The Investigation of Employing Supervised Machine Learning Models to Predict Type 2 Diabetes Among Adults

Tareq Alhmiedat[1*], and Mohammed Alotaibi[2]
[1] Industrial Innovation & Robotics Center, University of Tabuk
Tabuk, 47191 KSA
[e-mail: t.alhmiedat@u.edu.sa]
[1, 2] Faculty of Computer & Information Technology, University of Tabuk
Tabuk, 47191 KSA
[e-mail: mmalotaibi@ut.edu.sa]
[*]Corresponding author: Tareq Alhmiedat

Received February 28, 2022; revised May 18, 2022; revised June 28, 2022; revised August 16, 2022; accepted August 31, 2022; published September 30, 2022

Abstract

Currently, diabetes is the most common chronic disease in the world, affecting 23.7% of the population in the Kingdom of Saudi Arabia. Diabetes may be the cause of lower-limb amputations, kidney failure and blindness among adults. Therefore, diagnosing the disease in its early stages is essential in order to save human lives. With the revolution in technology, Artificial Intelligence (AI) could play a central role in the early prediction of diabetes by employing Machine Learning (ML) technology. In this paper, we developed a diagnosis system using machine learning models for the detection of type 2 diabetes among adults, through the adoption of two different diabetes datasets: one for training and the other for the testing, to analyze and enhance the prediction accuracy. This work offers an enhanced classification accuracy as a result of employing several pre-processing methods before applying the ML models. According to the obtained results, the implemented Random Forest (RF) classifier offers the best classification accuracy with a classification score of 98.95%.

Keywords: machine learning, medical diagnosis, type 2 diabetes, diabetic prediction.

1. Introduction

Diabetes Mellitus is categorized as one of the leading killer diseases, especially in developed countries. The National Institutes of Health (NIH) define diabetes as a disease that occurs when the blood glucose (blood sugar) is high. In this situation, insulin, which is responsible for controlling body sugar levels, is not adequately produced in the patient's body. In other instances, although the body produces the required level of insulin, it cannot be effectively utilized in the system, which increases the sugar levels in the blood. This is detrimental and can even lead to death.

Recent statistics link diabetes mellitus type 2 (DM 2) with over 90% of the overall cases of the disease reported globally. Global healthcare experts state that this population is expected to grow to around 592 million people by 2035 [6]. Therefore, having adequate information and means of learning and acting on these situations will create better patient experiences. Using ML proficiency, it is easy to learn from DM 2 health condition experiences and to undertake effective planning. Machine learning algorithms have been applied to solve real problems in other areas of operations and they can make a difference in healthcare with regard to how DM 2 is understood and managed.

The severity of DM 2 varies according to a number of factors, including the population demographics and age, among others. Usually, middle-aged or older people are most likely to develop this kind of diabetes. Diabetes is listed as one of the most prevalent chronic diseases across the globe. In 2019, there were a record 463 million diabetes patients globally between the ages of 20 and 79 years old. It is estimated that in 2019, 5 million people succumbed to diabetes and other related complications alone. In addition, a predictable 1.1 million children and adolescents under 20 years are diagnosed with type 1 diabetes. The total worldwide costs incurred for diabetes treatment and other correlated complications were roughly US$760 billion in 2019 [1].

There has been a common misconception that diabetes is only prevalent in the western world, and that it is rare in other settings. However, the situation has been changing significantly, and this health complication is now global. Over the past few years, diabetes has become very common in the Kingdom of Saudi Arabia (KSA), which has a social environment highly similar to several other developing countries. Moreover, the KSA is one of the nations with the highest prevalence of diabetes – one fifth of its total population ail from this disease. The cost of diabetes treatment and management has been estimated at roughly US$10,000 annually for each patient in the KSA. Additionally, there is an inadequate number of diabetes specialty centers in the KSA, especially in remote areas. Furthermore, according to Diabetes Atlas, around 250,000 children in Saudi Arabia have type 1 diabetes; this is the highest number recorded in the Middle East [1].

Over the last few decades, artificial intelligence (AI) has gained popularity in healthcare, specifically in the management of chronic diseases. Globally, dozens of pilot studies and research studies are being conducted in this field. Most of these studies show that the incorporation of AI technology into the management of diabetes helps with the early prediction of the development of this disease. Additionally, AI technology plays a vital role in reducing the indirect costs incurred by diabetic patients by reducing the number of times they visit the hospital. It also contributes significantly to the ability to predict and diagnose the disease early and its recommendations can be trusted by physicians and patients.

Furthermore, AI can improve diabetes management in five ways: 1) early detection of the disease; 2) early detection of diabetes-related conditions; 3) predictions of blood glucose levels and personalized insulin therapy; 4) diabetes prevention and management with food and

nutrition-centric approaches; and 5) efficient collection of patient data improves personalization, thus improving communication between patients and doctors and facilitating more personalized care. The growing trend of using AI techniques in health, especially for chronic disease prediction and management, was the inspiration for the present research project. This paper presents a diabetes diagnosis system that can diagnose type 2 diabetes among adults through employing preprocessing techniques and several machine learning models. The main contributions of this paper are outlined below:

A. Develop a diagnosis system using machine learning models for the detection of type 2 diabetes among adults.
B. Adopt two different diabetes datasets (one for training and the other for testing) to enhance diagnosis system's reliability.
C. Enhance the accuracy of the employed machine learning models by adopting pre-processing functions.
D. Discuss and analyze the results obtained from employing different machine learning models on two different diabetes datasets.

The rest of this paper is organized as follows: Section 2 discusses the recent development of ML models to predict the type 2 diabetes among adults. In Section 3, the diabetic prediction system is discussed, whereas Section 4 discusses the experiments setup and shows the results obtained from several experiments. In Section 5, a discussion part that discusses the obtained results and compares them with the previous developed systems, and finally, Section 6, concludes the work presented in this paper and presents future works.

2. Related Works

The severity of type 2 diabetes, especially in developing regions like Africa, has become more rampant than the other regions. The main problem is usually evident in the diagnosis of diabetes mellitus type 2 (DM 2) and planning the appropriate means of addressing or dealing with this challenge. There is vast literature based on scientific findings which provides a clear overview of the ML models application when dealing with the DM 2. During the last few decades and with the huge revelation in artificial intelligence technology, many studies focused on using AI, especially ML technology, in healthcare. Machine learning technology and AI, in general, provide an automatic or semiautomatic support tool for early detection of the disease and also a high contribution to the management of the disease. Recently, most of the research focused on using AI technology and ML in improving chronic diseases, for instance, hypertension and diabetes.

A study introduced in 2018 aimed to develop a system that could achieve early and accurate prediction of diabetes by linking the outcomes of diverse ML techniques. The study used three supervised ML methods: Artificial Neural Network (ANN), Logistic regression, and supervised vector machine (SVM). It focused on seven attributes (features): (1) blood glucose, (2) blood pressure, (3) skin thickness, (4) insulin, (5) BMI, (6) diabetes pedigree function, and (7) age. The study did not publish sufficient details about how these ML techniques were implemented. It concluded that ML could revolutionize diabetes risk prediction with the help of advanced computational methods and the availability of a large amount of epidemiological and genetic diabetes risk datasets [2].

Another study published in 2018 designed a model that should be able to predict the likelihood of diabetes in patients with maximum accuracy. The study used three well-known ML classification algorithms – Decision Tree (DT), SVM and Naïve Bayes (NB) – for the

early detection of diabetes. Moreover, it employed the popular Pima Indians Diabetes Database (PIDD) available at the UCI machine earning repository for free. The three algorithms' performances were evaluated for their precision, accuracy and recall, with accuracy being measured over correctly and incorrectly classified instances [3].

Furthermore, a study published in 2018 compared the ability of DTs, random forest (RF) and neural networks to predict diabetes mellitus. The authors used two datasets, the Pima dataset and a dataset that comprised 14 attributes collected from hospital physical examinations in Luzhou, China. This dataset covered: age, pulse rate, breath, left systolic pressure (LSP), right systolic pressure (RSP), left diastolic pressure (LDP), right diastolic pressure (RDP), height, weight, physique index, fasting glucose, waistline, low-density lipoprotein (LDL) and high-density lipoprotein (HDL). The models were examined through the use of five-fold cross validation. The training set consisted of data from 68,994 randomly selected healthy people and diabetic patients. Moreover, principal component analysis (PCA) and minimum redundancy maximum relevance (mRMR) decreased dimensionality. The study concluded that prediction with RF achieved the highest level of accuracy (ACC = 0.8084) when all the attributes were used. However, ML can be used for the prediction of diabetes when finding proper attributes, classifier, and data mining method. This study reported that the best result for the Luzhou dataset was 0.8084, and the best performance for the Pima Indians was 0.7721 [4].

