AN EXAMINATION OF THE USE OF DEEP LEARNING TECHNIQUES TO DIAGNOSE DIABETES MELLITUS

T.P.Udhayasankar[1], G.Suganya[2], Dr.O.Sarananan[3]

[1]Associate Professor, Department of Computer Science and Engineering,AEC,Salem

[2]Assistant Professor, Department of Computer Science and Engineering,AEC,Salem

[3]Professor, Department of Computer Science and Engineering,AEC,Salem

Abstract:

Uncontrolled diabetes is the cause of Diabetes Mellitus (DM), a condition that results in organ dysfunction in patients. Thanks to current developments in computer vision and artificial intelligence, early diagnosis and treatment of DM are more advantageous than manual evaluation through an automated method. This evaluation thoroughly examines and presents six aspects of DM acknowledgement, prognosis, and self-management methods, including DM data sets, pre-processing technologies, feature extraction, recognition through deep learning, classification and DM prognosis, and smart artificial intelligence-based DM assistant. The findings and recommendations of the previous research are interpreted. A comprehensive overview of DM diagnosis and self-administration technologies is also provided in this paper, which can be helpful to field researchers.

Keywords: Mellitus, Diabetes, detection, deep learning, algorithms, classification

I. INTRODUCTION

Significant advancements in medicine and medical sciences, particularly high-throughput sequencing, continue to aid in the generation of massive amounts of data at low prices, propelling analytical biology into the realm of big data [1], [2]. While these procedures generate a lot of data, they don't allow for any sort of analysis, characterization, or retrieval of information. The major goal is to delve deeper into the ever-increasing number of biological data in order to lay the groundwork for answers to basic medical and biological concerns. The power and effectiveness of comparable techniques determine their ability to isolate patterns and develop models from information. As a result, data availability has considerably aided data-driven study in biological science. Prognosis and diagnosis of diseases that threaten people or shorten their lifespan are two of the most important research fields in a hybrid DM is an example of such a condition. It has been noted as a growing health issue in both industrialised and developing countries in the twenty-first century. Diabetes was said to be more common as a result of western lifestyle, industrialization, and social progress [3]. It is a global epidemic with devastating personal, societal, and economic consequences that affects roughly 260 million individuals globally.

Type 2 dm is a severe form of diabetes defined by chronic hyperglycemia, which occurs when the pancreas does not produce enough insulin or when the glucose it does produce is not used adequately by the body. It can also be asymptomatic [4]. The time between starting treatment and receiving a diagnosis can be more than ten years, however prediction is improving[5]. To diagnose diabetes, a clinician must look at a number

However, variables like the experts' lack of experience or exhaustion may contribute to a misdiagnosis. Early therapy with exercise and diet or treatment strategies has been shown to significantly reduce or prevent Type 2 diabetic complications in humans.

For the prevention of chronic diseases, a detailed recommendation addressing dietary adjustments was published [7]. For the first detection of diabetes, various risk assessments have been developed. Schwarz et al. conducted a detailed review of these methods, including their accuracy and specificity, and concluded that the Finnish Insulin Risk Score was the most valuable tool for diabetes first diagnosis [8]. However, because this system relies on human intervention in determining criteria and scoring, it is vulnerable to human mistake [9]. Because DM is influenced by a variety of other variables and has severe socioeconomic consequences, it generates vast amounts of data. Figure 1 depicts the phases involved in forecasting that necessitate algorithm training. These are also topics of great interest in the clinical scientific community today, as these techniques are primarily aimed at improving the sensitivity and accuracy of disease identification and treatment. At the same time, these strategies reduce the possibility of human error throughout the decision-making process [9]. As a result, in the context of this research, an attempt was made to examine the most recent literature on methods to computer vision and data mining in insulin research.

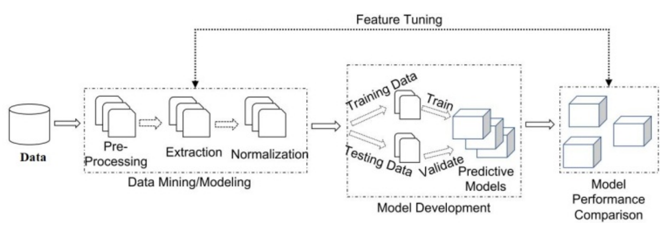

Figure 1. The use of deep learning techniques to extract features and forecast dm [10].

