Informatics in Medicine Unlocked 21 (2020) 100482

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

A comparative analysis on diagnosis of diabetes mellitus using different approaches – A survey

Fareeha Anwar [a,*], Qurat-Ul-Ain [a], Muhammad Yasir Ejaz [a], Amir Mosavi [b,c,d,**]

[a] Department of Computer Science and Software Engineering, International Islamic University Islamabad (IIUI), Islamabad, 44000, Pakistan
[b] School of Economics and Business, Norwegian University of Life Sciences, 1430 Ås, Norway
[c] John von Neumann Faculty of Informatics, Obuda University, 1034 Budapest, Hungary
[d] Department of Informatics, J. Selye University, 94501 Komarno, Slovakia

ARTICLE INFO

Keywords:
Diabetes mellitus
Deep learning
Machine learning
Accuracy

ABSTRACT

Diabetes Mellitus is commonly known as diabetes. It is one of the most chronic diseases as the World Health Organization (WHO) report shows that the number of diabetes patients has risen from 108 million to 422 million in 2014. Early diagnosis of diabetes is important because it can cause different diseases that include kidney failure, stroke, blindness, heart attacks, and lower limb amputation. Different diabetes diagnosis models are found in literature, but there is still a need to perform a survey to analyze which model is best. This paper performs a literature review for diabetes diagnosis approaches using Artificial Intelligence (neural networks, machine learning, deep learning, hybrid methods, and/or stacked-integrated use of different machine learning algorithms). More than thirty-five papers have been shortlisted that focus on diabetes diagnosis approaches. Different datasets are available online for the diagnosis of diabetes. Pima Indian Diabetes Dataset (PIDD) is the most commonly used for diabetes prediction. In contrast with other datasets, it has key factors which play an important role in diabetes diagnosis. This survey also throws light on the weaknesses of the existing approaches that make them less appropriate for a diabetes diagnosis. In artificial intelligence techniques, deep learning is widespread and in medical research, heart rate is getting more attention. Deep learning combined with other algorithms can give better results in diabetes diagnosis and heart rate should be used for other cardiac disease diagnoses.

1. Introduction

Among medical diagnosis, a diabetes diagnosis is one of the major challenges. The World Health Organization (WHO) report shows that the number of diabetes patients has risen from 108 million to 422 million in 2014. An estimate shows that by 2045, this number may reach 629 million. In 2016, the estimated 1.6 million deaths were reported due to diabetes. Early diagnosis of diabetes is significant in lowering the chances of different diseases like kidney failure, stroke, blindness, heart attacks, and lower limb amputation.

Many machine learning techniques have been used in the medical diagnosis system. They have proven to be accurate in diagnosis, successful in treatments, and cost-efficient. Diabetes Mellitus is a metabolic disorder in which the body is unable to use insulin or to store and use glucose for energy and does not make insulin [1]. Different classification techniques are used to deal with different medical problems. There are multiple types of diabetes, such as Type1, Type 2, and gestational diabetes. In type 1, the pancreas fails to produce sufficient insulin for the body. Whereas in Type 2, the body is unable to use insulin properly. It is the most common type of diabetes. The third type of diabetes is Gestational Diabetes. It occurs in pregnant women having high glucose levels in the blood [4].

Deep learning is a subset of machine learning in Artificial Intelligence (AI) that can self-learn from the data. It is also capable of unsupervised learning. It can learn a large amount of unstructured and unlabeled data that even a human brain can take years to understand. Deep learning uses multiple layers to extract features from raw data. Deep learning models are based on artificial neural networks, and Convolutional Neural Network (CNN) is one of them. Architecture of simple neural network is shown in Fig. 1.

* Corresponding author.
** Corresponding author.
E-mail addresses: fareehaanwar@iiu.edu.pk (F. Anwar), amirhosein.mosavi@nmbu.no, amir.mosavi@kvk.uni-obuda.hu (A. Mosavi).

https://doi.org/10.1016/j.imu.2020.100482
Received 7 April 2020; Received in revised form 12 November 2020; Accepted 13 November 2020
Available online 17 November 2020

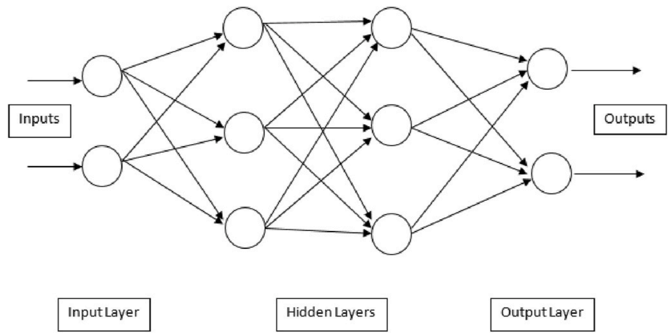

Fig. 1. Neural network architecture.

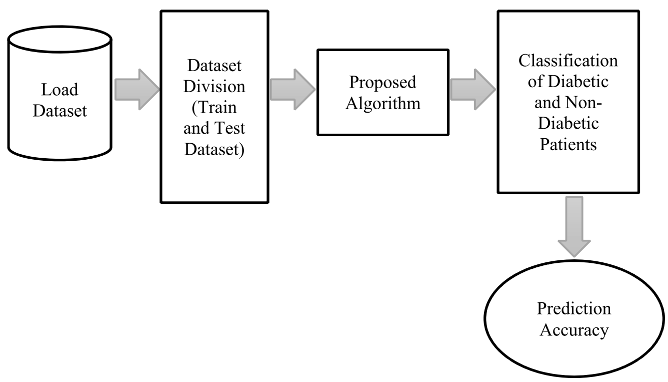

Fig. 2. General flowchart for diabetes diagnosis.

Fuzzy logic is a method of reasoning that is modelled upon human cognitive and analytical abilities. It involves both possibilities of YES or NO. A computer gives the output as TRUE or FALSE that in human language is equivalent to YES or NO.

Different techniques are used by researchers for diabetes diagnosis such as Backpropagation neural network (BPNN) [3]. Similarly, researchers in Ref. [4] show the performance of the Small World FANN model in diabetes diagnosis. Moreover, an artificial neural network-based approach is presented in Refs. [16].

Many researchers used Pima Indian Diabetes Dataset (PIDD) for a diabetes diagnosis. Pima Indian Diabetes Dataset consists of eight parameters. Those parameters include the number of times pregnancy has occurred, BMI, plasma glucose, diastolic blood pressure, systolic blood pressure, skinfold thickness, diabetic pedigree function, and Class 0 or 1 (0 means non-diabetic while 1 means diabetic patient). The literature review shows that PIDD might be the best dataset for diabetes diagnosis as it has a large number of values making it a standardized dataset. Other small datasets are also discussed in the literature for example data collected from patients directly, data collected through surveys, heart signals (ECG signals), CGM Signals, images dataset, Eye dataset, Skin dataset, and Ayurvedic dataset.

This paper presents a survey for diabetes diagnosis with some new contributions outlined below:

1) Literature Review is performed to analyze the existing latest approaches for diabetes diagnosis with some suggestions for future research.
2) Several related schemes from the last decade have been searched as per research questions and carefully studied to identify strengths and weaknesses.
3) The quality evaluation has been performed to verify articles linked with research questions.

The general flowchart for the diagnosis of diabetes is shown in Fig. 2.

The rest of the document is divided as follows: Section 2 throws light on the Literature Review of studied articles. Section 3 presents an analysis of the survey considering different evaluation measures, and Section 4 includes a comprehensive conclusion.

2. Literature review

The literature review helps us in identifying specific areas or research questions, or gaps in the literature that already exists.

2.1. Research question

The main objective of this research is to find a question for our research.