A recent study published in 2019 aimed to develop a prediction algorithm using ML, to find the optimal classifier that delivered the closest result to clinical outcomes and to identify the most effective features for diabetes prediction. The study employed the DT, RF and NB techniques. The authors used a dataset collected from the UCI machine repository that consisted of 2,500 data items containing 15 attributes. The dataset includes the most important clinical data related to diabetes. The study reported that the DT algorithm and RF had the highest specificity of 98.20% and 98.00%, respectively. Naïve Bayes achieved the best accuracy of 82.30%. Moreover, the study found that four attributes are not important for predicting diabetes – the plasma glucose postprandial, pregnancy, serum creatinine and HBAIC – due to their correlation value being small compared to other attribute values [5].

The authors of [6] gathered comprehensive data from the Murtala Mohammed specialist hospital, located in the region of Kano, to develop supervised machine models whose function was based on the SVM, RF and gradient boosting algorithms, as well as K-nearest neighbour and NB. These developments were based on the available or gathered DM2 diagnostic data. These ML techniques can be integrated into this research and used to assess the outcomes of different proficiencies. Therefore, this study can significantly contribute to this field by highlighting and recommending some of the most effective ML competencies for dealing with DM 2. The predictive learning model RF stands out as one of the best and most reliable models, with an accuracy rate of around 88.76. In the receiver operating characteristic curve, the gradient boosting and random learning models were established by the study to be the most effective models for predictive purposes with a significant predictive capacity of 86.28% [6]. This study is highly reliable as it incorporated different predictive learning models and distinguished the most effective for DM 2.

The work presented in [7], the latest study published in 2020, made a comparison between ML-based models, including XGBoost, RF, LightGBM, and Glmnet, and the commonly utilized models of regression for the prediction of undiagnosed DM 2. Therefore, this particular study creates an understanding of the possible changes that ML technologies will bring in comparison with the traditional techniques. With the available data for a 6-month period, the simple regression models recorded very low average performance. For instance,

the Glmnet recorded an average RMSE of 0.859. However, with additional data, Glmnet recorded a further improvement of +3.4%. The highest stability level was evident in the LightGBM models. Therefore, these outcomes confirm that the use of more sophisticated clinical models of prediction bring no significant clinical benefit. This study helps to explain how the stability of the chosen variables influences the model interpretation. Therefore, the calibration of models should be a key consideration when developing clinical models of prediction.

The authors of [8] presented a research study conducted in 2019, and they examined the data-driven technique of diabetes prediction. This technique was applied to identify specific patients who were diabetic. The study confirms the capability of ML models to detect the patients at risk, using ML models based on the available laboratory data obtained through a survey. The XGBoost model attained a score of 86.2% without incorporating the laboratory data and 95.7% with the data. On the other hand, the ensemble attained a 73.7% accuracy rate without data and 84.4% after incorporating data from the laboratory. This study provides further insights into the research on how data gathering can influence the model. The researchers further provide additional information on possible indicators, including age, size of waist, weight and sodium intake, which can have direct health implications.

In an alternative study conducted in [9], the authors explored the vector ML model concept to predict diabetes at different stages. Both techniques are applicable in the diabetic and pre-diabetic stages. This process can be effectively advanced using web-based tools for diabetes classification. This study affirms that how data is usually classified significantly influences the viability of the outcomes of the ML technologies. It is an indication that when using the algorithms for diabetes detection, the data used significantly influences the type of outcomes to be obtained. Therefore, this information will help in this project by enhancing the efficiency and guiding the selection of ideal applications for handling data.

The authors of [10] explored the classification methods for DM 2 prediction through ML. The algorithms have a high efficiency level, which is usually required to advance the quality of healthcare. This study focused on diabetes assessment based on family background and lifestyle. This approach adds credibility to this study by providing a broad lens for the DM 2 assessment. When the ML model is proved to be accurate, people can use it to carry out a self-assessment of their risks associated with diabetes. The researchers used 952 data samples obtained through questionnaires. Random Forest was found to be a highly effective data classifier. Therefore, RF, in this case, attained very high accuracy and it can be used for diabetes prediction.

In an exploratory study, the authors of [11] assessed another effective ML-based technique to predict diabetes. The main aim of the research was also to explore the ideal means of predicting and diagnosing diabetes with ease and accuracy. This study is highly reliable as it provides another viable option that can be used to make diabetes testing methods more accurate. The researchers assessed different ML algorithms, including the DT, genetic algorithm, logistic regression, RF, NB and SVM. The performance categories included recall, accuracy, precision and the F-measure. The algorithms were integrated to enhance the robustness of the model. The researchers established that the best means of identifying the risk of disease is by using a genetic algorithm, due to its high level of efficiency. This study contributes to this research by showcasing the need to assess different algorithms and to ascertain detection efficiency.

The authors of [12] employed SVM to early classify the diagnosis of diabetes from a medical dataset that is high dimensional, and they demonstrated the success of ML. They pointed out that an increase in the use of ML systems would make physicians' work easier

with regard to diagnosing diabetes. The machine systems explored were DT, SVM, NB and logistic regression.

In [13], the researchers reported that ML algorithms can successfully classify patient data, which helps to predict a model for the prevention of DM 2. They point out that ML can reliably detect diabetes mellitus. Furthermore, the authors employed the Pima diabetes dataset to classify and detect demographics in diabetes, and they demonstrated that women are at a higher risk of developing diabetes than men. Similar to the work presented in [12], the authors also demonstrated that ML has shown significant successes in predicting and preventing diseases.

The authors of [14] applied the K-Means clustering algorithm. They first applied PCA and later the algorithm of K-Means to fine-tune the results. Additionally, they recognized the weaknesses of such MLs by pointing that they have limitations because they do not give the same results if applied to a different dataset, such as the Pima diabetes dataset.

The work presented in [15] employed the firefly and cuckoo search-based attribute selection algorithm with the objective of achieving higher accuracy and lower training overheads for the Pima Indian diabetic database. This application classifies and predicts diabetes and identifies solutions. They emphasize that ML has been used by many researchers, mining data from patients to develop a model or system of diabetes prevention. For example, the Pima Indian diabetes is a set of data recorded from the Pima tribe, and its purpose is to improve accuracy in the prediction characteristics of DM 2. The K-Nearest Neighbor (KNN) classifier has been used in classification by calculating different learning accuracies, and it has been used for the automatic detection of diabetes from the Pima Indiana database. Furthermore, algorithms such as cuckoos and firefly have also been applied as classifiers of diabetes in the Pima diabetes database.

In [16], the researchers reported that diabetes can be managed during its earlier stages if it is predicted early enough from the changes of lifestyle and diet of patients. They predict diabetes mellitus occurrence with the help of available risk features based on an ANN known as feed-forward. They employed the ANN, a ML model computation based on the functional and structural network of biological neural. The network learns through the input-output; it is statistical data modelling that identifies patterns.

Authors of [17] proposed a diagnosis system using machine learning models for the detection of diabetes among adults. In addition, authors proposed a filter method based on decision tree to extract the most significant features in the diabetes dataset, in addition to the employment of two ensemble learning algorithms: Random Forest and Ada Boost The obtained results are efficient due to the employment of different combinations of selected features set.

As presented above, the literature review identified several works that revealed critical information on the application of machine-based prediction models for the early detection of DM 2. Additionally, data mining has become a common operational prompt in the healthcare and medicine sectors. Therefore, data mining techniques are recognized as being highly reliable for detecting and classifying diseases like DM 2 that affect a significant number of people globally. Many categories of data mining techniques can be applied to detect DM 2. When technically integrated, they are all competence-based proficiencies that can promptly guide the detection and reliable treatment of DM 2. Therefore, this information from different scientific researchers serves as a reliable evidence-based and practical basis for developing functional and effective DM 2 prediction models. The researchers, in this case, directly contribute to the advancement of research credibility by showcasing possible patterns that would not have been visible if all the methods were applied utilizing a similar technique or prompt.