II. ML and KDD

Deep learning was the science which works with the ways in which computers learn and develop in its most basic form. For many academics, the terms "deep learning" and "artificial intelligence" are interchangeable, provided that the capacity to learn is a significant feature of someone who is intelligent in the broadest sense of the term. The goal of deep learning is to create computer devices that can adapt and gain from their experiences [11]. KDD is a field that comprises hypotheses, techniques, and procedures for making sense of information and extracting useful information from it [12]. Figure 2 depicts the stages involved in KDD for demonstrative purposes only.

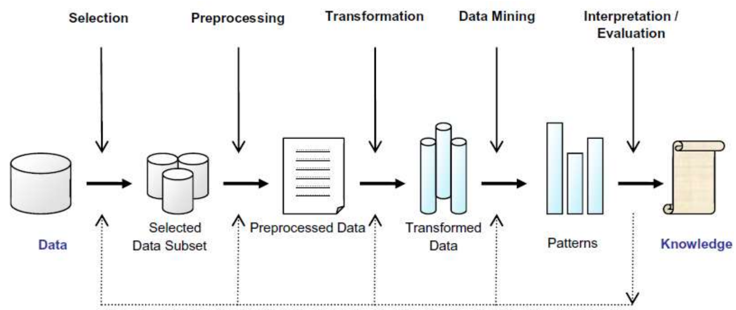

Figure 2. The KDD technique's main steps [12].

A. Categories of ML Task

Deep learning activities are typically divided into 3 categories [13], including (a) reinforcement methods, in which a training examples data feature is construed from the programme, (b) unsupervised learning, wherein the compute cluster attempts to perceive an unlabelled file formats, and (c) reinforcement learning, in which the device interacts with a changing situation.

In reinforcement methods, the computer must inferential "learn" a function called the target functional, which is a formula of a model that represents the data. The optimization problem is used to analyse the value of a variable from a set of variables known as attributes. The domain of an example is the list of input variables method datasets. Each case is defined by a collection of features. Training set are a sample of all situations for which a number for the single output is known. In order to find the optimum linear model from a provided training array, the training model considers alternative values, referred to as possibilities. In supervised learning, there are 2 kinds of learning assignments: categorization and extrapolation.

In unsupervised learning, the machine tries to uncover the hidden distribution of data or relationships between variables. In that situation, the training data consists of cases with no associated tags. The term Reinforcement Learning refers to a set of tactics in which the system attempts to learn to optimise some concept of accumulated compensation through immediate communication with the environment[14].

B. DM

Diabetes Mellitus (DM) would be a group of metabolic diseases characterised by inconsistent insulin secretion [15]. Excess blood glucose levels and improper carbohydrate, fat, and protein metabolism result from insulin insufficiency. DM is one of the most serious endocrine illnesses, affecting over 250 million individuals each year. The incidence of diabetes is expected to rise dramatically in the future years. DM can be classified into several categories. T1D and T2D are the two most common clinical forms. T2D, which is characterised by insulin sensitivity, is the most serious type of diabetes (affecting 85% of all diabetic patients). Regular exercise, way of life, eating behaviours, and inheritance are the main factors of T2D.

C. Detection and Prediction of Diabetic Metabolic Syndrome Biomarkers

Biomarkers, also known as biological markers, are observable signs of an illness that reflect diagnosis and treatment status. Biomarkers are substances found in body fluids that are used to monitor the severity of clinical disorders and how they respond to treatments. Biomarkers might be direct outcomes or indirect indications of the illness's concomitant consequences. Biomarkers can be used to reflect the presence and severity of hyperglycemia or the presence and severity of related diabetes and its complications in the case of DM [16].

The predictive performance of the features extracted is then tested using a classification method. The second group is concerned with prediction and diagnosis. The system is implemented in MATLAB using SQL server as the information, as illustrated in Fig. 3.