Question: "Is there any algorithm that has better accuracy using large dataset/Pima Indian Diabetes Dataset?"

2.2. Databases

Digital Libraries used are:

(i) Science Direct (www.sciencedirect.com/)
(ii) IEEE (www.ieeexplore.ieee.org/)
(iii) Springer (www.springerlink.com/)
(iv) Others (https://scholar.google.com.pk/)

2.3. Collection of study

This collection of articles for study are collected based on:

(i) Research articles on diabetes diagnosis
(ii) Research articles with available PDF
(iii) Research articles vary from the last decade
(iv) Articles based on surveys if required

Early Diabetes diagnosis is important for human health to saves them from the fatal effects of diabetes. In the past few years, different techniques have been introduced using a variety of models and approaches to diagnose diabetes. Those techniques include neural network-based approaches, deep learning approaches, and machine learning approaches, decision making approaches, k-NN approach, retinal images-based approaches and face image-based diagnosis techniques.

2.4. Neural network approaches

Researchers in Ref. [3] proposed Back Propagation Neural Network (BPNN). Graphical User Interface (GUI) was built in MATLAB. Pima Indian Diabetes Dataset is used by researchers to test their proposed methodology. Once loading of the dataset is completed, parsing was performed. After reading values one by one, they were stored to train ANN using Back Propagation Neural Network. In feature extraction phase values were classified with similar features and also arranging of groups was done in the column. Normalization was the next step of the proposed technique. Data values were represented within 0 and 1. Normalization removes data redundancy and guarantees data dependencies. The training was the last step of the proposed technique. Up to 9 iterations were performed to train the proposed system. The minimum error was found in the 3rd iteration. Best results were obtained at lower epoch values. Results were created using the regression plot and validation plot.

Feed Forward ANN (FFANN) becomes prominent in today's world because of its computational speed and efficiency. Researchers in Ref. [4] presented the performance of the Small World FANN model in diabetes diagnosis. For the investigation, researchers considered four-layered FFANN. There were eight inputs in the network that includes one output neuron. They used two hidden layers in FFANN. Two

Table 1
Results comparison table for neural network approaches.

Technique	Diabetes Diagnosis Accuracy
BPNN [3]	81%
SW-FFANN [4]	91.66%
ANN [16]	87.3%
ANN [12]	100%

different network topologies were used for FFANN. SW-FANN Activation function used by researchers in the proposed methodology is bipolar-sigmoid function. Backpropagation learning algorithm with training was used for the training process of SW-FFANN. The data set used for this research was PIDD taken from the UCI repository. The rewiring process was applied to the best regular topology for SW-network construction. DGlobal and DLocal parameters were calculated for each rewiring step.

The artificial neural network-based approach was presented in Ref. [16]. The artificial neural network has three main layers: input, hidden, and output layers. The input layer gets raw data. Hidden layers function is determined using inputs and weights assigned to them. The data was entered into a JNN tool that determines the values of attributes. Afterwards, training, testing, and validation of data were performed. The proposed system provided output in binary numbers. 0 as a diabetic patient and 1 as a healthy person. An average error rate of the proposed system was 0.010. The number of epochs performed on the dataset was 158,000. Samples for training data were 767, and samples to validate the system were 237.

Researchers in Ref. [12] used skin impedance and heart rate variability for the detection of diabetes. Artificial neural networks were used for classification. Skin impedance data were collected from 11 patients having diabetes that include six females and five males with an average age of 40 ± 8 years. Also, data of eight normal persons were collected that includes five females and three males with an average age of 24 ± 3 years. To measure signal power at different frequencies, Welch Method was used. ECG data was collected from 20 healthy persons including fourteen males and six females with an average age of 22 ± 7 years. Also, data of 20 diabetic patients were collected including eight females and twelve males with an average age of 40 ± 8 years. Preprocessing was performed on raw ECG signal removing baseline drift in a signal using median filtering. The noise of high frequencies was also removed using butter worth a lowpass filter. Then smoothing of the ECG signal was performed using the Savitzky-Golay filter.

Table 1 briefly explains different neural network approaches for the diagnosis of diabetes. All approaches show better results, but ANN [12] outperforms all other neural network approaches.

2.5. Machine learning approaches

ANFIS was proposed in Refs. [1] that was based on Sugeno FIS. The proposed methodology was the hybridization of an artificial neural network and Fuzzy inference system having a learning ability. Features were adapted from the Artificial Neural Network. ANFIS comprised of two parts antecedents, and conclusion. It consists of five layers having its own functionality. X and Y were values of input against nodes, while fuzzy sets were represented as Ai and Bi. The triangular membership function was used in the proposed techniques. The output of the first layer becomes the input of the second layer. Normalization of data was performed in the third layer. Datasets used to perform experiments were taken from the locals of Bhubaneswar, Odisha, India. Levenberg-Marquardt backpropagation algorithm was used to train the ANN system.

Researchers in Ref. [11] used Pima Indian Diabetic Dataset to classify diabetic patients and diabetes diagnosis using different machine learning techniques. To classify diabetic patients and normal persons, some sets of characteristics were used that are selected according to WHO criteria. Researchers use those sets of characteristics as features vectors. Feature vectors were composed of all eight features from a selected dataset. Three stages of the evaluation were performed by researchers [11]. The first one showed a comparison of the state of diabetic and non-diabetic peoples. The second evaluation stage used hypothesis testing to check if the feature vector showed different distributions for diabetic and non-diabetic patients. In the last stage classification, the analysis was performed to assure whether all features can discriminate between diabetic patients and non-diabetic patients. Machine learning classification algorithms like J48, JRip, MultilayerPerceptron, RandomForest, HoeffdingTree, and BayesNet were used. Weka tool was used for performing classification analysis. The null hypothesis got rejected by all eight features, statistically showing that all these features can distinguish between diabetic and non-diabetic patients.

Five different techniques of machine learning were used in Ref. [15] for diabetes diagnosis and preprocessing of data. Those techniques include DNN, Logistic Regression, Decision Tree, SVM, and Naïve Bayes. Those techniques were used on Pima Indian Diabetic Dataset to calculate the accuracy of cross-validation. Five preprocessing steps were performed on the dataset. After each step, the accuracy of all algorithms was calculated and compared. Those data preprocessors include imputation, scaling, normalization separately, imputation and scaling, imputation and normalization. Imputation was the process to calculate the missing values of a dataset. After performing data preprocessing steps, the comparison of the results showed that Naïve Bayes and Decision Tree performed the same on the original dataset and the scaled dataset in terms of accuracy. All other classifiers also showed good results in terms of accuracy on a scaled data set.

Different machine learning models: k-NN, Naïve Bayes, Decision Tree, Random Forest, SVM, and logistic regression were used in Ref. [13] to identify type 2 diabetes using electronic health records. From the total number of 23,281 diabetes-related patients, 300 samples were selected. All samples were un-labelled. Two clinical experts were called to label the dataset. From 300 samples, 161 were typed 2 diabetic patients, 60 were non-diabetic patients and 79 samples were unconfirmed. 78.3% of samples were incomplete those 79 samples were dropped. The feature construction model was used to convert that electronic health records (raw data) into statistical features so that it can be used as input for classification models. Related features were summarized using summation to form new features. From 36 features, eight features were extracted using feature summarization. These features were used as input for classification models like k-NN, Naïve Bayes, Decision Tree, Random Forest, SVM, and logistic regression. Also, the ability to diagnose type 2 diabetes was tested using the same classification models. Weka tool was used to apply those classification models on a dataset. Proposed classification model performance based on parameters such as accuracy, precision, specificity, sensitivity, and AUC.

