform Assoc* 2003; 10(5): 470-7. [http://dx.doi.org/10.1197/jamia.M1290]
- [6] Tang PC, McDonald CJ. Electronic health record systems. In: Shortliffe EH, Cimino JJ, Eds. *Biomedical Informatics Health Informatics*. New York, NY: Springer 2006. [http://dx.doi.org/10.1007/0-387-36278-9\_12]
- [7] Lee M, Heo E, Lim H, Lee JW, Weon S, Chae H. Developing a common health information exchange platform to implement a nationwide health information network in south korea. *J Healthc Inform Res* 2015; 21(1): 21-9.
- [8] Elyse O, Steve EA, Alan RS, *et al.* Clinical decision support for recognition and management of hypertension: a randomized trial. *Pediatrics* 2018; 141(2): e20172954.
- [9] William R. The electronic medical record: Promises and problems. *J Am Soc Inf Sci* 1995; 46(10): 772-6.
- [10] Eta SB. Clinical decision support systems. *Health Informatics book series, Springer Science+ Business Media,LLC* 2007; 233: 3-22.
- [11] Sweeney J. Healthcare informatics. *Online J Nurs Inform* 2017; 21(1): 1-4.
- [12] Davis D A. Translating guidelines into practice: A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guideline. *CMAJ* 1997; 157(4): 408-16.
- [13] Kwok R, Dinh M, Dinh D, Chu M. Improving adherence to asthma clinical guidelines and discharge documentation from emergency departments: Implementation of a dynamic and integrated electronic decision support system. *Emerg Med Australas* 2009; 21(1): 31-7. [http://dx.doi.org/10.1111/j.1742-6723.2008.01149.x]
- [14] Jagadamba G, Krishna ELS, Amogh JP, Abhishek BB, Manoj HN. Adaptive electronic health records management and secure distribution using blockchain. In: *Lecture Notes in Networks and Systems*. Singapore: Springer 2022; 481.
- [15] Allen S. Adaptive electronic health records management and secure distribution using blockchain. *Proceedings of International Conference on Network Security and Blockchain Technology* 2019; 312-24.
- [16] Evans RS, Lloyd JF, Pierce LA. Clinical use of an enterprise data warehouse. *AMIA Annu Symp Proc* 2012; 2012: 189-98. [PMID: 23304288]
- [17] Vonbach P, Dubied A, Krahenbuhl S. Prevalence of drug- drug interactions at hospital entry and during hospital stay of patients in internal medicine. *Eur J Intern Med* 2008; 19(6): 413-20.
- [18] Richard S. *The Computer-Based Patient Record: An Essential Technology for Health Care*. Washington: National Academies Press 1997.
- [19] Eye Q. Available from: <https://www.sankaranethralaya.org/patient-care-eye-q.html>
- [20] Claudia A. Use of a checklist and clinical decision support tool reduces laboratory use and improves cost. *J Pediatr* 2016; 137(1): 2014-3019.
- [21] Christopher M. A decision support tool for using an ICD- 10 anatomographer to address admission coding inaccuracies: A commentary. *Online J Public Health Inform* 2013; 5(2): 222.
- [22] Turchin A, Shubina M, Gandhi T. NLP for patient safety: splenectomy and pneumovax Article with American Medical Informatics Association Annual Symposium.
- [23] Sim P, Greenes RA, Haynes RB, *et al.* Clinical decision support systems for the practice of evidence-based medicine. *J Am Med Inform Assoc* 2001; 8(6): 527-34.
- [24] Judith JA, Muhammad AI, Yasir A. Towards interoperability: A new resource to support nursing terminology standards. *Comput Inform Nurs* 2015; 33(12): 515-9.
- [25] Talha MA, Muhammad AI, Yasir A. A model for early prediction of diabetes. *Informatics in Medicine Unlocked* 2019; 16: p. 100204.
- [26] Sarah JJ. Remembering Lawrence weed: A pioneer of the soapnote. *J Acad Med* 2019; 94(1): 11.