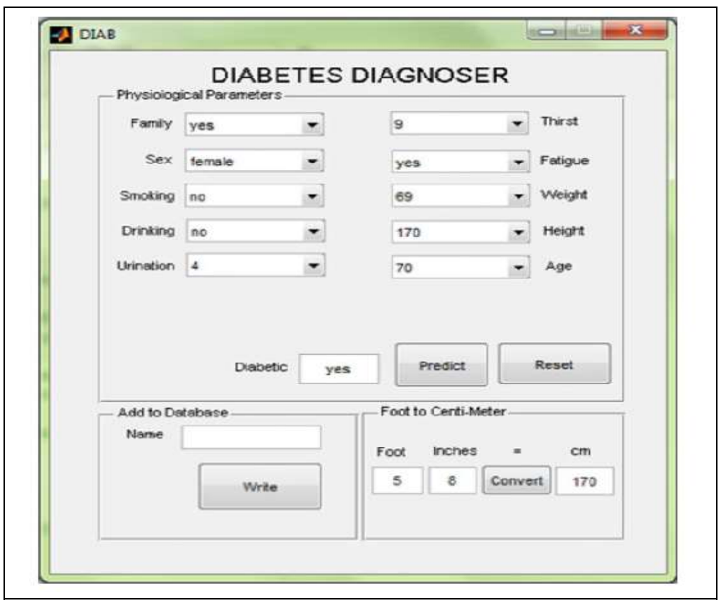

Fig. 3.MATLAB Program for Diabetic Medication Diagnosis[31]

Zhang et al [32] proposed a non-invasive method for detecting DM and NPDR in the early stages of DR based on three kinds of features extracted from tongue pictures. These include colour, design, and form. Figure 4 shows the tongue capture equipment that was constructed in-house. For a DM sample, they discovered a greater proportion of Deep Red colour. While normal samples have a better textural quality (Figure 5-8). Figure 9 shows three typical Healthy and DM samples. Finally, by combining features from each of the three classifications, they were capable of distinguishing healthy or DM tongue from NPDR tongue with an accuracy rate of 79 percent.

.

Figure 4.Tongue-capture apparatus.

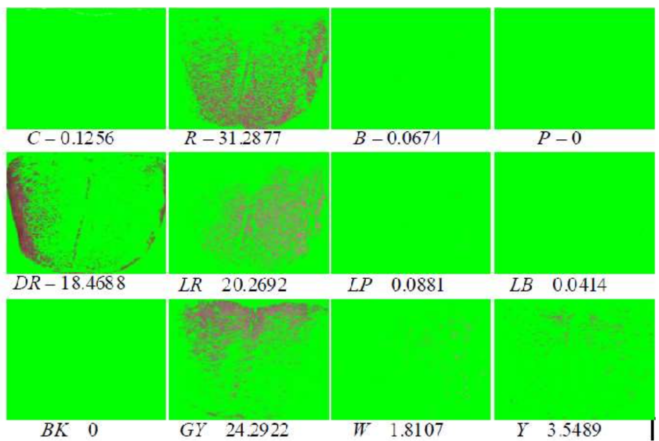

Fig. 5. The DM tongue specimen, its tongue colour feature space, and the 12 colour makeup that corresponds to it, with the majority of the pixels categorized

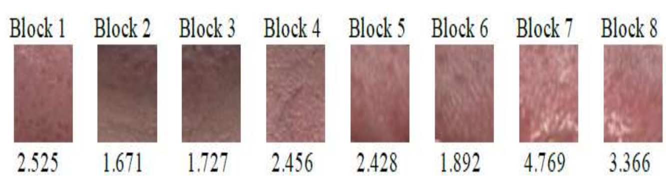

Fig. 6.Texture blocks with a good texture value

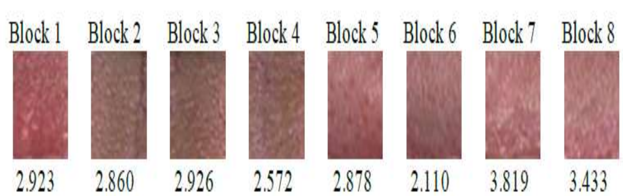

Fig. 7. Texture values for DM texture blocks

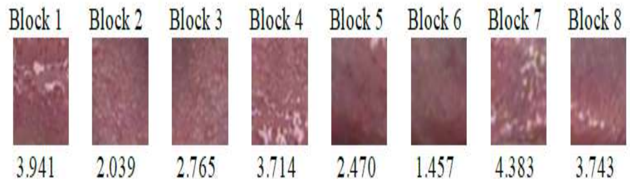

Fig. 8. The texture value of an NPDR texture block

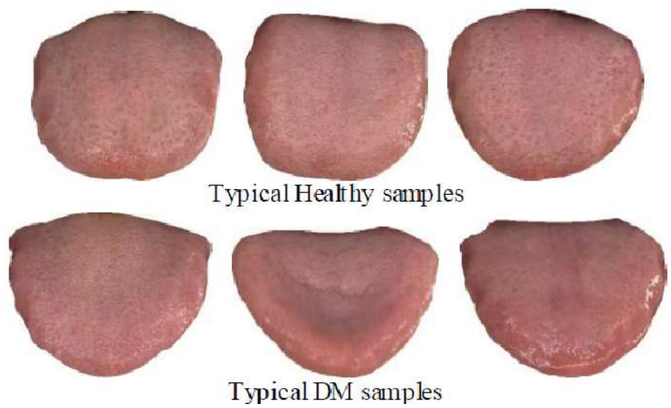

FIG. 9. Tongue samples from healthy and diabetic people.