The graph-based approach was proposed in Ref. [9] to classify the retinal image. Retinal vessels are of 2 types' veins and arteries. The most important phase is the extraction of retinal vessels to detect vascular changes. The retinal image of a patient was used to calculate the artery vein ratio. Diabetes recognition was done using an artery-to-vein ratio. The implementation of the proposed system was done in different stages. The first one was the preprocessing. Extraction of the green channel from scanned retinal images was done in this stage. This stage improved unprocessed image quality by removing noise and eliminating irrelevant information. The Green channel image was calculated using equation Eqn (1) [9].

$$g = \frac{G}{R + G + B} \tag{1}$$

Enhancement was used to clear an image. Edge detection was the next stage of the proposed technique. Edge detection techniques were applied to the retinal image to extract blood vessels. Researchers in Ref. [9] used the canny edge detection technique. Kirsch template was

Table 2
Results comparison table for machine learning approaches.

Technique	Results
ANFIS [1]	90.32% Accuracy of ANFIS
J48, MLP, HoeffdingTree, JRip, BayesNet, RF [11]	HoeffdingTree precision 0.770 and recall 0.775
DNN, SVM [15]	DNN accuracy 77.87%
k-NN, Naïve Bayes, DT, RF, SVM, LR, Expert algorithm [13]	AUC 0.98
RF, SVM, Binary Tree, Adaptive Boosting, Generalized Linear, NN [19]	89.63% accuracy using RF
AdaBoost algorithm with base classifiers [35]	80.729% accuracy using AdaBoost algorithm with Decision Stump classifier

used to identify the presence of edge and finally to extract blood vessels from the retinal image. Graph-based methods were applied for retinal vessel classification. Graphs were represented using links and nodes. Object detection was done from different images after extracting unique features. To detect damaged parts, the MSER algorithm was used. Classified images and extracted features were considered as input. Row and column-wise values of the image were compared. Any part having maximum value was considered diabetes. The proposed methodology showed 88% accuracy.

Iris images and machine learning techniques were used in Ref. [19] to diagnose type 2 diabetes. For this purpose, 338 subjects were considered, 180 out of them were diabetic, and 158 were non-diabetic patients. Subjects were selected on three factors that include: gender ratio, standard deviation, and diabetes duration age (vary from 1 to 25 years), and average age. Iris images were attained using I-SCAN-2. Gray infra-red images of size (left and right iris) 640 × 480 were acquired. Using the iris image, suitable features were extracted from regions of interest. Inner and outer boundaries of iris were used in segmentation. Rubber-sheet normalization was used to plot extracted iris into a fixed rectangle. Region of interest was cropped from iris according to the tail, head, and body pancreas organ. A threshold was then applied to generate the edge map. Centre point and radius of pupil were considered as main parameters. For each feature, the scoring criteria was calculated. Different machine learning algorithms were used by researchers for classification. Those algorithms include SVM, Naïve Bayes, Random Forest, NN, Adaptive boosting model, and generalized linear model.

A decision support system was proposed in [35] that used the AdaBoost algorithm with Decision Stump as a base classifier for classification. The proposed methodology was implemented in four different phases. Local and global dataset collection was performed. The global dataset was used for training and testing, a local dataset was used. The dataset used for this research was collected from various places in Kerala, India. Pima Indian Diabetes Dataset was considered as a global dataset, while the dataset collected from Kerala was considered as a local dataset. Missing values in the local dataset were fulfilled by replacing them with the mean value. In the second phase, AdaBoost algorithm was applied to a global dataset to train the proposed system. Different base classifiers (SVM, NB, Decision Stump, and DT) were also used along with the AdaBoost algorithms. In the third phase, the validation of the proposed system was achieved using local dataset. Finally, the accuracy of AdaBoost algorithm with base classifiers was calculated. AdaBoost algorithm with Decision Stump as a base classifier showed the best accuracy of 80.729% for diabetes prediction. Also, it showed less error rate.

Table 2 briefly explains different machine learning approaches for a diabetes diagnosis. ANFIS [1] found to be more accurate than others, but with a smaller dataset.

2.6. Deep learning approaches

Researchers in Ref. [21] proposed a diagnosis of diabetes using HRV signals taken from ECG signals. CNN and CNN-LSTM were employed in combination with the automatic detection of diabetes. Deep learning networks have embedded feature extraction, feature selection, and classification. Deep Learning has the advantage of self-learning using data. CNN consists of three layers: Convolutional Layer, Pooling Layer, and the last one is the fully connected layer. The last layer has a ReLU activation function. Convolutional layer output is given to the pooling layer. The activation function used by the convolutional layer is ReLU that applies max (0, x) to every input to ReLU symbolized by x. The main function of the pooling layer is to perform a downsampling operation. LSTM is an improved form of RNN. To handle vanishing and exploding gradient problems, it uses memory blocks instead of convolutional simple Recurrent Neural Network units. Long Short-Term Memory can handle long term dependencies in a better way than traditional RNN.

Table 3
Results comparison table for deep learning approaches.

Technique	Results
CNN and CNN-LSTM [21]	CNN-LSTM accuracy of 95.1% using 5-fold cross-validation
LR, MLP and CNN [6]	CNN accuracy 77.5%
Deep Neural Network [24]	5-fold cross-validation: 98.35% accuracy, F1 of 98 and MCC of 97. 10-fold cross-validation: 97.11% accuracy, the sensitivity of 96.35% and specificity of 98.80%

Researchers proposed a deep learning approach in Ref. [6] for the detection of Type 2 diabetes. Logistic Regression, Multi-Layer Perceptron, and Convolutional Neural Network were applied to CGM signals collected from 9 patients. After producing CGM signals, the dataset was divided into training and testing data. 1–6 patients CGM signals were used as a training set for all three classifiers, and after that, CGM signals of 7–9 patients were used as a test dataset for all classifiers. ReLU was used as an activation function in a hidden layer in MLP. For CNN, convolutional layers consist of three layers, with every layer having ReLU as its activation function shadowed by max-pooling layers. In the feature selection phase, investigated filter sizes and filters were 6, 12, 18, and 8, 16, 32, 64 respectively. 10 and 50 units were used in fully connected layers. Among different values and combinations for every model, the best combination was in the CNN model that has LR = 10–4, convolutional layer = 2, Filter size of 18, 8 number of filters, and no. of units in fully connected layer = 10.

The patient's largest dataset of diabetes was introduced in Ref. [22]. This dataset includes records of over 14 thousand patients. Different deep learning models were applied to a dataset that includes LSTM and GRU for Type 2 diabetes detection. Dataset was collected from 2010 to 2015. Data preprocessing was performed on the dataset. A set of measures were used to describe every patient's visit. Episodes were used to represent those measures. Each sequence had 30 features. KAIMRC dataset was used to train LSTM and GRU. Results were compared with MLP models. LSTM and GRU achieved better results than MLP models for different data inputs with sizes ranges from 3 to 15. For longer dependencies, LSTM outperforms others. While on short sequences, GLU performed better.

To reduce overfitting, a prediction system with dropout was proposed [23] for diabetes prediction. A deep learning neural network model that had fully connected layer plus dropout layers; was proposed by researchers. Pima Indian diabetic dataset was used to train and test the proposed system. Firstly, the dataset was entered as an input to the input layer. After that, two fully connected layers were used, and each layer had a dropout layer. After passing the dataset from these layers the result of the system was obtained through output. Those layers made that system as MLP.

A deep neural network with training in five-fold cross-validation and ten-fold cross-validation was proposed in Refs. [24] to diagnose diabetes. Pima Indian Diabetes Dataset was used for diagnosis. Data was collected from the UCI machine repository database. After data collection, a dataset was divided into five-fold and ten-fold cross-validation. The proposed methodology had four hidden layers, and the number of neurons in hidden layers were 12, 16, 16, and 14. The best outcome was achieved using this combination. There were eight input layers consist of eight attributes and one output layer was used to predict diabetic or no-diabetic patients in binary form. ReLU was used as an activation function.

