
Y. Sharmila Begum¹, P. Uma Maheswari²

¹Research Scholar, Department of Computer Science and Engineering, Anna University Regional Campus, Madurai, INDIA
²Assistant Professor, Department of Computer Science and Engineering, Anna University Regional Campus, Madurai, INDIA

Abstract

The quality of life for people with diabetes can be remarkably considered to improve continuous monitoring which can save human and other healthcare expenditures further deal with health-related issues. This is possible by combining a number of technologies including distinct domain like Internet of Things (IoT), Sensor Networks, contemporary cloud computing and Artificial Intelligence (AI). Simultaneously, proposed research is focuses on advancement framework for intelligent healthcare system which is especially suited for smart e-health applications to address the existing obstacles. The availability of several communication channels has made remote that are responsible for healthcare treatment at higher potential range. Further proposed research highlights an inherent support of machine learning (ML) enabled tracking system is mainly designed for diabetic patients. On the other side, smart phone’s, sensors and various smart gadgets can measure physical characteristics that are suited elements of simulation design. The gathered data is further tested for preprocessing and normalization processes in order to reduce duplication. Moreover construction of advanced automation system with IoT framework for healthcare monitoring system smoothly integrates with ML techniques. Finally simulation result proven that technology utilization ensures remedial diagnosis through real-time monitoring, endless connectivity and decision-making.

Keywords: Machine Learning, Internet of Things, Classification, Sensors, Data Processing, Classification.

1. Introduction

In general, diabetes is a major disorder characterized by increase amount of blood sugar levels further it is one of the top 10 worldwide disease that lead to death [1]. As per International Diabetes Federation (IDF) estimation nearly 537 million persons globally affect with diabetes in past 2021 which can cause 6.7 million deaths. In addition, World Health Organization (WHO) predicted that by 2030 approximately 643 million people can affect by diabetics and for 2045, it will rise into 783 million and more [2]. In row, diabetes is a chronic
disease that develops either by two cases, pancreas produces insufficient insulin or when the body cannot effectively use the insulin it generates. The important phenomenon of hormone insulin is in charge of controlling blood sugar level in addition uncontrolled diabetes frequently causes hyperglycemia which is defined by increased blood sugar level. Hyperglycemia can gradually harm individual parts of body including neurons and other blood artery and so on.

In order to expansion of diabetes in an individual under the circumstance of multiple interaction between number of risk aspects including sleep quality, alcohol intake, dyslipidemia, physical inactivity, serum uric acid levels, obesity, hypertension, cardiovascular disease, ancestry, depression, age and gender. Diabetes can worsen and lead to major problems if it is not treated in a timely manner [4]. Diabetes is a disorder that may change a person's life and be difficult for certain people to manage, perhaps leading to depressed symptoms. When compared to people without diabetes, those with diabetes have two to three time’s higher chance of developing depression, and only around 25 to 50 percent of those who do obtain effective diagnosis and treatment [5]. Remote tracking and monitoring systems collect patient data outside of regular healthcare facilities by utilizing the most recent developments in information technology. The most effective remote patient monitoring systems frequently make use of technology that is easily accessible and of consumer quality. These health monitoring systems frequently use equipment created with patient comfort in mind [6]. The use of cell phones and modern information technology has a lot of potential to improve diabetic self-management. A smartphone app created for diabetic self-management can successfully cut HbA1c levels in patients, according to promising findings from the study carried out by Kirwan and his colleagues [7]. To create a better mobile app for diabetes self-management, considering the age of the target user demography, medical professionals, academics, and app developers must collaborate [8].