For diabetes diagnosis, a decision - making support scheme based on the AdaBoost algorithm and Decision Stump was applied. SVM and decision tree were also included in this algorithm to boost accuracy. Figure 10 depicts the suggested system. Figure 11 illustrates the operation of a decision tree for prediction and classification. They were able to achieve an efficiency of 75 percent.

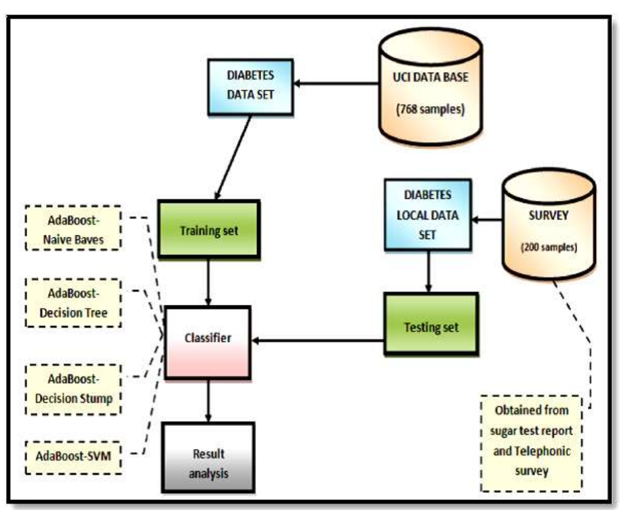

Figure. 10. The suggested system is depicted as a block diagram.

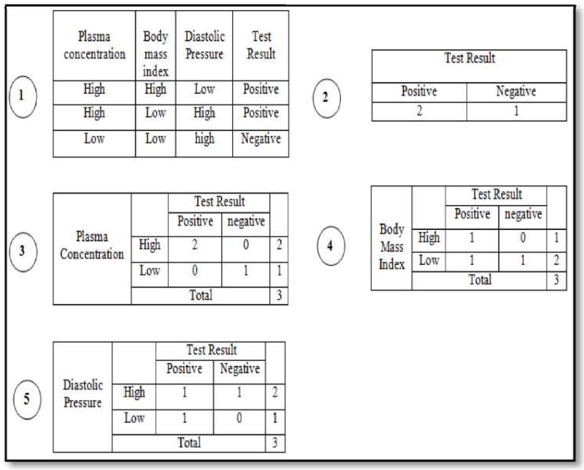

Figure 11. Decision tree in action for diabetic prognosis

In terms of multi-dimensional sample data, researchers used a dataset of 41 lakh persons from pharmacy records from 2005 to 2009 to make predictions for different T2D prediction. Ensemble approaches, which combine various ml algorithms, have proven to be an effective way to enhance classification accuracy. Unique strategies are also used in DM forecasting. developed a multi-layer categorization ensemble design that integrated seven different classifiers, and proposed Rotation Forest, an unique ensemble technique for combining thirty different deep learning algorithms. Finally, an ensemble learning technique was suggested that converts the SVM decision "black box" into intelligible and explicit laws.

Table 1:A summary of the various algorithms employed and the performance metrics evaluated.

Type of diabetes	Algorithms used	Performance metrices	Regression/classification	Reference
T1D	Random Forest and RReliefF	Prediction horizon (min)/RMSE (mg/dl), standard deviation of the importance of features based on RF algorithm, RMSE rate of SVR regression	Classification and Regression	[28]

		models		
T2D	Electromagnetism-like mechanism (EM) algorithm	Non-parametric statistical tests are conducted to justify the performance of the methods in terms of classification accuracy and Kappa index	Classification	[21]
Pre-diabetic females	Wrapper method, symmetrical uncertainty (filter methods).	Akaike information criterion (AIC) and area under the curve (AUC)	Classification	[19]
Onset of DM	ANFIS	Accuracy (%) Specificity (%) Sensitivity	Classification	[38]
Onset of DM	k-NN	Accuracy (%) Specificity (%) Sensitivity	Classification	[39]
T1D	Novel, clustering-based feature extraction framework	Prediction horizon (min)/RMSE (mg/dl), -30/5.7 ±1.5	Classification	[20]
T1D	Feed-forward neural network and first-order polynomial model	Prediction horizon (min)/RMSE (mg/dl), 30/14.0 ± 4.1	Classification	[40]
T1D	Jump neural network model	Prediction horizon (min)/RMSE (mg/dl),	Classification	[41]