3. System Design and Implementation

This section discusses the diabetic diagnosis prediction system, which consists of several phases as presented in **Fig. 1**, starting with analyzing the diabetes datasets to minimize the prediction errors in the classification models, then dividing the diabetes datasets into two subsets: training and testing. Several ML models were trained using the training dataset, and then the trained model was tested using the testing dataset.

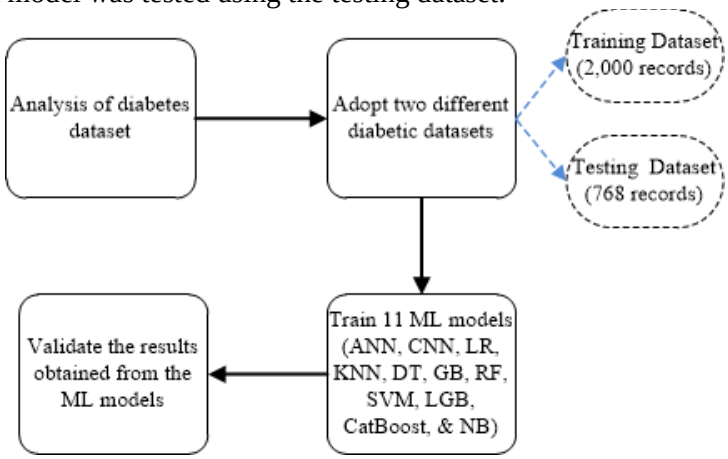

Fig. 1. ML Diabetes Prediction System

3.1 Diabetes datasets

In this project, we employed two different datasets, as follows: First, the Pima Indians Diabetes Dataset [18], originally from the National Institute of Diabetes and Digestive and Kidney Diseases. And second, the diabetes dataset, obtained from the hospital Frankfurt in Germany [19].

Pima dataset is from 768 women from a population near Phoenix, Arizona, USA. The main stimulus behind the utilization of the Pima dataset is that most of the population in today's world follow a similar lifestyle with a large dependence on processed foods and a decrease in physical activities. The outcome of this dataset is diabetes; 268 tested positive and 500 tested negatives.

On the other hand, the second dataset was collected from patients in the hospital Frankfurt in Germany with 2,000 records in total. The outcome of this dataset is diabetes; 684 tested positive and 1,316 tested negatives.

The main objective of the selected datasets is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements located in the dataset. Several research works such as [3, 4, 10, 12, 13, 15, 16] adopted the selected datasets in their studies. Both datasets consist of eight features (pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, diabetes prediction function, and age) and a single label (outcome). Each attribute (label/feature) is listed below:

- Pregnancies: this indicates the total number of pregnancies.
- Glucose: this refers to the plasma glucose concentration 2 hours in an oral glucose tolerance test.
- Blood pressure: this refers to the diastolic blood pressure (mm Hg).
- Skin thickness: this refers to triceps skin fold thickness (mm).

- Insulin: this indicates the 2-hour serum insulin (mu U/ml).
- BMI: this refers to the body mass index.
- Diabetes pedigree function: this function offers some data on diabetes mellitus history in relatives and the genetic relationship of those relatives to the patient.
- Age: this refers to the women's age in years.
- Outcome: this indicates the health status of the women (0: normal, 1: diabetes).

Fig. 2 shows the distribution of the outcome attribute for the Pima dataset, where the percentage of diabetic women is 20.05% and the percentage of non-diabetic is 79.95%. As presented, the Pima dataset is unbalanced, as three-quarters of the dataset are data for normal non-diabetic women. The distribution of the outcome attribute in the second dataset is as follows: 34.20% for the diabetic people, and 65.80% for non-diabetic people, as presented in **Fig. 3**. An imbalanced dataset poses a challenge for prediction, because most ML methods were designed with the assumption of an equal number of examples for each class. Therefore, the presented diabetes datasets lead to a poor prediction performance.

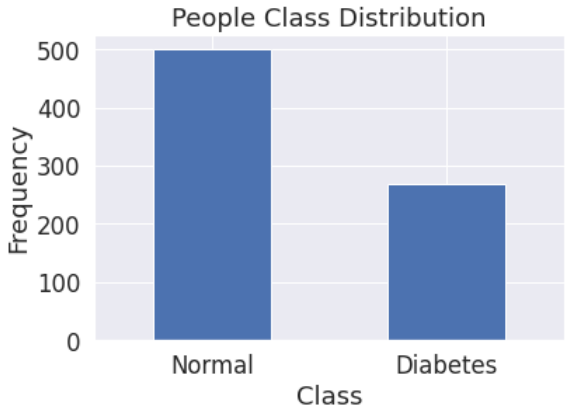

Fig. 2. The distribution of the Pima diabetes outcome attribute

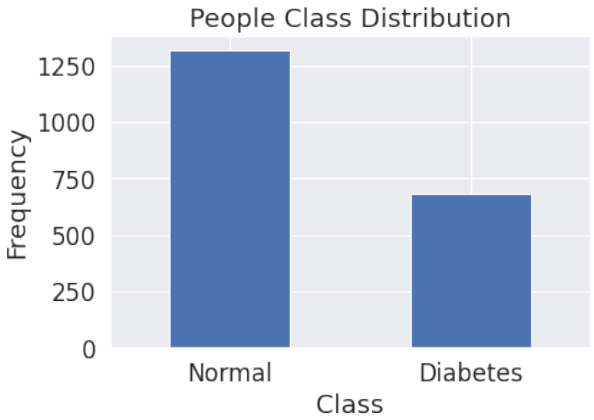

Fig. 3. The distribution of the Germany diabetes outcome attribute

3.3 Preprocessing of the Diabetes datasets

It is generally accepted that data quality affects an ML model's accuracy. The selected datasets are corrupted with various errors including missing, incorrect, or inconsistent values; therefore,

the errors that exist in datasets can drastically affect the prediction accuracy. Usually, data scientists perform several data cleaning processes before considering training the model [14]. Data cleaning is an important step for every ML model; it needs to be performed in order to enhance the prediction accuracy. This stage includes removing columns that are referred to zero-variance predictors, removing empty rows, and providing values for the empty records in the dataset.

Therefore, we obtained the correlation maps for the selected datasets (before and after applying the cleaning process), in order to present the effect of data cleaning on the diabetes dataset. Correlation is useful in data analysis and modelling as it helps us to better understand the relationship between the dataset features. **Fig. 4** shows the correlation map for features for the selected diabetes datasets.

In addition, we performed several data cleaning processes in order to remove the incorrect, inaccurate, incomplete, and missing parts of the data in both datasets. In addition, we performed several methods including modifying, replacing, and deleting, as required. The rows with null values have been removed from the employed datasets, to enhance the classification accuracy. **Fig. 5** shows the heat map for the diabetes dataset after considering cleaning operations. For instance, the correlations between the age attribute and the glucose, blood pressure, skin thickness and insulin attributes have been enriched. Likewise, the correlations between the BMI attribute and the pregnancies, skin thickness, insulin and age attributes have been enhanced, as have the correlations between the insulin attribute and the glucose, skin thickness, age, and outcome attributes. Therefore, most of the correlations between the attributes have been improved after considering several data cleaning processes.