Type 2 diabetes, formerly known as noninsulin-dependent or adult-onset diabetes, develops as a result of the body's ineffective use of insulin [9]. The main causes of this kind of diabetes, which is the most common in the world, are having an excessively big body and not exercising enough [10]. American Indians with diabetes can benefit from a personalized recommendation system that was developed for their study. By fusing users' ontological profiles with accepted clinical diabetic recommendations and guidelines, this technology may be used to treat diabetes. The ability to create a mobile application that can effectively propose meals and include food recognition features falls short. It also lacks an interactive visual interface and activity tracking for diabetes patients[11]. The researchers developed a diabetes management system using Java for Android, Objective C for the iPhone and PHP and MySQL for the online application. This system had a module that reminded users to record their readings by sending them an SMS notice. Additionally, it had a recording interface for readings and used a fuzzy logic-based artificial intelligence system to determine the health of users. While this system performs an excellent work of informing medical personnel about the patient's status whether missing a module or not by keep track of the patient's food and exercise routine. The major drawback is symmetric thus managing diabetes requires quality of food and exercise which must be track accordingly[12]. By the use of low-cost, low-energy sensor generation of data for classification is recommend for glucose monitoring level. Daily patient data is gathered by the sensor and it is routinely transferred to the cloud storage platform. This information further used for medical personnel to monitor changes in the patient's blood glucose levels and, if necessary, to carry out the proper
treatment actions. To maximise accuracy, our method includes a detailed analysis, comparison, and evaluation of several categorization algorithms with various parameters. The best machine learning algorithms for making predictions were ultimately chosen for the proposed methodology.

The rest of this paper is organised as follows: The analysis of related works is covered in detail in Section 2. Entire description of the proposed architecture is stated in Section 3, which makes use of an IoT-enabled system to monitor diabetes patients. More information on the system's implementation is provided in Section 4. Section 5 provides a brief overview of the data collection process and Section 6 covers the presentation of the findings and the conversations that follow. Finally, in Section 7 summarise our findings and provide conclusion with prospective directions for further study.

2. Related works

In this part, we provide an overview of earlier studies in the field of diabetes patients' blood sugar, pressure and temperature monitoring. This area also includes research on big data and predictive analytics in the field of healthcare, using classification to anticipate when blood sugar levels, blood pressure and temperature may increase or fall in the future. For improving the condition's therapy, categorization plays a crucial role in e-health monitoring.

In order to assess the effectiveness of a basic outpatient diabetes self-management education program, Zheng et al conducted a study inquiry. In this study, a total of 60 individuals with type 2 diabetes mellitus were randomly assigned to one of two groups; the intervention group had an equal number of participants as the control group did. It should be noted that a meal suggestion system was not integrated into this study[13].

According to the study presented in [14], real-time data may be used to improve prediction accuracy by combining the Internet of Things (IoT) with machine learning (ML). The suggested hardware and software solution help patients anticipate early heart illness. In order to address the unique requirements and difficulties faced by older people and the people who care for them, [15] examines cutting-edge medical procedures. A thorough review of methods for recognizing, detecting, and self-managing diabetes mellitus is also conducted by [16].

The existing system tracks the user's walking activity and saves the route, but it is unable to link the recorded route to the number of calories expended. [17] Describes how to perform a literature study that focuses on the benefits of combining telemedicine with AI, which offers substantial development potential. The article also discusses the difficulties these methods attempt to solve and looks at how telemedicine and AI have been used to improve continuous monitoring. In [18], the author uses patient blood pressure and glucose data to estimate their risk for diabetes and hypertension using supervised machine learning classification techniques. The support vector machine classification approach is identified to the most accurate after assessing multiple classification methods. The system is taught to forecast the patient's blood pressure and diabetes condition. Last but not least, [19] proposes a model that uses Li-Fi technology, which is faster and more effective than conventional Wi-Fi networks, to estimate the body's glucose levels. In order to evaluate the effectiveness of various monitoring topologies for patient tracking, Paganelli et al. [20] did a comparison study, which increases the likelihood that the virus would be found using one of these designs. Three levels make up the bulk of the proposed architecture: the application layer, the data collecting layer, and the data transmission layer. IoT-based healthcare solutions do, however, have some drawbacks, such as communication lags, latency problems, and other restrictions. In order to
effectively assess and manage the data stored in the cloud and enable the development of exact illness predictions, Kondaka [21] emphasises the significance of machine learning. Deep learning technology integration may be essential for reducing mistakes in IoMT smart health systems.