		30/16.2 ± 3.1		
DM diagonsis	LDA–MWSVM	sensitivity specificity, and accuracy,	Regression	[26]
T1D	SVR	accuracy,average prediction errors	Regression	[28]

D. DT, RF, SVM, NB

attempted to develop a system for predicting a patient's level of diabetic risk. The models are developed using decision trees, ANNs, Naive Bayes, and SVM categorization methods. Precision was achieved with 85 percent for decision trees, 69 percent for Naive Bayes, and 67 percent for SVM. These studies revealed a high level of precision. This study analyses crucial information, creates a deep learning-based predictive model, and identifies the best classifier to provide the closest outcome to medical data. Using crucial features, constructed a prediction algorithm using deep learning to determine the best classifier to generate the nearest result instead of medical outcomes

For diabetes detection, DA, DT, LR, and SVM, k-NN series of data mining algorithms, as well as ensemble learners, were applied. The data were assessed using validation set criteria, with average classification accuracy serving as the criterion for success. The average accuracy levels reached ranged from 64.48 percent to 78.05 percent. The LR approach achieves 78.95 percent, whereas a Coarse Gaussian SVM method achieves the poorest 66.15 percent.

Using multiple timeframes and data types sets from DM, various Ml models were evaluated for classification result. The algorithms were then combined to create a weighted ensemble method that could enhance detection accuracy by combining the efficacy of the individual models. Tree-based algorithms were used to identify critical parameters within patient data, allowing data-learned models to identify risk patients for every illness class. Using the information supplied, the suggested cardiovascular disease ensemble model produced an AU-ROC score of 83 percent without lab results and 86 percent with lab results.

Using multiple timeframes and data types sets from DM, various Ml models were evaluated for classification result. The algorithms were then combined to create a weighted ensemble method that could enhance detection accuracy by combining the efficacy of the individual models. Tree-based algorithms were used to identify critical variables within patient data, allowing data-learned models to identify risk patients for every illness class. Using the information supplied, the suggested cardiovascular disease ensemble model produced an AU-ROC score of 83 percent without lab results and 86 percent with lab results.

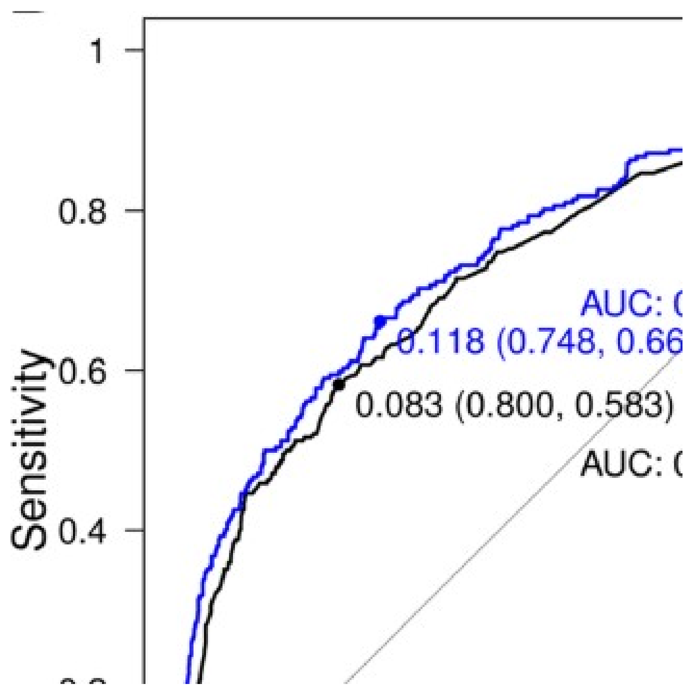

Figure 12. To forecast GDM, a multivariate LoR-based ROC was used.