Table 3 briefly explains different deep learning approaches for a diabetes diagnosis.

2.7. Hybrid approaches

Fuzzy Deep Learning approach adopted in [26] for the prediction of the diabetic. First, Fuzzification of data was done, and then that data was given as input to CNN. Ayurvedic dataset of Indian Population was used. That dataset was collected by taking interviews of various patients. A total of 150 samples were collected. In the preprocessing phase, normalization was performed on the dataset. Subsequently, Fuzzification applied to assign a range of values to each attribute. Every sample converted into the matrix. Columns of the matrix represented features and rows of the matrix fuzzy values of features. Fuzzification was performed in a way that, in the matrix, every feature provided a minimum of 10% overlapping. After converting all data into matrix form, the fuzzy matrix was given as input to CNN. Researchers performed three experiments. Two experiments were performed using the Neural Network, while the last one was performed using CNN. Values of α set to 2 and 5 and a total of 2000 iterations were performed for every experiment. Matrix size for convolutional and pooling layers of CNN was 3 × 3 and 2 × 2. Hybrid Fuzzy-CNN performed better than the Neural Network approach.

Researchers in Ref. [25] also used a hybrid fuzzy deep learning approach for diabetes detection. Dataset was taken from the National Institute of Diabetes and Digestive and Kidney Diseases. At the start, the fuzzification of data was performed. Afterwards Fuzzy matrix of 5 × 5 was formed where columns of the matrix represented features and rows of the matrix represented the fuzzy value of features. The fuzzy matrix was then given to CNN as input. Three experiments were performed. Two experiments were conducted using the Neural Network, while the last one was done using CNN. Values of α were considered as 2 and 5, and a total of 2000 iterations were performed for every experiment. Matrix size for convolutional and pooling layers of CNN was 3 × 3 and 2 × 2 respectively. Hybrid Fuzzy-CNN performed better than the Neural Network approach.

Hybridization of SVM and statistical modelling of Naïve Bayes [36] was used for the prediction of a diabetic. Dataset was composed of 402 patients. New attributes were also introduced that were not used previously. The SVM algorithm was used to represent the occasion's occurrence in space as points. In that way, different classes were displayed with strong gaps. The main aim of SVM was to minimize weight. A statistical model of Naïve Bayes was used for prediction which used a linear function. Dataset collected for the proposed system comprised of independent attributes having equal importance. The probability of record Y that belongs to a class C can be calculated as Eqn (2) [36].

$$P\ (Y = C) = \prod_{i=1}^{q} P\ (Xi\ |Y = C) \tag{2}$$

The output of the system was in binary form either 0 or 1. Where 0 showed normal, and 1 showed a diabetic patient. The data was considered as unclassified and in the gray zone if the output occurred other than 0 or 1. SVM achieved 95.52% accuracy while Naïve Bayes had 94.53% accuracy.

2.8. Using Heart Rate Signals

Researchers in [39] proposed digital signal processing methods for automatic detection of diabetes using ECG signals. Digital signal processing methods are used to extract features from heart rate (HR) signals and those features are used to diagnose diabetes. Useful features identified through a statistical analysis were Poincare geometry properties (SD2) and recurrence plot properties (REC, DET, LMean). Those important features contributed well in differentiating diabetic and non-diabetic features using HR signals. For validation of the proposed method, the AdaBoost classifier was used in combination with a perceptron weak learner. After that, a novel diabetic integrated index DII was developed. The accuracy of the proposed system was 86%. DII showed that the HR signal was of diabetic patients. It also helped in automatic diabetes detection.

A dataset of 15 diabetic and 15 non-diabetic persons ECG was used for diabetic prediction [40]. Time-domain extracted parameters were HR (Mean HR), HF, statistical parameters (NNN50, PNN50), and histogram parameters (HRV Δ Index). Time-domain cannot differentiate if the HRV signal is sympathetic or parasympathetic. So, frequency domain analysis was used to overcome this disadvantage. It included a power spectrum density (PSD) estimate to analyze HRV signals. Nonlinear methods including Poincare plots, recurrence plots, correlation dimension, approximate and ample entropies were used to quantify the dynamics of HR fluctuations. Results showed that nonlinear analysis of HRV was best among all three analyses. Clinically significant nonlinear parameters were correlation dimension approximate entropy, sample entropy, and recurrence plot properties.

Linear methods are unable to find hidden information in signals, to overcome this a new nonlinear method based on empirical mode decomposition (EMD) was proposed [41]. This method was used to differentiate between diabetic and non-diabetic patients using RR-interval signals. Mean frequency using Fourier-Bessel series expansion along with two bandwidth parameters, amplitude modulation bandwidth and frequency modulation bandwidth, used in research. These bandwidth features were extracted from intrinsic mode functions obtained from EMD of RR-0 interval signals. A unique representation was also given to differentiate between 2 groups. Overall, five features were extracted using IMFs. Results proved that those features can efficiently differentiate between diabetic and non-diabetic patients.

A non-invasive diagnosis support system was used for a diabetes diagnosis [42]. Examination of the heart health of a person using HRV analysis can identify a diabetic patient or not. Nine nonlinear features i.e. approximate entropy, largest Lyapunov exponent, detrended fluctuation analysis, and recurrence quantification analysis were used to perform this analysis. Clinically significant measures used as input for classification algorithms (AdaBoost, DT, Fuzzy Sugeno classifier, k-NN, probabilistic neural network, and SVM). 10-fold cross-validation was used to identify the best classifier among all. AdaBoost achieved the best accuracy of 90%, the sensitivity of 92.5%, and specificity of 88.7%.

2.9. Other approaches

Researchers in Ref. [20] proposed a model having three different machine learning algorithms for diabetes prediction. Those machine learning algorithms include Decision Tree, SVM, and Naïve Bayes. Decision trees predicted target class using rules of previous data. It chose each node calculating the highest information gain. Pima Indian Diabetic Dataset taken from the UCI repository available online was used in this research.

Automatic Multi-Layer Perceptron (MLP) with a combination of outlier detection methods was proposed in Ref. [14]. In this model, small MLPs were ensemble having auto-tuning quality, that automatically adjusts the parameters. Pima Indian Diabetes Dataset was used to test proposed systems. The preprocessing of the diabetes dataset was done by detecting outliers with the help of enhanced class outlier-based methods. Ten outliers were detected with 12 nearest neighbors. After data preprocessing, Automatic MLP was used for the classification of diabetic patients. A total of 4 MLPs with several numbers of hidden layers and learning rates were used. The error rate was determined after ten training cycles.

A model having K-means for data reduction and J48 (Decision Tree) as a classifier was introduced in Ref. [10]. In the first step missing and impossible values of the dataset such as BMI = 0 etc. were replaced by calculating mean values. In the second step, for the removal of incorrectly classified samples, the K-means algorithm that was implemented using WEKA. In the third step, classification of patients was performed using J48 Decision tree algorithm having a 10-fold cross-validation method. Lastly, the performance was evaluated. Accuracy, specificity, and sensitivity were used as evaluation measures using True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Another performance measure was the confusion matrix. TP and TN were used to represent correctly classified samples while FP and FN represent misclassified samples. For the proposed methodology, 10-fold cross-validation was used. Pima Indian Diabetes Dataset (PIDD) available at UCI repository was used.