3. Proposed Methodology

Traditional machine learning systems are largely used for data collecting tools by which transmission of physiological data in real time from a single sensor to a distant server at certain time period. As compared with standard statistical approaches, the proposed model handles this data on its own and improves prediction accuracy while making fewer assumptions. Because of this, proposed technique attempts to address for periodic monitoring and controlling the systems.

The proposed monitoring system is demonstrated in Figure 1. Data collecting using sensors is the first step of the process, which is followed by data categorization using data acquisition. Training and denoising are two aspects of data assessment. Before using them as inputs for a classifier, they must undergo preprocessing, feature extraction, and maybe merger with other datasets. Following training, the classifier is put through a number of stringent tests. A decision-making process results in the alarm being set off.

With the end objective of improving user comfort and procedure efficiency the proposed investigation determines feasible to incorporate with mobile platform into the structure of a remote monitoring system.

There are four crucial decision points that might have a big impact on the effectiveness of the final system to create machine learning system. These four important checkpoints are as follows (1) the method utilized for labeling training data (2) data segmentation (3) classifier selection (4) modifying classifier requirements to take advantage of the available computing resources. Automation of training data labeling is the first step in the proposed work. Additionally, it suggests testing classifiers on mobile platforms to create accuracy-computational profiles for each candidate model. Depending on the available computing power the system will dynamically allocate a model.
Figure 1: The proposed Diabetic Monitoring System

Data acquisition module

Data acquisition

Data evaluation - Denoising

Preprocessing and Features selection by multi objective

Prediction by Classification Algorithm

Decision making and send alert to personnel
a. **Data Acquisition/Collection Layer**

The suggested diabetic monitoring system combines the data from two major databases. Numerous physiological measurements such as blood pressure, glucose and sugar levels, body temperature, respiratory rate due to motion, heart rate are gathered as part of routine health monitoring.

These data are sent through wireless medium like Bluetooth into distant gateway device, where they go through with preliminary processing which can be used for health risk assessment. The patient’s complete medical information including observation reports, full clinical records and other diabetes histories that are safely stored in the cloud platform.

b. **Stages of Pre-Processing Layer**

Effective data pre-processing is now a must before using machine learning techniques since real-world data is inherently inconsistent, incomplete, and noisy. To treat of removed data, standardization of the information is done with accurate feature selection which is required for effective prediction of diabetic disease and sugar level monitoring utilizing with sensed dataset. To overcome this missing values and noise can make it difficult to predict diabetic level working with data from wearable sensors which can increase the risk of inaccurate diagnoses or misclassifications. The data cleansing approach is used to above specific context to overcome these difficulties.

This method of data cleansing involves removing extraneous data such as noise, duplicates, and outliers. This filtering method is made for managing large amounts of real time sensing data since it generates data that is noise free and more closely matches the sensor's actual measurements. Furthermore, the outcomes closely match the actual sensor values. Two more unsupervised filters are also included in this level of processing to deal with gaps in the data and redundant values, respectively. The initial filter removes superfluous characteristics while keeping up to 90% of the maximal variance. The second filter uses the mean and median to replace any missing values in the organized dataset.

c. **Random Forest based Feature Selection**

Prior to the actual learning process, feature selection is on demand procedure in machine learning techniques. Further machine learning system can be made more exact by using feature selection. Also reduce the chance of under fitting and over fitting by using a random forest for feature selection.

The likelihood of any one decision tree falls to random prediction of the target value which is almost decreased over ensemble classifiers. Consequently, numerous decision trees representing various classifiers that are employ in static condition. To reach its conclusion, the random forest (RF) aggregates the data from all the remaining trees.

The RF margin function is represented by equation (1) which is mentioned below:

\[ mg(X, Y) = \text{avg}(I_k(h_k(X) = Y)) - \max_j(h_k(X) = j) \]  \hspace{1cm} (1)

Where I denoted as indicator function, the error function is generalization that can be expressed as follows in Equation (2):

\[ PE_\_ = P_{X,Y}(mg(X,Y)>0) \]  \hspace{1cm} (2)

Generally, I serve as an indication. Equation (2) provides the following expression for the error generalization:
This matters if you want to depict probabilities using an X- and Y-axis coordinate system. The rule \( h_k(X) = h(X, k) \) retain true for all repetition sequence (TR) as random forest classifiers (decision tree) which can improve regularly.