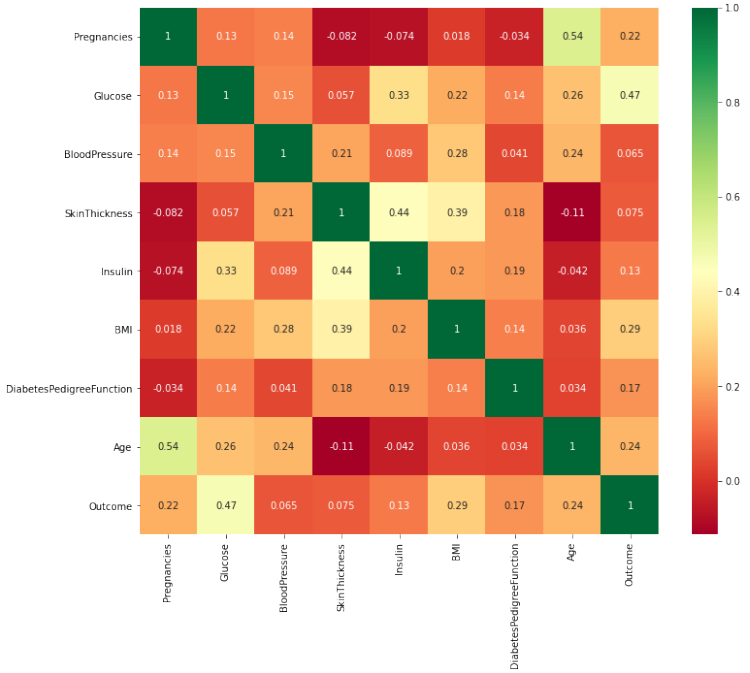

Fig. 4. Heat map for the selected diabetes datasets

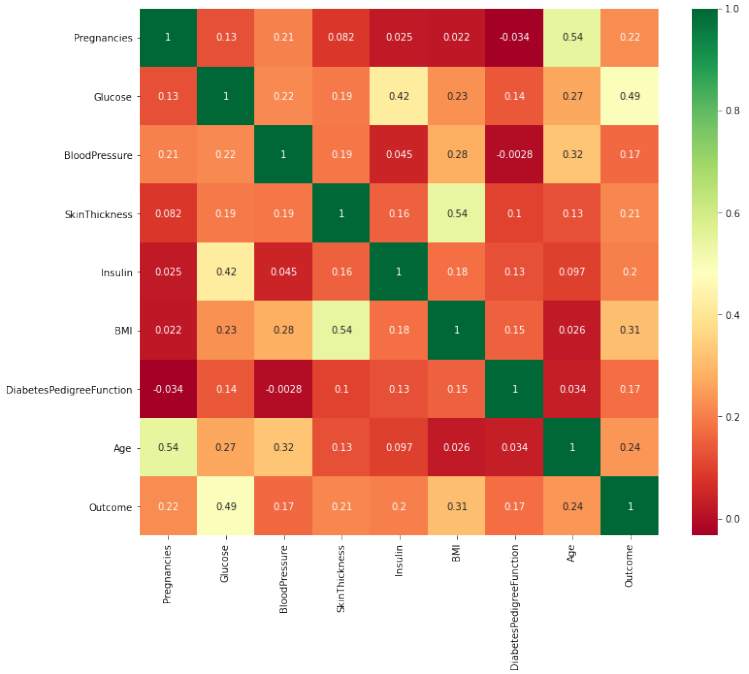

Fig. 5. Heat map for the selected diabetes dataset after applying the cleaning process

On the other hand, we employed a feature extraction method using Random Forest (RF) model to study the impact of the feature importance on the prediction accuracy. Therefore, after considering several experiments on the selected datasets, we have found that the most significant features are: Glucose, BMI, Age, and DiabetesPredictionFunction (DPF). **Fig. 6** presents the importance percentage for all features. As noticed, the Glucose is the most significant feature in both datasets, whereas the body mass index (BMI) level comes in the next place. The age feature occupies the third place, and finally, the DPF feature occupies the fourth place. According to [20], the Age, Glucose, and BMI factors are the most significant ones that can affect any person with diabetes. Therefore, it is important to study and analyze these factors for every diabetic person.

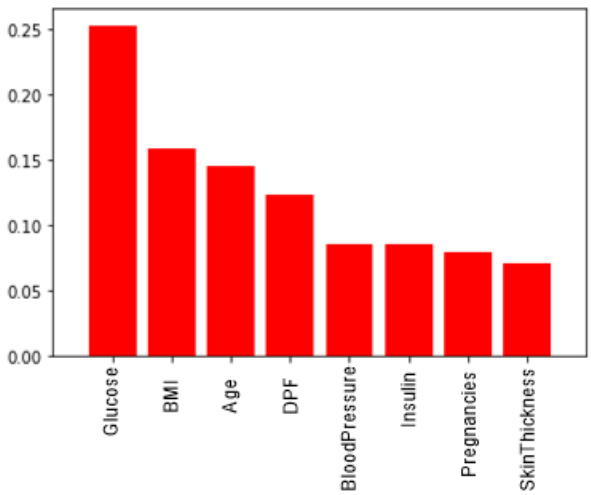

Fig. 6. The feature importance for the selected diabetes datasets

3.4 Machine Learning Models

Supervised ML algorithms rely on labelled input data to learn a function that offers an appropriate output when new unlabeled data are given. In this paper, we adopted ML models (Logistic Regression, K-Nearest Neighbor, Decision Tree, Gradient Boosting, Random Forest, Support Vector Machine, Light Gradient Boosting, and CatBoost) and deep learning models (Neural Networks, and Convolutional Neural Network), and we compare the results using two different diabetes datasets, as detailed below:

1. Artificial Neural Network (ANN): is a deep neural network model that consists of computer algorithms which aim to recognize the relationships in a group of data through a process that simulates how the human brain functions. The neural network (NN) can adapt to the changing input; therefore, the network achieves the finest results without redesigning the output criteria. The ANN model architecture is presented in **Fig. 7**, where it consists of a single input layer, 3 dense layers, and a single output layer. **Table 1** shows the ANN model summary, that involves number and capacity for each layer. The input layer contains artificial neurons which receive input from the outside world. The hidden layer consists of nodes that placed between the input and out layers, where its main job is to transform the input into something that the output unit can use. Finally, the output layer consists of units that respond to the information about how it's learned any task.
2. Convolutional Neural Network (CNN): is a class of deep neural networks, where CNN is a multilayer perceptron, and where the network is fully connected. CNN employs a mathematical convolution operation, where convolution is a specialized linear operation. **Fig. 8** shows the architecture for the CNN model, which consists of a single input layer, number of hidden and pooling layers, and a single output layer, whereas **Table 2** presents the CNN model summary with the total size of each layer. As noticed, the CNN model consists of several layers, as follows:
    - Convolutional layer: this layer aims to derive features from fixed-length segment for the diabetic datasets.
    - Batch normalization: is a technique used to automatically standardize the input data to a layer in a deep learning neural network, that aims to accelerate the training process.
    - Pooling layer: this layer intents to reduce the size of the processed data through merging the output of the neuron clusters at one layer into a single neuron in the next layer.
3. Logistic Regression (LR): a supervised learning classification algorithm used to estimate the probability of an output variable. LR is used to estimate the probability of a binary event occurring. The parameters of the LR model can be estimated using a probabilistic function called maximum likelihood estimation. The LR classifier can be derived by analogy to the linear regression hypothesis, where the LR formula is presented below:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

The tunning parameters in our experiments are as follows:

$$\theta = \begin{bmatrix} 0.629 \\ 0.931 \end{bmatrix}$$

4. K-Nearest Neighbor (KNN): a ML method that can be employed to solve both classification and regression problems. The KNN algorithm supposes that similar

things are near to each other. KNN works by estimating the distance between a query and all the examples in the data, choosing a specific number of examples closest to the query, and then votes for the most frequent label. For our experimental testbed, the *K* value was set to 41, where *K* refers to the number of nearest neighbors to a specific data point.

5. Decision Tree (DT): a ML method where the data in the considered dataset are continuously split up according to a certain parameter. The tree consists of two entities, decision nodes and leaves, where the leaves are the outcome, and the decision nodes are where the data is split up. The nodes in a tree are split up based on the concept of impurity. Impurity is a measure of the labels' homogeneity on a certain node. Information gain employs an entropy measure as the impurity measure and divides a node. Gini is a function that is employed to measure the quality of a split. The supported criteria are "gini" for the Gini impurity and "entropy" for the information gain, as presented below:

$$Gini: Gini(E) = 1 - \sum_{j=1}^{c} p_j^2$$

$$Entropy: H(E) = -\sum_{j=1}^{c} p_j \log p_j$$

6. Gradient Boosting (GB): a ML method for classification and regression, which builds a prediction model in the form of an ensemble of weak prediction models, normally DTs.
7. Random Forest (RF): is a ML algorithm. The forest is an ensemble of DTs, normally trained with the bagging method. The bagging method includes a combination of learning models to increase the overall result. In general, RF builds multiple DTs and combines them together to obtain a more accurate prediction.
8. Support Vector Machine (SVM): considered as one of the most robust prediction methods, and based on statistical learning frameworks. The SVM training algorithm builds a model that assigns new examples to one category or the other. The presented SVM model creates decision boundary that makes the distinction among two or more classes. However, as discussed earlier, the presented datasets are noisy and not linearly separable in most cases, where the standard SVM tries to distinct all positive and negative examples, and dose not permit any points to be misclassified. Therefore, SVM involves adopting two different parameters: C and Gamma. C parameter adds a penalty for every misclassified data point. Gamma parameter on the other hand, controls the distance of influence of a single training point.
9. Light Gradient Boosting (LightGB): an open-source implementation of GB that is designed to be effective and more efficient than other implementations.
10. CatBoost: a ML algorithm for GB on DTs. The CatBoost method has been employed in searching, recommendation, self-driving cars and personal assistant systems.
11. Naïve Bayes (NB): a classification technique based on Bayes' theorem with an assumption of strong independence between features. An NB classifier assumes that the features are independent from each other.