A new methodology composed of DM Genetic Algorithm, Gray-Scale Histogram features, and k- Nearest Neighbor classifier was proposed in Ref. [5]. Detection of diabetes was performed using four facial blocks extracted from the facial image. Genetic Algorithm (GA) has proved itself as an efficient and effective searching method. GA helped to select new features using a Gray-Scale Histogram. To extract features from facial blocks, GHF was used. Those features kept most of the block information. Distribution of four facial was as follows: Block A represented forehead, Block B and D represented left and right eye area respectively while Block C lied in the middle of block B and D that is the nose. Four facial blocks contained the skin of the contestant. there No other shapes or edges were present in those blocks. The values of all those blocks should be in the same range. GHF was used to extract the range of every block. It calculated the frequency of every gray-scale value of blocks. DGMA was proposed to remove redundancy from GHF and to keep valuable information. Genetic Algorithm uses population fitness to select individuals. Crossover and mutation were performed in each generation to generate new children. To detect diabetes mellitus, k-NN classifier along with weights named k–NN–W was used. Each block of the facial image was assigned with weights.

To classify Gestational Diabetic and Gestational Non-Diabetic patients using real-time data, RBFNN was proposed in Refs. [2]. As ANN are adaptive, they learn using examples. RBFNN is widely used for the control and classification of curve-fitting problems. RBF network is a type of ANN that uses RBF as an activation function. Feed Forward NN consists of an input, hidden, and output layer. In the proposed method, inner layer outputs were determined by calculating the distance formula between hidden layer centres and input. I didn't have a nonlinear function in the hidden layer. Instead, it had a linear function at the output layer. Gaussian bell function was used in it. Changing these two things in RBF network architecture made its performance different from that of the RBF network. It used a single hidden layer to exhibit nonlinear functions. Some of the advantages of the proposed technique were fast training, simpler architecture, powerful mapping capability, and cost-effectiveness. No need to take blood tests as using records of patients of different hospitals as they collected real-time data. The real-time dataset consisted of 188 records and ten parameters. Data was taken from the patient's records from Jan 2013 to May 2013.

A model based on data mining techniques for the prediction of Type 2 diabetes mellitus was proposed [17]. The proposed model had two parts: detection using the k-means algorithm, and by logistic regression. Pima Indian Diabetes Dataset was used to test proposed systems. Preprocessing of data was performed in WEKA using different built-in filters. Firstly, to reduce dataset complexity, the medical implication of every attribute was analyzed along with correlation with diabetes mellitus. Missing and incorrect values that occur because of errors were also removed. An unsupervised normalized filter was used for attributes to normalize the data. To remove incorrectly clustered data, and the improved k-means algorithm was used. A logistic regression algorithm was used to predict diabetic and non-diabetic patients. The proposed model was evaluated on k-fold cross-validation, detailed accuracy, and Kappa statistics. Researchers in Ref. [16] used a 10-fold cross-validation method.

Table 4
Results comparison of other approaches for diabetes diagnosis.

Techniques	Results
Decision Tree, SVM, Naïve Bayes [20]	Naïve Bayes Accuracy 76.30%
K-means for data reduction with J48 decision tree as a classifier [10].	Accuracy 90.04% Sensitivity 87.27% Specificity 91.28%
k-Nearest Neighbors) with weights to detect DM using four facial blocks extracted from the facial image [5]	Accuracy 99.48%
RBFNN [2]	The efficiency of RBF networks is 68.23%
Improved k-means and LR [17]	The model achieves 3.04% higher prediction accuracy
RF and Gradient Boosting Classifiers [18]	90% accuracy, specificity, sensitivity and AUC
Empirical Mode Decomposition Technique [8]	Accuracy 95%
E-Nose Hardware [7]	Accuracy 95.0% The precision of Diabetes 91.30% Precision of Healthy 94.12% Kappa statistic's value 0.898
Bayesian Network [37]	Accuracy 99.51%
Feature Selection, SVM [38]	Accuracy 98%

A tree-based ensemble learning model was introduced for automatic diabetes prediction [18 Random Forest and Gradient Boosting used for classification. Pima Indian Diabetes Dataset was used to test proposed systems. Dataset consisted of 768 instances, among which 268 were positive diabetic, and 500 instances were of non-diabetic patients. After collecting the dataset, preprocessing, and cleaning of data was performed. In the preprocessing step, data points having zero or null values for features$\geq$3 were removed. Secondly, all those zero values were replaced by mean values calculated from all other data. Outliers were also detected and removed using the k-NN approach.

A nonlinear method based on EMD was proposed to distinguish between diabetic and normal R-R interval signals [8]. The SVM was used for prediction. Parameters acquired from ECG signals used as a feature set for SVM classifiers. Unwanted noise in ECG was removed using a bandpass filter. Pan and Tompkins Algorithm made a great impact on R-R interval 15 detections. In the proposed algorithm, a special digital bandpass filter was implemented. This helped in reducing false detection that occurs due to different interference types in ECG signals. The proposed algorithm can automatically modify parameters and thresholds to get used to the changes in QRS complexes. SVM was used as a classifier to detect diabetes. The dataset consisted of 50 ECG signals from which 33 were healthy, and 17 were of diabetic patients.

An E-Nose technique was proposed in Ref. [7] where human breath analysis of gas signal data was used to detect diabetes. That gas signal was captured using electrochemical sensors which were connected to microcontrollers (e-Nose). The proposed technique had seven stages: Making of E-Nose, collecting ground-truth data, data preprocessing, feature extraction, feature selection, classification, and evaluation. A collection of ground-truth data was performed to collect training data. The blood glucose level (BGL) of non-fasting patients was calculated for diabetic prediction. Patients who had BGL below 120 mg/dL were considered healthy, while patients having BGL above 150 mg/dL were considered as diabetic patients. To collect ground-truth data, the patient breathed for about 150 s using an e-nose, and it was recorded using a laptop connection. Preprocessing was further divided into two phases: Signal Diagnosis and Feature Scaling (Normalization). Signal diagnoses helped in making the sensitivity and accuracy of e-Nose better. Normalization helped to make features rescaled to have standard

Table 5
Comparison of Dataset and Tools used by Researchers.