Following is a description of the convergence probability function (PE) Equation (3) demonstrates predicted value by maximum law of large number combined with tree topology:

\[
P_{X,Y}(P_{\theta}(h(X, \theta) = Y) - \max_j(P(h(X, \theta) = j) < 0) \quad -(3)
\]

Where, \( \theta \) - random vector, \( X \) - input vector and \( Y \) - random vector distribution.

This set of equations and ideas explains how to define and compute the margin function, generalization error, and forecast confidence of random forests, which are used for feature selection.

In this procedure, we feed a collection of classifiers known as decision trees, labelled as \( h_1(x), h_2(x) \), and so on, training data made up of \( X \) and \( Y \) vectors. The margin function is described in equation (1). Along with these fundamental classifiers, the RF bagging method is incorporated into the bootstrapping process. This ensemble learning method produces many trees instead of just one. The robust classifier often achieves a reduced error rate as the similarity between these trees rises. The choice of features and the quantity of decision trees used are two important aspects in training the RF classifier. With this method, overfitting is efficiently avoided, noise and outliers are taken into account, and good accuracy is maintained.

### d. Data Prediction Layer

The difficulties with sequence prediction have longer been acknowledged as some of the most challenging problems in the area. The machine learning method used by the data prediction layer has four essential parts which are listed below:

The following elements make up a standard classification system: a classifier, an inference engine, a knowledge base, and a declassifier. These parts are made to handle inputs that can be either numerical data or fuzzy collections of textual descriptions.

- Firstly, the process of converting chunks of \( m \) input into suitable fuzzy group is referred to as classification.
- The inference engine makes use of expert information, which is often represented in the knowledge base as a set of classification conditional rules. The values of the input variables are converted using these rules into linguistic values for the output variables.
- Domain-specific knowledge is organized and formalized in knowledge bases using a predefined database which composed of set of rules. In addition, database stores language control rules along with rule set which incorporates domain expert knowledge.
- Declassification follows discrete data into the fuzzy logic which highly produces a result of the procedure then combining numerical data with linguistic values.

Machine learning models are used in the data prediction layer to produce predictions or classifications based on the chosen attributes. Classification techniques such as decision tree, random forest, support vector machine and other could be employed for health monitoring. The historical data used to train the machine learning model comprises of goal labels (such as risk level, health status, and illness presence) as well as input attributes (such as sensor readings). In this training phase, the models discover links and patterns in the data. The model might need to be updated or retrained with fresh data over time to adjust to altering
health conditions or to boost its precision. These elements work together to create a complete system that uses classification algorithm as the foundation to address issues in sequence prediction.

Classification is the process by which the input like maximal heart rate, ECG, temperature and blood pressure are converted into classification sets using predetermined classifier value range. In contrast with the value of maximum glucose rate such variable indicates the danger hence immediate action is required. Individual patient greatest and lowest recorded glucose rate as well as blood pressure, temperature, metabolic rate and heart rate readings are entered into standard member function that addressed for further decision making. Additionally, these inputs are transformed into classifier sets with appropriate value ranges. The created classifier sets are entered into the classifier inference system which is used to classify patients based on their medical information. Within the database (DB), these operations are represented as classifier conditional strategies.

4. Results and Discussion

The RF-FL is trained and tested with i7 processor enabled Windows 10 PC with MATLAB R2020b. A new training technique is now available for software developers to accumulate the overall performance of the classified value as they construct through test bed experiments. The purpose of classifiers training method is increased by using the proposed method to train classifiers for a variety of mobile applications. It's vital to recommend the proposed method which can require more training time than the speedier 5-fold method. The advancement of the proposed framework is capable of transforming the classified information to the mobile device and assessing multiple computational effectiveness range is necessary to follow the strategies that need to outline into practice. In order to assess the effectiveness of sequential prediction models, this research used machine learning technique (classification algorithm) mainly for dataset pertaining for diabetic person in real time environment.