- [27] Eigner I, Hamper A, Wickramasinghe N, Bodendorf F. Success factors for national eHealth strategies: A comparative analysis of the Australian and German eHealth system. *Int J Netw Virtual Organ* 2019; 21(4): 399-424. [http://dx.doi.org/10.1504/IJNVO.2019.103681]
- [28] Heimhalt Diana. E-health law, in Germany. Taylor Wessing, Lexology 2016.
- [29] Giaedi T. The impact of electronic medical records on improvement of health care delivery. *Libyan J Med* 2008; 3(1): 4.

- [30] Bahga A, Vijay KM. A cloud-based approach for interoperable electronic health records (ehrs). *IEEE J Biomed Health Inform* 2013; 17(5): 894-906. [http://dx.doi.org/10.3402/ljhm.v3i1.4739]
- [31] Max health care. Available from: <https://www.maxhealthcare.in/about-us>
- [32] Osheroff J, Pifer E, Teigh J, Sittig D, Jenders R. *Improving Outcomes with Clinical Decision Support: An Implementer's Guide*. Routledge and CRC Press 2005.
- [33] Manohara MMP. Standard electronic health record (EHR) framework for the Indian healthcare system, health services, and outcomes research methodology. *Health Serv Outcomes Res Methodol* 2021; 21(3): 339-62.
- [34] Kumar N, Narayan Das N, Gupta D, Gupta K, Bindra J. Efficient automated disease diagnosis using machine learning models. *J Healthc Eng* 2021; 2021: 1-13. [http://dx.doi.org/10.1155/2021/9983652] [PMID: 34035886]
- [35] Pieter JH, Bas OS, Prashant VNP. Drug-drug interaction checking assisted by clinical decision support: A return on investment analysis. *J Am Med Inform Assoc* 2015; 22(4): 764-72. [http://dx.doi.org/10.1155/2021/9983652]
- [36] Eta S. Performance of four computer-based diagnostic systems. *NEJM* 1994; 330(25): 1792-6.
- [37] Wright Adam. Development and evaluation of a comprehensive clinical decision support taxonomy: Comparison of front-end tools in commercial and internally developed electronic health record systems. *J Am Med Inform Assoc* 2011; 18(3): 232-42.
- [38] Available from: <https://vikaspedia.in/health/nrhm/national-health-policies/electronic-health-record-standards-for-india>
- [39] Scott Evans R, Rick Carlson, Kyle V Johnson, Brent K Palmer, James F Lloyd, "Enhanced notification of infusion pump programming errors". *J Study Health Tech Inform* 2010; 160: 734-8.
- [40] Nuria C. Machine learning applied to diagnosis of human diseases: A systematic review. *Appl Sci MDPI Acad Comm* 2020; 10(15): 5135. [http://dx.doi.org/10.32628/CSEIT206463]
- [41] Asan O, Henry NY, Betty C, *et al.* How physician electronic health record screen sharing affects patient and doctor non-verbal communication in primary care. *Patient Educ Couns* 2015; 98(3): 310-6.
- [42] Rani KMJ. Diabetes prediction using machine learning. *IJARCSST* 2020; 6: 294-305.
- [43] Fitriyani NL, Syafrudin M, Alfian G, Rhee J. Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. *IEEE Access* 2019; 7: 144777-89. [http://dx.doi.org/10.1109/ACCESS.2019.2945129]
- [44] Laila UE, Mahboob K, Khan AW, Khan F, Taekeun W. An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors* 2022; 22(14): 5247. [http://dx.doi.org/10.3390/s22145247]
- [45] diabetes. Available from: <https://www.kaggle.com/johndasilva/diabetes>
- [46] An electronic health record (EHR) is more than a digital version of a patient's paper chart. Available from: <https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/benefits-ehrs>
- [47] Thakkar M, Diane CD. Risks, barriers, and benefits of EHR systems: A comparative study based on size of hospital. *J AHIMA* 2006; 3: 3-5.
- [48] Omidian Z, Hadianfard AM. The role of clinical decision support systems in healthcare (1980-2010): A systematic review study. *Jentashapir J Sci* 2011; 2(3): 125-34.