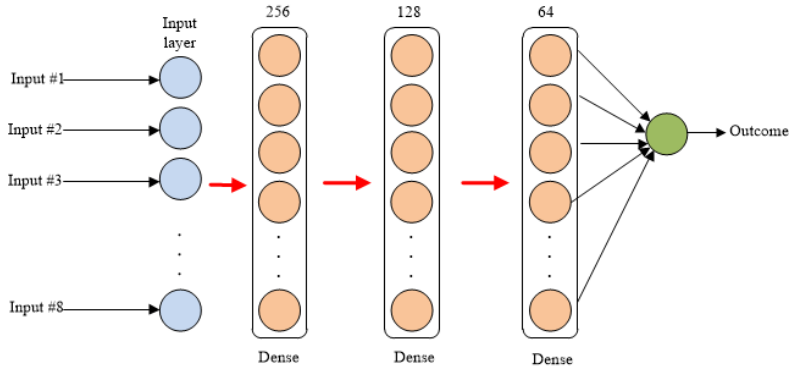

Fig. 7. The designed ANN model

Table 1. The ANN model summary

Layer (type)	Output Shape
dense_28 (Dense)	(None, 256)
dense_29 (Dense)	(None, 128)
dense_30 (Dense)	(None, 64)
dense_31 (Dense)	(None, 1)

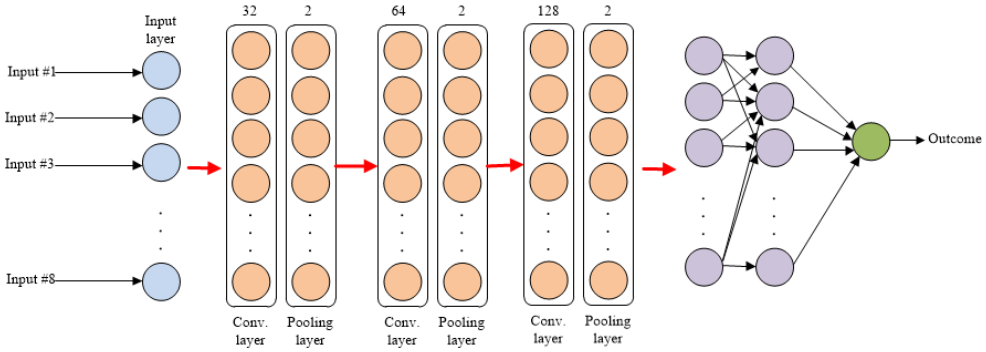

Fig. 8. The designed CNN model

Table 2. The CNN model summary with Pooling layer

Layer (type)	Output Shape
conv1d	(None, 29, 32)
batch_normalization	(None, 29, 32)
max_pooling1d	(None, 14, 32)
dropout (Dropout)	(None, 14, 32)
conv1d_1	(None, 13, 64)
batch_normalization_1	(None, 13, 64)
max_pooling1d	(None, 6, 64)
dropout_1	(None, 6, 64)
conv1d_2	(None, 5, 64)
batch_normalization_2	(None, 5, 64)
max_pooling1d_2	(None, 2, 64)
dropout_2	(None, 2, 64)
flatten	(None, 128)
dropout_3	(None, 128)
dense	(None, 128)
dropout_4	(None, 256)
dense_1	(None, 1)

4. Experimental Results

This section discusses the experimental testbed, and presents the results obtained from applying several ML models on both diabetes datasets.

4.1 Experimental Testbed

For our experiments, we employed the Colab for evaluation purposes. It is a hosted Jupyter notebook environment that is free to use, runs in the Cloud, and supports free GPU. In addition, Colab supports many ML libraries, including Tensorflow and Keras, where they can be loaded into Colab.

On the other hand, the train-test division procedure is essential to estimate the performance of the ML method. As mentioned earlier, we have employed two different datasets: one of training with a total number of 2,000 records, and the other one for testing with total number of 768 records. **Fig. 9** shows the structure of the training and testing datasets, employed in the training and testing phases. The first dataset was used to fit the model and was named the training dataset, whereas the second dataset was used to test the model, and it was named the testing dataset.

As mentioned above, both datasets are is unbalanced, and therefore it is critical to investigate several factors other than the accuracy, to correctly assess the efficiency of the employed ML classification model. The efficiency of prediction models was evaluated by assessing several parameters, as presented in **Table 8**.

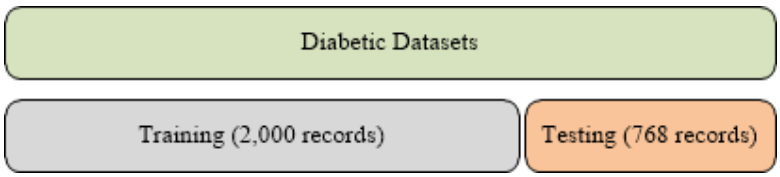

Fig. 9. The structure of the training and testing datasets

The tuning parameters for each ML model is presented below. For instance, **Table 3** shows the presents the tunning parameters for the GB classifier, whereas **Table 4** presents the tuning parameters for the RF classifier. **Table 5** includes the tuning parameters for the SVM classifier, **Table 5** shows the tuning parameters for the LGB classifier, and finally **Table 7** presents the tuning parameters for the CatBoost classifier. On the other hand, the training parameter for the decision tree classifier is the max depth (max_depth = 16).

Table 3. Tuning parameters for the GB classifier

Parameter	Value
Max depth	8
Learning rate	0.4
Iterations	55

Table 4. Tuning parameters for RF classifier

Parameter	Value
Number of estimators	250
Minimum sample leaf	80

Table 5. Tuning parameters for the SVM classifier

Parameter	Value
Gamma	1
Kernel	rbf
C	0.15

Table 6. Tuning parameters for LGB

Parameter	Value
Maximum depth	5
Number of leaves	100
Min sample in leaf	25
Learning rate	0.1
Number of iterations	120

Table 7. Tuning parameters for the CatBoost classifier

Parameter	Value
Depth	8
Learning rate	0.05
Iterations	65

Table 8. The performance metrics

Metric	Definition	Equation
Accuracy	refers to how often the classifier is correct overall. In other words, accuracy is the total number of cases correctly predicted over the total number of cases	$\boldsymbol{Accuracy} = \frac{\boldsymbol{cp}}{\boldsymbol{t}}$ where *cp* and *t* refer to the correct predictions and the total number of samples, respectively
Precision	the percentage of the correctly predicted over the total prediction cases, in other words, precision refers to how often the classifier predicts the correct answer	$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}}$ where *TP* and *FP* refer to the number of true positive cases and false positive cases, respectively
Recall	the ratio of true positives to the total of true positives and false negatives.	$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}$ where *FN* refers to the number of false negative cases.
F1-score	the weighted harmonic mean of precision and recall	$\boldsymbol{F1\ score} = \boldsymbol{2} \times \frac{\boldsymbol{Precision} \times \boldsymbol{Recall}}{\boldsymbol{Precision + Recall}}$
Misclassification Rate (MCR)	this refers to how often the classifier is wrong	$\boldsymbol{MCR} = \frac{\boldsymbol{FP + FN}}{\boldsymbol{TP + TN + FP + FN}}$
True Negative Rate (TNR)	this refers to how often the classifier predicts "No" when its actually "No".	$\boldsymbol{TNR} = \frac{\boldsymbol{TN}}{\boldsymbol{TN + FP}}$
Cohen's Kappa	is a metric used to evaluate the agreement between two raters and to assess the performance of a classification model	$\boldsymbol{k} = \frac{\boldsymbol{p_o - p_e}}{\boldsymbol{1 - p_e}}$, where $\boldsymbol{p_o}$ and $\boldsymbol{p_e}$ refer to the overall accuracy of the ML model, and hypothetical probability of chance agreement, respectively
Receiver Operating Characteristic	this refers to the overall performance of a classifier at all classification thresholds. The ROC curve summarizes the trade-off between the true positive rate and the false positive rate through different probability thresholds.	$\boldsymbol{TPR} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}, \boldsymbol{FPR} = \frac{\boldsymbol{FP}}{\boldsymbol{FP + TN}}$
Normalized Mutual Information (NMI)	this refers to a normalization of mutual information score to scale the results between 0 value (no mutual information) and 1 value (perfect correlation).	$\boldsymbol{NMI(Y, C)} = \frac{\boldsymbol{2 \times I(Y; C)}}{\boldsymbol{[H(Y) + H(C)]}}$ where *Y* is class labels, *C* is cluster labels, *H(.)* is the entropy, *I(Y:C)* is the mutual information *b/w Y* and *C*.