The digital glucose sensor offered reliable readings of both blood pressure and blood glucose levels. A pulse sensor further served as a heart rate monitor to track heart rate. In order to develop the sensor into sensible with the human body and user can determine the body's temperature. A pedometer was used to track physical activity, and a specially created application was used to immediately extract location information from the smartphone. The ESP8266-12F module was used to process each and every piece of data gathered by these sensors. The ESP8266 system on a chip served as the system's foundation. It provides a comprehensive and independent Wi-Fi networking solution.

In order to collect data from IoT devices simultaneously sensing value is transmitted to the cloud platform where it is organized and analyzed by the field of Wireless Body Sensor Networks (WBSNs) which is more essential in recent days. Using apache spark and Cassandra will act as the server and storage architecture is fixed for the tensor flow ML package. The overall experiment is carried out through I2K2 cloud resource which is crucial to remember that proposed system has various significant feedbacks, one of which is the use of a finite number of IoMT devices. On the other side, increased data volume has potential to reduce pre-processing efficiency which in turn affects the overall system accuracy.

a. Evaluation Indices

Despite, accuracy, precision, sensitivity specificity and F1-score are the recent parameter indicators used to determining the system efficiency. Accuracy evaluates how well the
The proposed random forest classifier model performs by comparing the actual outcomes with the expected ones. In connection to this proposed research, measure the classifier model performance in order to determine if a patient has a high level of sugar or not through fixed optimal value by comparing true positive (TP = 15) and true negative (TN = 60) rates. From this end, user can easily predict a model which attempt to address the overall proportion rate.

The few indicators utilized to determine a model value in terms of accuracy, sensitivity, specificity, and F1-score. By contrasting the actual results with the predicted one, accuracy assesses how well the suggested random forest classifier model runs periodically. With inherent support of true positive (TP) and true negative (TN) rates, diverse range of assess exploits how well the classifier model performs in order to detect whether a patient have diabetic disease or not. An indicator proves that how effectively a model handles false positive (FP = 20) and false negative (FN = 5) which is directly proportion of accurate predictions at final solution. The residual values are referred to the designations false positive (FP) and false negative (FN). Further, the accuracy may be determined over at the ratio of real positive observations to all positive instances. Therefore, diverse parameters like accuracy, sensitivity and specificity along with F1 measure is plotted in Figure 2a and by row figure 2b and 2c denoted both sensitivity and specificity 2b finally 2d shows F1 score measure respectively.

The following statement encapsulates in terms of accuracy statement which attempts to apply in the proposed mechanism and general correctness is:

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

The following phrase uses sensitivity to describe whether or not a given class was correctly classified:

\[
\text{Sensitivity} = \frac{\text{TP}}{\text{FP} + \text{TP}}.
\]

As seen in the following phrase, specificity indicates how frequently a certain class is properly classified:

\[
\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}
\]
F1-score represents the centered mean value of the precision and recall, followed by the expression is mentioned below:

\[ F1 \text{ score} = \frac{2 \times TP}{2 \times TP + FP + FN} \]

5. Conclusion

In conclusion, unlike many of the suggestions, which are specialised to certain phases, this study conducted a thorough classification algorithm taking into an account in terms of both accuracy and computing performance in a general context. Here with relevance to the random forest characteristics to prioritize their influence on model correctness while taking the availability of system resources into account. A unique formation of systematic predictive monitoring system can built on the significant outcomes which address to recommend for future researchers. When compared to other approaches, classification algorithm use in
classification showed greater classification skills, and the accuracy of classification findings was improved by choosing the right characteristics. In comparison to existing approaches, the simulation results show that the suggested proposed mechanism proves a greater degree of accuracy and prediction. In future, real-time application for diabetic patient monitoring is frame into AI techniques over various types of dataset.

6. References