Author	Techniques	Dataset	Tools/ Languages
Aparimita Swain et al., 2016	ANFIS [1]	Self-collection of data of 100 peoples	MATLAB 2013
Priya Shirley et al., 2016	RBFNN [2]	Self-collection of data of 188 records	MATLAB R2010a
Miss. Sneha Joshi et al., 2016	BPNN [3]	Pima Indian Diabetes Dataset	MATLAB 2015
Okan Erkaymaz et al., 2016	SW-FFANN [4]	Pima Indian Diabetes Dataset	–
Ting Shu et al., 2016	*k*-Nearest Neighbors) with weights to detect DM using four facial blocks extracted from the facial image [5]	Facial Images Dataset	–
Ali Mohebbi et al., 2017	LR, MLP and CNN [6]	CGM Signals	MATLAB
Hariyanto et al., 2017	E-Nose Hardware [7]	E-Nose	4 sensors/ Arduino MEGA 2560 + MATLAB
Reena Musale et al., 2017	Empirical Mode Decomposition Technique [8]	ECG Signals of 50 persons	
R.S. Mangrulkar et al., 2017	A graph-based approach for retinal image classification [9]	Scanned Retinal Images	MATLAB
Wenqian Chen et al., 2017	K- Means for data reduction with J48 decision tree as classifier [10]	Pima Indian Diabetes Dataset	WEKA
Francesco Mercaldo et al., 2017	J48, MLP, Hoeffding Tree, JRip, BayesNet, RF [11]	Pima Indian Diabetes Dataset	WEKA
Tarak Das et al., 2017	ANN [12]	ECG Signals + Skin Impedance	Kubios Software (Version 2.2) + MATLAB
Tao Zheng et al., 2017	k-NN, Naïve Bayes, DT, RF, SVM, LR, Expert algorithm [13]	Electronic Health Records	WEKA
Maham Jahangir et al., 2017	Auto MLP [14]	Pima Indian Diabetes Dataset	–
Sidong Wei et al., 2018	DNN, SVM [15]	Pima Indian Diabetes Dataset	–
Nesreen Samer et al., 2018	ANN [16]	Pima Indian Diabetes Dataset	JNN Tool Environment
Han Wu et al., 2018	Improved k-Means and LR [17]	Pima Indian Diabetes Dataset	WEKA
Sujit Kumar Das et al., 2018	RF and Gradient Boosting Algorithm [18]	Pima Indian Diabetes Dataset	–
Piyush Samant et al., 2018	RF, SCM, Binary Tree, Adaptive Boosting, Generalize Linear, NN [19]	338 samples of Iris Images	–
Deepti Sisodia et al., 2018	Decision Tree, SVM, Naïve Bayes [20]	Pima Indian Diabetes Dataset	WEKA
Swapna G et al., 2018	CNN and CNN-LSTM [21]	ECG of 20 diabetes and 20 normal persons and 71 datasets of diabetic and 71 datasets of normal persons where each dataset has 1000 samples	–
Zakhriya Alhassan et al., 2018	LSTM and GRU [22]	KAIMRC	–
Akm Ashiquzzaman et al., 2018	Deep Learning Neural Network with dropout [23]	Pima Indian Diabetes Dataset	Python
Safial Islam Ayon et al., 2019	Deep Neural Network [24]	Pima Indian Diabetes Dataset	Python
Tushar Deshmukh et al., 2019	Fuzzification and CNN [25]	National Institute of Digestive and Kidney Diseases	–
Tushar Deshmukh et al., 2020	Fuzzification and CNN [26]	Ayurvedic dataset of Indian Population	–
Veena Vijayan V. et al., 2015	AdaBoost algorithm with base classifiers [35]	Pima Indian Diabetes Dataset	MATLAB and WEKA
Zhilbert Tafa et al., 2015	SVM and statistical modelling of Naïve Bayes [36]	Dataset of 402 instances taken from 3 different locations in Kosovo	MATLAB
Mukesh Kumari et al., 2014	Bayesian Network [37]	Self-collected from hospital	WEKA
Prof. Neilesh B. et al., 2014	Feature Selection, SVM [38]	Pima Indian Diabetes Dataset	–

properties distribution. Four statistical features, standard deviation, minimum value, average and maximum value were calculated. K-NN classifier was used for diabetes detection.

Researchers in [37] proposed a Bayesian network classifier to diagnose diabetes. The dataset used was collected from hospitals. The Bayesian network a graphical model was used that works on probability. WEKA tool was used to implement proposed techniques. The dataset consisted of nine attributes and 206 records. Preprocessing was performed on a dataset to identify attributes and selection of those attributes. Data normalization was performed. 99.51% accuracy was achieved using a Bayesian network classifier algorithm. The error rate is also reduced to 0.48%.

A new feature selection method along with SVM was proposed in [38]. The feature selection method was considered as one of the best methods to improve prediction accuracy, prediction efficiency, and reduce complexity. Methods used for feature selection include K-means clustering and F-score. Pima Indian Diabetes Dataset taken from the UCI Repository was used to test and train the proposed techniques. F-score showed better classification performance than other methods of feature selection. SVM achieved 98% accuracy, 97.77% sensitivity and 97.79% specificity.

Table 4 briefly discusses different approaches used for diabetes diagnosis and the comparison of their results.

3. Analysis

This section presents the analysis of various methods used for diabetes diagnosis discussed in Section 2. Different approaches performed better in terms of accuracy on different datasets. Still, a few disadvantages were there that are discussed in this section. Many researchers used the Pima Indian Diabetes Dataset for a diabetes diagnosis. Thus, this section also shows results comparing different approaches using the same Pima Indian Diabetes Dataset.

3.1. Dataset and tools comparison

Different research articles are considered for this research. Among them, few researchers used a dataset collected by themselves or different kinds of datasets. Fourteen researchers used the Pima Indian Diabetes Dataset taken from the UCI repository [31] for a diabetes diagnosis. It comprises eight variables having 768 instances. All feature values are

Table 6
Analysis of neural network approaches for diabetes diagnosis.

Technique	Drawback/Comments
BPNN [3]	The actual performance of back-propagation on a particular program is dependent on input data. It can be sensitive to noisy data and outliers. Fully matrix-based approaches to back-propagation over a mini-batch. Slower to train data and a lot of iterations are there [3].
SW-FFANN [4]	This study was the first attempt to construct an SW-FFANN for the diagnosis of diabetes. Therefore, it seems crucial for the diagnosis of diabetes in human life [4].
ANN [16]	Classification boundaries of ANN are hard to understand intuitively, and ANNs are computationally expensive.

Table 7
Analysis of machine learning approaches for diabetes diagnosis.

Technique	Drawback/Comments
ANFIS [1]	a) Grid partitioning used to generate FIS Structure is only suitable for applications with a small no. of input variables. b) In ANN, performance reduces if data is too long [1].
J48, HoeffdingTree, MLP, BayesNet, RF [11]	Slow initial learning is an issue with HoeffdingTree [32]
DNN, SVM [15]	DNN has shortcomings like comparatively more computational time and more adjustment of parameters [15].
k-NN, Naïve Bayes, DT, RF, SVM, LR, Expert algorithm [13]	Manual validation can be time-consuming and expensive. [29]
RF, SVM, Binary Tree, Adaptive Boosting, NN Generalized Linear [19]	SVM shows better accuracy but expensive in terms of computational time [30]

Table 8
Analysis of deep learning approaches for diabetes diagnosis.

Technique	Drawback/Comments
CNN and CNN-LSTM [21]	Dataset used is small in size, so the proposed system may not give perfect results on a large dataset [21].
LR, MLP and CNN [6]	Not feasible in terms of Accuracy, Scalability, and System Optimization [28]
Deep Neural Network [24]	DNN has shortcomings like comparatively more computational time and more adjustment of parameters [15].

taken from females having age 21 and above. Those eight variables include:

1. Number of Times Pregnant
2. Plasma Glucose Concentration
3. Diastolic Blood Pressure
4. Skin Fold Thickness
5. 2-hour Serum Insulin
6. Body Mass Index (BMI)
7. Diabetes Pedigree Function
8. Age

The glucose level in the human body, body mass index, and age, along with other factors, play the most important role in the prediction of diabetes. Skinfold thickness and number of pregnancies are the least important factors in diabetes prediction. Pima Indian Diabetes Dataset has all these three main factors in it, that is why it is considered as a standard dataset for diabetes prediction, and most of the researchers used PIDD for diabetes prediction and achieved better results, as shown in Table 10. Few researchers used a dataset which they collected themselves from different hospitals or surveys, etc.

Comparison of Dataset and Tools/languages used by different researchers for diabetes detection is discussed in Table 5.

Table 9
Analysis of other approaches for diabetes diagnosis.

Techniques	Drawback/Comments
ANN [12]	Early detection of diabetes is not possible using heart rate variability and skin impedance [12].
Decision Tree, SVM, Naïve Bayes [20]	SVM: It is a binary classifier. For classification of multiclass, it can use pairwise classification. Computational cost is high, so it runs slow. Naïve Bayes: It is lazy as they store entire training examples. [27]
K-means for data reduction with J48 decision tree as a classifier [10].	a. The proposed model is proposed to apply to the Type 2 diabetes diagnosis that is a two-class classification problem. It may not work properly on multiclass classification problems. b. Proposed model is only applied to numeric data [10].
k-Nearest Neighbors) with weights to detect DM using four facial blocks extracted from the facial image [5]	The proposed model gives perfect results on a small facial image [5].
RBFNN [2]	For classification problems, traditional NN can get better classification results with much more efficient networks than RBF networks. RBF networks are recommended for surfaces with regular peaks and valleys. An unnecessary increase in basis function can increase computational complexity [2].
Improved k-means and LR [17]	Using LR problems occurs when a large number of features and a good chunk of missing data is present in the dataset. Too many categorical variables are also a problem for LR [33]
RF and Gradient Boosting Classifiers [18]	Using RF training, a large no. of deep trees can have high computational costs (but can be parallelized), use lots of memory [34]
Empirical Mode Decomposition Technique [8]	Early detection of diabetes based on the ECG signal is not possible [12].
E-Nose Hardware [7]	The proposed system cannot detect pre-diabetes patients having blood glucose between 120 mg/dL to 150 mg/dL due to a lack of ground truth data [7].