Table 9 shows a comparison of 11 classification models by evaluating several significant parameters: accuracy, precision, recall, F1-score, MCR, Kappa and NMI statistics. The accuracy is the most significant performance measure and its simply refers to the ratio of correctly predicted observations (normal and diabetic) to the total observations (normal and diabetic). A certain ML model is considered as the best classification model when that model offers high accuracy. However, accuracy is a great measure when the dataset is balanced, when the amount of positive and negative values is the same, but in the selected datasets, the data are unbalanced, as stated earlier. Therefore, it is important to study other attributes/parameters, to assess the performance of the classification models. **Fig. 10** presents the accuracy, precision, recall, F1-score, MCR, TNR, Kappa, and NMI for each ML mode.

Table 9. A comparison of the different prediction models

Model	Accuracy	Precision	Recall	F1-score	MCR	TNR	Kappa	
NN	68.83%	74.52%	79.00%	76.69%	31.26%	61.38%	0.3574	0.1539
CNN	86.96%	81.65%	89.00%	85.16%	19.58%	71.54%	0.4921	0.1774
LR	77.34%	77.00%	77.00%	77.00%	22.65%	53.73%	0.4672	0.1858
KNN	97.52%	78.00%	75.00%	76.00%	20.57%	60.82%	0.5258	0.2262
DT	98.30%	99.00%	98.00%	98.00%	01.56%	95.89%	0.9624	0.8647
GB	90.10%	90.00%	88.00%	89.00%	09.89%	82.83%	0.7791	0.5080
RF	98.95%	99.00%	99.00%	99.00%	01.17%	97.38%	0.9711	0.9049
SVM	77.73%	77.00%	72.00%	73.00%	22.26%	52.23%	0.4710	0.1951
LGBM	98.69%	99.00%	98.00%	99.00%	01.30%	97.01%	0.9711	0.8970
CatBst	98.43%	99.00%	98.00%	98.00%	01.56%	95.89%	0.9653	0.8857
NB	75.13%	73.00%	72.00%	72.00%	24.86%	59.70%	0.4405	0.1514

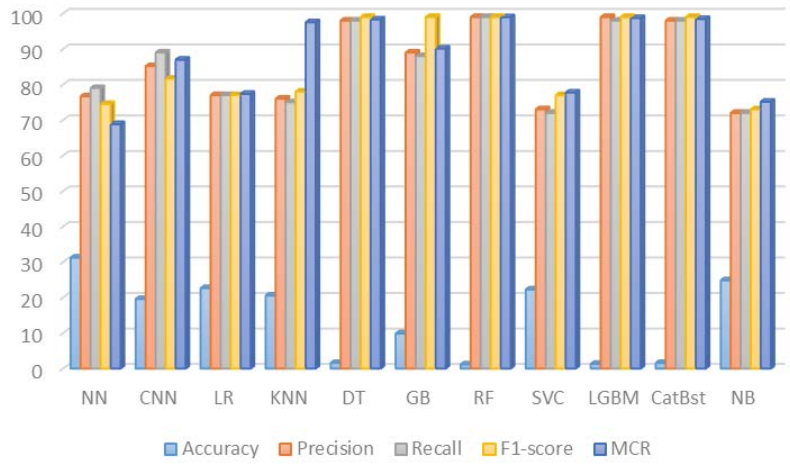

Fig. 10. Accuracy, precision, recall F1-score, and MCR for each ML model

The RF ML model achieves the best classification accuracy (98.95%), and this indicates that the prediction system was able to predict 98.95% of all cases in the selected dataset. On the other hand, the precision attribute was also assessed for each classification model. Precision refers to the ratio of correctly predicted positive observations (normal persons who were correctly predicted as normal) to the total predicted positive observations (overall persons who were predicted as normal persons). DT, GB, RF, LightGB, and CatBoost offer the best precision value (99.00%). In addition, the recall parameter was estimated for each ML model,

where recall is the ratio of the correctly predicted observations (persons who were correctly predicted as normal) to all the observations in the actual class. According to the results obtained from our experiments, RF offers the best recall value (99.00%). This means that the ratio of normal persons in the selected dataset is greater than the ratio of diabetic persons.
Furthermore, we assessed the stability of each ML model based on the adoption of the cross-validation technique. **Fig. 11** presents the cross-validation for each ML model. As noticed, the RF classifier offers the best cross-validation value among all competitors.

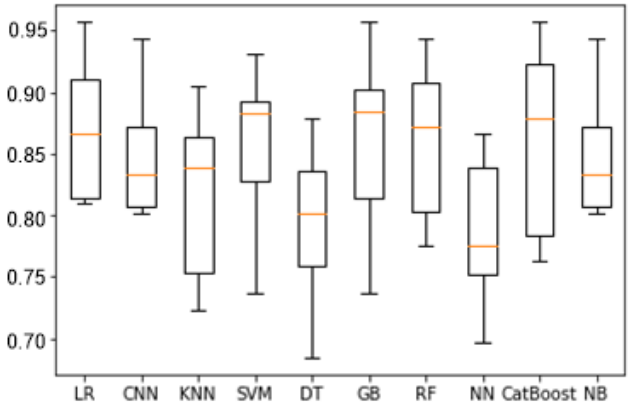

Fig. 11. The cross-validation for each ML model

The F1-score was also estimated for each ML model. The F1-score is more useful than accuracy, especially when the dataset is unbalanced as with the employed datasets. In general, accuracy is essential when false positives (health persons who were wrongly predicted as diabetic) and false negatives (diabetic persons who were wrongly predicted as healthy) have a similar cost. The F1-score is the weighted average of precision and recall; this score takes both the false positive and false negative observations into account. In general, the F1-score is more useful than accuracy, especially when the dataset has an uneven class distribution as with the employed diabetes datasets. The prediction accuracy works well whenever the false positive observations (the diabetic persons who were classified as normal persons) and false negative observations (the normal persons who were classified as diabetic persons) have a similar cost. However, if the cost of false positive observations and false negative observations are different, as with the selected datasets, then it is better to estimate the precision and recall. RF and LighGb models had the best F1-score (99.00%).

In addition, we assessed the performance of the diabetic prediction systems using the ROC probability curve. ROC can tell to what degree the model is capable of distinguishing between diabetic and non-diabetic person. The ROC curve is plotted with true positive rate (TPR) (the rate of persons who were correctly predicted as non-diabetic persons) against the false positive rate (FPR) (the rate of diabetic persons who were wrongly predicted as non-diabetic persons). ROC is the baseline for determining the performance of the diabetic prediction model. In general, ROC separates the area into two subareas (good or poor). An ROC curve of a perfect classifier is the combination of two straight lines moving away from the baseline towards the top-left corner. As presented in **Fig. 12**, RF is the best classifier performance in terms of ROC.

The Misclassification Rate (MCR) parameter is also investigated for all ML models. RF achieves the best MCR for the selected datasets with 1.17%. RF model offers the minimum ratio of classifications that were misclassified.

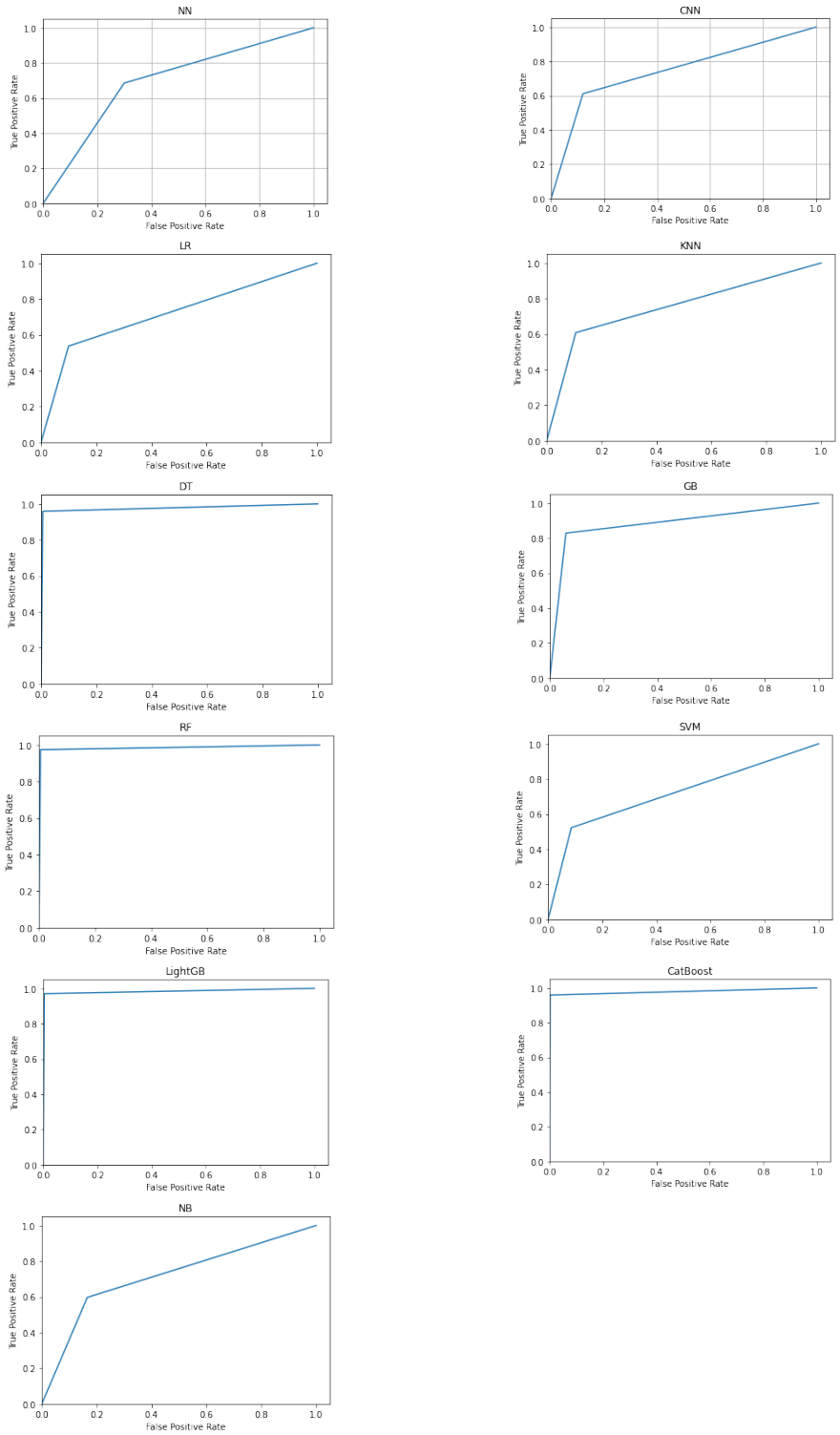

Fig. 12. ROC curve for 11 ML models

A significant evaluation attribute was assessed, the True Negative Rate (TNR). TNR represents how often the classifier predicts that the person has diabetes when the person has a diabetic disease. This is one of the most significant parameters, since the classifier must be able to predict people who have diabetes symptoms in an efficient way. Among all the implemented ML models, the RF offers the best TNR, as 261 persons were classified correctly as diabetic and only 7 persons were wrongly classified as normal persons. Therefore, in our study the TNR is a critical parameter. Finally, we discuss the Cohen's Kappa statistics for each ML model. As presented in **Table 3**, RF and LightGB offer the best Kappa value with 0.9711. in addition, we measure the score of NMI, where the developed RF classifier offers the best NMI value (0.9049) among all the competitors.

In order to add clarification to the obtained results, we estimated the prediction accuracy for each category: normal and diabetic, as presented in **Table 10**. The best classification accuracy for classifying normal people was 99.80%, whereas for diabetic persons it was 97.83%. As presented in **Fig. 13**. the prediction accuracy for normal people is greater than for diabetic ones. This is due to the structure of the dataset, which is unbalanced, and this affects the overall prediction accuracy. DT and CatBoost classifiers achieve the best prediction accuracy for normal persons, whereas the RF offers the best prediction accuracy for diabetic persons.

Table 10. The prediction accuracy for each ML model

Model	Prediction Accuracy	
	Normal	*Diabetic*
NN	79.00%	50.00%
CNN	89.00%	62.96%
LR	90.00%	53.73%
KNN	89.40%	60.82%
DT	99.80%	95.89%
GB	94.00%	82.83%
RF	99.60%	97.83%
SVC	91.40%	52.23%
LightGBM	99.60%	97.01%
CatBoost	99.80%	95.89%
NB	83.40%	59.70%

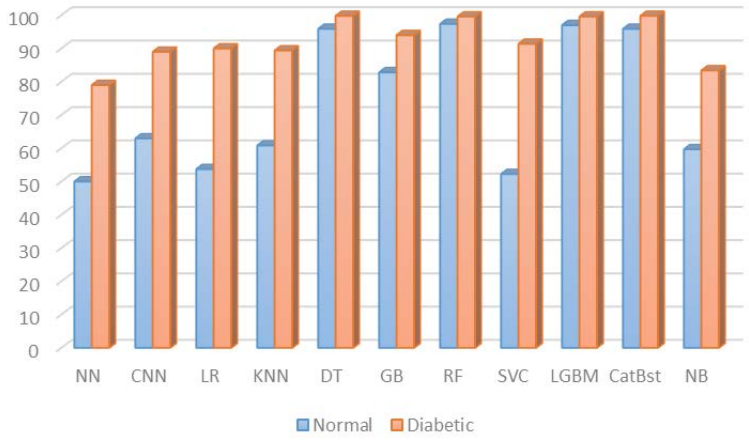

Fig. 13. Prediction accuracy for each class (normal/diabetic)

5. Discussion

This section discusses the results obtained from Section 4, and compares them with the results obtained from previous research works that employed the diabetes datasets. In this paper, we performed several data cleaning processes, in order to enhance the classification results, since data cleaning improves the quality of data and therefore increases the overall productivity. As soon as the employed datasets are cleaned, all of the outdated data and incorrect information are eliminated, leaving the highest quality data. Quality is the major concern, and dirty data may lead to incorrect predictions and inefficient data analysis.

Usually, CNN models beat the RF model in many problems, however, in our case, the employed RF model achieves better classification accuracy than the implemented CNN model. In general, RF is less computationally expensive and does not require GPU to complete the training phase. However, RF offers various interpretation of a decision tree but with better performance. In addition, according to [21, 22] RF offers much better classification accuracy especially with tabular data, as with the employed diabetic dataset.

The success of CNN is not universal across all domains. For learning problems without any special structure, or in cases where the dataset is somehow limited, CNN models are not able to perform well with respect to traditional ML models, such as RF [23]. According to [24], there are four criteria can be investigated to assess the performance of developed ML model: performance, robustness, cost and time expenditure, and comprehensibility. However, according to the obtained results in the previous section, the developed RF model offers better performance, robustness, cost, and time expenditure than the developed CNN model.

As presented in the previous section, the RF classification model achieved the best classification accuracy, precision and F1-score. RF offers the best classification accuracy for predicting both normal and diabetic persons. In addition, RF offers the best precision result and F1-score. DT, GB, LightGB and CatBoost offer similar precision results (99.00%) to the RF model.