Table 10
Analysis of different approaches for Pima Indian diabetes dataset (PIDD).

Techniques	Results in terms of accuracy
BPNN [3]	81%
SW-FFANN [4]	91.66%
K-Means for data reduction with J48 decision tree as classifier [10]	90.04%
J48, MLP, Hoeffding Tree, JRip, BayesNet, RF [11]	–
Auto MLP [14]	88.7%
DNN, SVM [15]	77.8%
ANN [16]	87.3%
Improved K-Means and LR [17]	3.04% higher
RF and Gradient Boosting Classifiers [18]	90%
Decision Tree, SVM, Naïve Bayes [20]	76.30%
Deep Learning Neural Network With Dropout [23]	88.41%
Deep Neural Network [24]	98.35% for 5-fold cross-validation 97.11% for 10-fold cross-validation
Feature Selection, SVM [38]	98%

3.2. Critical analysis of previous techniques

Table 6 shows the analysis of different neural network approaches used by researchers for a diabetes diagnosis. These approaches performed better in terms of accuracy but still have drawbacks.

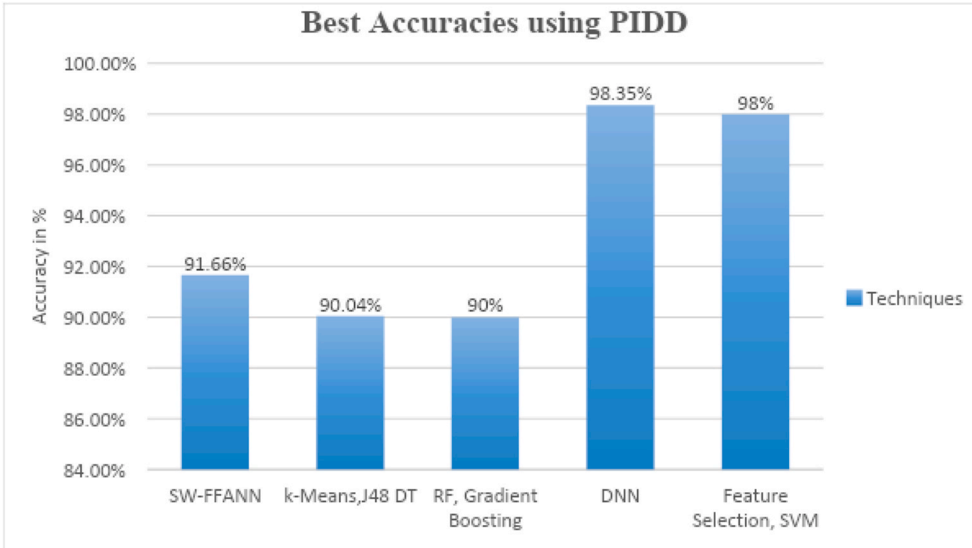

Fig. 3. Best Algorithms Accuracies for Diabetes Prediction using PIDD.

Table 7 shows the analysis of different Machine Learning approaches used by researchers. Different machine learning approaches used for diabetes diagnosis are considered for this research. Those machine learning approaches also have some drawbacks that are discussed in Table 6.

Analysis of different deep learning approaches is performed in Table 8. Deep learning approaches used for diabetes diagnosis also have few drawbacks.

Researchers also used some other approaches for the diagnosis of diabetes. Those approaches also performed better in terms of accuracy but still, they suffer at some point with different drawbacks that are briefly discussed in Table 9.

Table 10 shows results in comparison in terms of the accuracy of different techniques applied to the same Pima Indian Diabetes Dataset. Results comparison shows that Deep Neural Network [24] achieves the highest accuracy among all. Deep Neural Network faces shortcomings like computational time, and it also requires more time for parameter adjustments [15].

Literature Review of Heart Rate Signal papers shows that nonlinear methods are better for diabetes diagnosis using HR signals. As linear methods are unable to find hidden information in HR signals. Researchers in [40] showed that clinically significant nonlinear parameters were correlation dimension approximate entropy, sample entropy, and recurrence plot properties. The AdaBoost classifier is considered as the best classifier for diabetes diagnosis using HR signals. As researchers in [42] applied different classification algorithms on clinically significant measures, and among all of them, AdaBoost achieved the highest accuracy of 90% for Heart Rate signals and HRV analysis.

3.3. Results comparison in terms of accuracy

Many researchers used Pima Indian Diabetes Dataset to test and train their proposed techniques. Here is the graphical representation of their results in terms of accuracy using the same Pima Indian Diabetes Dataset (PIDD). Fig. 3 shows the best accuracies achieved using different approaches for diabetes prediction, among them Deep Neural Network [24] achieved the highest accuracy. It has four hidden layers, and different combinations of no. of neurons are applied to hidden layers to achieve the best result. Along with that, SVM [38] also achieved the best accuracy as there is a slight difference in their accuracies.

4. Conclusion

Among different challenges in the medical diagnosis system, diabetes detection is one of the major technical challenges. Early diagnosis of diabetes is important as delayed detection may lead to different diseases that include kidney failure, stroke, blindness, heart attacks, and lower limb amputation. Machine learning techniques have been introduced in the medical diagnosis system as they have proven to be accurate in diagnosis, successful in treatments, and more cost-efficient. Deep learning is a subset of machine learning in AI, which has the capability of self-learning from data. It is also capable of unsupervised learning. It can learn large amounts of unstructured and unlabeled data that for the human brain may take years to understand. This research is done on existing techniques to perform a survey of the diagnosis of diabetes. This study includes papers from the last decade.

Diabetes is one of the fatal diseases, and its early and accurate detection is important to save humans from other fatal effects. Many researchers have proposed methodologies that showed better results in terms of accuracy. Among all approaches which use the same dataset (PIDD) for their model training and testing, Deep Neural Network performed better. On the contrary, DNN has shortcomings like it requires more computational time and frequent adjustment of parameters [15]. It is a well-known fact that deep learning performs more accurately on image datasets. Therefore, image datasets should also be considered for a diabetes diagnosis. Hence, a model must be introduced in the future that should be able to overcome these issues.

These different methods (PIDD, ECG, E-Nose and facial images, etc.) for diabetes diagnosis have the clinical advantage as there is no need for a blood sample for diabetes diagnosis as using these methods, diabetes can be diagnosed without any pain.

As deep learning is getting more attention nowadays, deep learning should be combined with different algorithms to achieve better accuracy as deep learning can learn large amounts of unstructured and unlabeled data that even the human brain might take years to understand. Also, as there are only a few diabetes datasets available on the internet, so more public data should be available to do research. More research should be performed using Heart Rate as it takes less bandwidth, and its computational complexity is also less. It can also be used in the cloud or mobile devices. HR signals should also be used to detect other cardiac diseases.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

None Declared.

References

[1] Swain Aparimita, Mohanty Sachi Nandan, Das Ananta Chandra. Comparative risk analysis on prediction of diabetes mellitus using machine learning approach. IEEE 2016:3312–7.
[2] Nirmala Priya Shirley Muller and M. Diagnosis of gestational diabetes mellitus using radial basis function. IEEE 2016;4.
[3] Sneha Joshi Miss, Borse Megha. Detection and prediction of diabetes mellitus using back-propagation neural network. IEEE 2016;4.
[4] Erkaymaz Okan, Ozer Mahmut. Impact of small-world network topology on the conventional artificial neural network for the diagnosis of diabetes. Sci. Direct 2016:8.
[5] Shu Ting, Zhang Bob, Tang YY. Using k-NN with weights to detect diabetes mellitus based on genetic algorithm feature selection. IEEE 2016:6.
[6] Ali Mohebbi, Aradottir Tinna B, Johansen Alexander R, Bengtsson Henrik, Fraccaro Marco, Mørup Morten. A deep learning approach to adherence detection for type 2 diabetics. IEEE 2017:4.
[7] Hariyanto Riyanarto Sarno, Rehman Wijaya Dedy. Detection of diabetes from gas analysis of human breath using E-nose. IEEE 2017:6.
[8] Musale Reena, Paithane AN. Design and develop an algorithm for a diabetic detection using ECG signal. IEEE 2017:6.
[9] Mangrulkar RS. Retinal image classification technique for diabetes identification. In: 2017 Int. Conf. Comput. Methodol. Commun. ICCMC Erode IEEE; 2017. p. 961–6.
[10] Chen Wenqian, Chen Shuyu, Zhang Hancui, Wu Tianshu. A hybrid prediction model for type 2 diabetes using K-means and decision tree. In: 2017 8th IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS Beijing IEEE; 2017. p. 386–90.
[11] Mercaldo Francesco, Nardone Vittoria, Santone Antonella. Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Sci. Direct 2017;112:2519–28.
[12] Das Tarak, Guha Sayanti, Ghosh Arijit, Basak Piyali. Early detection of diabetes based on Skin impedance Spectrogram and heart rate variability noninvasively. In: 2017 1st Int. Conf. Electron. Mater. Eng. Nano-technol. IEMENTech Kolkata IEEE; 2017. p. 1–5.
[13] Zheng Tao, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. Sci. Direct Jan. 2017;97:120–7.
[14] Jahangir Maham, Afzal Hammad, Ahmed Mehreen, Khurshid Khawar, Nawaz Raheel. An expert system for diabetes prediction using auto-tuned multi-layer perceptron. In: IEEE, vol. 2017 intelligent systems Conference (IntelliSys). London: IEEE; 2017. p. 722–8.
[15] Wei Sidong, Zhao Xuejiao, Miao Chunyan. A comprehensive exploration to the machine learning techniques for diabetes identification. In: 2018 IEEE 4th world forum internet things WF-IoT IEEE; 2018. p. 291–5.
[16] El_Jerjawi Nesreen Samer, Abu-Naser Samy S. Diabetes prediction using artificial neural network, vol. 478. *Springer*; Oct. 2018. p. 12.
[17] Wu Han, Yang Shengqi, Huang Zhangqin, He Jian, Wang Xiaoyi. Type 2 diabetes mellitus prediction model based on data mining. Sci. Direct 2018;10:100–7.
[18] Das Sujit Kumar, Mishra Arnab Kumar, Roy Pinki. Automatic diabetes prediction using tree-based ensemble learners. In: Int. Conf. Comput. Intell. IoT ICCIIoT 2018. SSRN ELSEVIER; 2018. p. 6.
[19] Samant Piyush, Agarwal Ravinder. Machine learning techniques for medical diagnosis of diabetes using iris images. Comput. Methods Programs Biomed. 2018 Sci. Direct Apr. 2018;157:121–8.
[20] Sisodia Deepti, Singh Sisodia Dilip. Prediction of diabetes using classification algorithms. Sci. Direct 2018;132:1578–85.
[21] Swapna G, Soman KP, Vinayakumar R. Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. Sci. Direct 2018;132:1253–62.
[22] Alhassan Zakhriya, McGough A Stephen, Alshammari Riyad, Daghstani Tahani, Budgen David, Moubayed Noura Al. Type-2 diabetes mellitus diagnosis from time series clinical data using deep learning models. In: Artif. Neural netw. Mach. Learn. – ICANN 2018, vol. 11141. Springer; 2018. p. 468–78.
[23] Akm Ashiquzzaman, et al. Reduction of overfitting in diabetes prediction using deep learning neural network. In: IT Converge. Secure. 2017 Lect. Notes Electr. Eng, vol. 449. Springer Singap.; 2018. p. 35–43.
[24] Islam Ayon Safial, Milon Islam Md. Diabetes prediction: a deep learning approach. In: Int. J. Inf. Eng. Electron. Bus, vol. 11; 2019. p. 21–7.
[25] Deshmukh Tushar, Fadewar HS. Fuzzy deep learning for diabetes detection. In: Adv. Intell. Syst. Comput. Springer, vol. 810; 2019. p. 875–82.
[26] Deshmukh Tushar, Fadewar HS, Shukla Ankur. The Detection of Prameha (Diabetes) in Ayurvedic Way with the Help of Fuzzy Deep Learning. Springer 2020: 152–8.
[27] Fatima Meherwar, Pasha Maruf. Survey of Machine Learning Algorithms for Disease Diagnostic. J. Intell. Learn. Syst. Appl. 2017:1–16.
[28] Al-Turjman Fadi, Baali Ilyes. Machine learning for wearable IoT-based applications: A survey. Wiley; 2019. p. 16.
[29] Ching T, et al. Opportunities and obstacles for deep learning in biology and medicine. J. Roy. Soc. 2018:47.
[30] Kaur Pavleen, Kumar Ravinder, Kumar Munish. A healthcare monitoring system using random forest and internet. of things (IoT). Springer 2019:19905–16.
[31] https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes.
[32] https://www.cms.waikato.ac.nz/~abifet/book/chapter_6.html.
[33] https://www.edvancer.in/logistic-regression-vs-decision-trees-vs-svm-part2/.
[34] https://www.oreilly.com/library/view/hands-onmachinelearning/9781789346411/e17de38e-421e-4577-afc3-efdd4e02a468.xhtml.
[35] Veena Vijayan V, Anjali C. Prediction and Diagnosis of Diabetes Mellitus - A Machine Learning Approach. IEEE 2015:6.
[36] Tafa Zhilbert, Pervetica Nerxhivane, Karahoda Bertan. An Intelligent System for Diabetes Prediction. IEEE 2015:5.
[37] kumari Mukesh, Vohra Rajan, Arora Anshul. Prediction of Diabetes Using Bayesian Network. Int. J. Comput. Sci. Inf. Technol. 2014;5:5.
[38] Neilesh B, Khyati Gandhi K. Diabetes prediction using feature selection and classification. Int. J. Adv. Eng. Res. Dev. 2014;1(5):7.
[39] Acharya UR, Fausta O, Sreeb SV, Ghista DN, Dua Sumeet, Joseph P, et al. An integrated diabetic index using heart rate variability signal features for diagnosis of diabetes. Taylor & Francis April 2015:14.
[40] Oliver Faust, Acharya, Filippo Molinari, Chattopadhyay S, Tamura T. Linear and non-linear analysis of cardiac health in diabetic subjects. Sci. Direct 2012:295–302.
[41] Pachori Ram Bilas, Avinash Pakala, Shashank Kora, Sharma Rajeev, Rajendra Acharya U. Application of empirical mode decomposition for analysis of normal and 4 diabetic RR-interval signals. Sci. Direct 2015:15.
[42] Rajendra Acharya U, Faust Oliver, Kadri Nahrizul Adib, Suri Jasjit S, Yu Wenwei. Automated identification of normal and diabetes heart rate signals using nonlinear measures. Sci. Direct 2013:7.