In addition, we discuss the classification accuracy for the existing ML models that have recently been implemented with the employed datasets, as presented in **Table 11**. The best classification accuracy was obtained by the work presented in [25] with 82.00% classification accuracy. However, in this paper, we investigated several ML models' ability to predict diabetic persons. According to the obtained results, the RF ML model achieves the best classification accuracy (98.95%) after considering several data cleaning processes. Therefore, the developed diabetic prediction system offers a significant improvement over the recent developed diabetic prediction systems.

Table 11. A comparison between the existing ML models using the diabetes datasets

Research work	ML model	Accuracy
[3]	SVM	65.10%
	Decision tree	73.82%
	NB	76.30%
[4]	RF	77.21%
	J48	75.34%
	NN	73.90%
[10]	LR	74.40%
	KNN	70.80%
	SVM	74.40%
	NB	68.90%

	DT	69.70%
	RF	70.00%
[12]	SVM	78.00%
[13]	PCA, K-Means algorithm	72.00%
[15]	Firefly & Cuckoo Search algorithms	81.00%
[16]	Feed-forward NN	82.00%

6. Conclusion and Future work

Through this paper, we developed a diagnosis prediction system for diagnosing type 2 diabetes among adults, through investigating the classification accuracy of 11 different ML models, using two different datasets. The employed diabetes datasets have been preprocessed before investigate the performance of 11 ML models. As results, the implemented RF classifier achieves the best accuracy score (>98%). For future works, continuous studies on this topical issue are crucial for establishing the most effective form of algorithm. Additionally, it is significant to use advanced forms of classifiers such as the evolutionary algorithm for diabetes detection. Moreover, we aim to employ embedded based feature selection approach to select the most significant features in the dataset.

Acknowledgement

The authors would like to acknowledge the financial support for this work from the Deanship of Scientific Research (DSR), University of Tabuk, Tabuk, Saudi Arabia, under grant number S-1440-0262.

References

[1] "International Diabetes Federation," *IDF Diabetes Atlas, 9th edn.*, 2019. [Online]. Available: https://www.diabetesatlas.org.

[2] N. J. Tejas and Prof. P. M. Chawan, "Diabetes Prediction Using Machine Learning Techniques," 2018. Article (CrossRefLink)

[3] D. Sisodia and D. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Computer Science*, vol. 132, pp.1578-1585, 2018. Article (CrossRefLink)

[4] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in Genetics*, vol. 9, 2018. Article (CrossRef Link)

[5] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," *Journal of Big Data*, vol. 6, no. 1, 2019. Article (CrossRefLink)

[6] L. J. Muhammad, A. A. Ebrahem, Sani Sharif Usman, "Predictive supervised machine learning models for diabetes mellitus," *SN Computer Science*, vol. 1.5, pp. 1-10, 2020. Article (CrossRefLink)

[7] L. Kopitar et al., "Early detection of type 2 diabetes mellitus using machine learning-based prediction models," *Scientific Reports*, vol. 10.1, pp. 1-12, 2020. Article (CrossRefLink)

[8] A. Dinh et al., "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Medical Informatics and Decision Making*, vol. 19.1 pp. 1-15, 2019. Article (CrossRefLink)

[9] W. Yu et al., "Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes," *BMC Medical Informatics and Decision Making*, vol. 10.1, pp. 1-7, 2010. Article (CrossRefLink)

[10] Tigga and S. Garg, "Prediction of type 2 diabetes using machine learning classification methods," *Procedia Computer Science*, vol. 167, pp. 706-716, 2020. Article (CrossRefLink)
[11] S. Kaul and Y. Kumar, "Artificial intelligence-based learning techniques for diabetes prediction: challenges and systematic review," *SN Computer Science*, vol. 1, no. 6, pp. 1-7, 2020. Article (CrossRefLink)
[12] V. A. Kumari and R. Chitra, "Classification of diabetes disease using support vector machine," *Int J Eng Res Appl.*, vol. 3, no. 2, pp. 1797-801, 2013. Article (CrossRefLink)
[13] R. N. Patil and R. N. Patil, "A novel scheme for predicting type 2 diabetes in women: using K-means with PCA as dimensionality reduction," *International Journal of Computer Engineering and Applications*, vol. XI, no. viii, pp. 76–87, 2017. Article (CrossRefLink)
[14] P. Li, X. Rao, J. Blase, Y. Zhang, X. Chu, C. Zhang, "Cleanml: A benchmark for joint data cleaning and machine learning [experiments and analysis]," *arXiv preprint arXiv:1904.09483*, 2019. Article (CrossRefLink)
[15] R. Haritha, D. S. Babu, P. A. Sammulal, "A hybrid approach for prediction of type-1 and type-2 diabetes using firefly and cuckoo search algorithms," *International Journal of Applied Engineering Research*, vol. 13, no. 2, pp. 896-907, 2018. Article (CrossRefLink)
[16] Y. Zhang et al., "A feed-forward neural network model for the accurate prediction of diabetes mellitus," *International Journal of Scientific and Technology Research*, vol. 7, no. 8, pp. 151-155, 2018. Article (CrossRefLink)
[17] Haq, A.U., Li, J.P., Khan, J., Memon, M.H., Nazir, S., Ahmad, S., Khan, G.A. and Ali, A., "Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data," *Sensors*, 20(9), p.2649, 2020. Article (CrossRefLink)
[18] J. W. Smith, J. E. Everhart, W. C. Dickson, W.C. Knowler, R.S. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," in *Proc. of the Symposium on Computer Applications and Medical Care. IEEE Computer Society Press*, pp. 261-265, 1988. Article (CrossRefLink)
[19] [Online]. Available: https://www.kaggle.com/johndasilva/diabetes. Last accessed: 13/05/2022.
[20] Patil RN, Patil RN, "A novel scheme for predicting type 2 diabetes in women: using K-means with PCA as dimensionality reduction," *International Journal of Computer Engineering and Applications*, XI(Viii), 76–87, 2017. Article (CrossRefLink)
[21] Khan, G.A., Hu, J., Li, T., Diallo, B. and Wang, H., "Multi-view data clustering via non-negative matrix factorization with manifold regularization," *International Journal of Machine Learning and Cybernetics*, 13(3), pp.677-689, 2022. Article (CrossRefLink)
[22] Diallo, B., Hu, J., Li, T., Khan, G.A., Liang, X. and Zhao, Y., "Deep embedding clustering based on contractive autoencoder," *Neurocomputing*, vol.433, pp.96-107, 2021. Article (CrossRefLink)
[23] Roßbach, P., "Neural networks vs. random forests–does it always have to be deep learning," *Germany: Frankfurt School of Finance and Management*, 2018. Article (CroffRefLink)
[24] Wang, S., Aggarwal, C. and Liu, H., "Random-forest-inspired neural networks," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 9(6), pp.1-25, 2018. Article (CrossRefLink)
[25] Castro-Rodríguez, M., Carnicero, J.A., Garcia-Garcia, F.J., Walter, S., Morley, J.E., Rodríguez-Artalejo, F., Sinclair, A.J. and Rodríguez-Mañas, L., "Frailty as a major factor in the increased risk of death and disability in older people with diabetes," *Journal of the American Medical Directors Association*, vol. 17(10), pp.949-955, 2016. Article (CrossRefLink)

Tareq Alhmiedat is an Associate Professor in the Department of Computer Science, and the head of Robotics & Artificial Intelligence Unit, in the Industrial Innovation & Robotics Center at University of Tabuk, Tabuk, Saudi Arabia. His research interests include tracking mobile targets through Wireless Sensor Networks, robot navigation and orientation, and robot vision systems. Dr. Alhmiedat has been involved in various research projects including: SafetyNET, IndoorTrack, Diabetic Robotics, and WSN environment monitoring.

Mohammed Alotaibi received the B.S. in computer Science from King Saud University in July 2008 and M.S. degrees in Computer and Information Networks, Essex University, UK (Jan, 2011)., respectively He got Ph.D. in health informatics from Kingston University London in March 2015. He has served as assistant Prof and then associate prof at the faculty of computers and information technology at the University of Tabuk since August 2015 until now. He owned three research project funds from different organizations. He is now a research leader of a research group that focus on using robotic and AI in chronic diseases which owned a research fund worth 100.000 Saudi riyals.