768 PIMA Indian Healthcare Data installations were used to examine the accuracy of predicted data mining algorithms by Varma and Panda [45]. They attempted to predict early diabetes using NB, LoR, C5.0 DT, and SVM. The models were evaluated in terms of accuracy, reliability, sensitivities, applicability, and F1 Score measures.

Table 2: Algorithms employed and results obtained

Algorithm	**Accuracy**
Decision Tree	86 %
Gaussion NB	93 %
LDA	94 %
SVC	60 %
Random Forest	91 %
Extra Trees	91 %
AdaBoost	93 %
Perceptron	76 %
Logistic Regression	96 %
Gradient Boost Classifier	93 %
Bagging	90 %
KNN	90 %

For the early GDM forecasting by LoR, deep NN, and, 17 variables were used. To facilitate clinical adoption, seven factors from the 17-variable array were chosen. The 7-variable data and the 73-variable data were had used to generate simulation results of Initial GDM for different conditions using advanced deep learning methodologies. With 73 parameters, the Accuracies were 1.23, indicating that the deep artificial neural network had a high level of discriminating power. With the 7-variable LR model, great discriminatory ability was also attained.

Intensive Care III (MIMIC-III) data were used in a secondary investigation. It is necessary to employ a medical expertise mart. For several Ml techniques, deep learning and NLP methodologies were applied. In the healthcare industry, domain knowledge is centred on dictionaries established by clinical terminology experts who have characterised drugs or clinical manifestations. Figure 13 shows the optimum configuration of the employed ML models with a competitive AUC of 0.87. ML models of medical documentation, when combined with NLP, have the potential to aid health care practitioners in forecasting the risk of mortality of critically ill individuals.

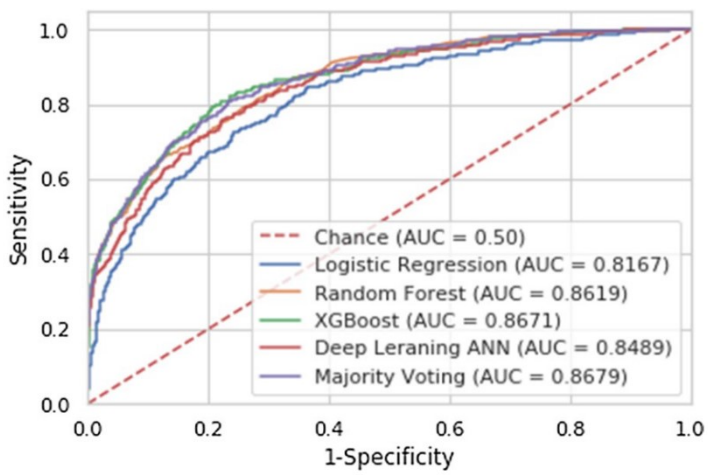

Figure 13: AUC of several deep learning models

This paper studied diabetic patients as well as how to diagnose diabetes using a range of deep learning approaches to create a system with some PIMA dependant dependencies. A portion of PIMA was evaluated, as well as a dataset from Kurmitola Medical Center in Dhaka, Bangladesh. The trained data was also used to test the model. DT, KNN, RF, and NB are the algorithms used. The purpose of the study is to demonstrate the output of several classifiers that are taught in a diabetes set of data in one nation and evaluated on patients in another country. Figure 14-15 shows the correlation vector and confusion derived from several deep learning techniques.

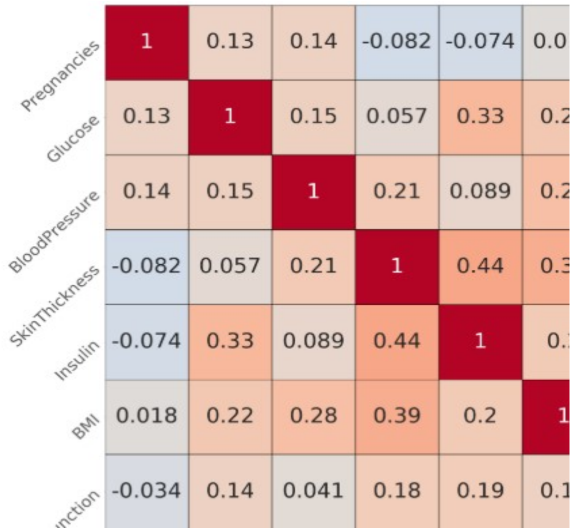

Figure 14: Dataset correlation matrix

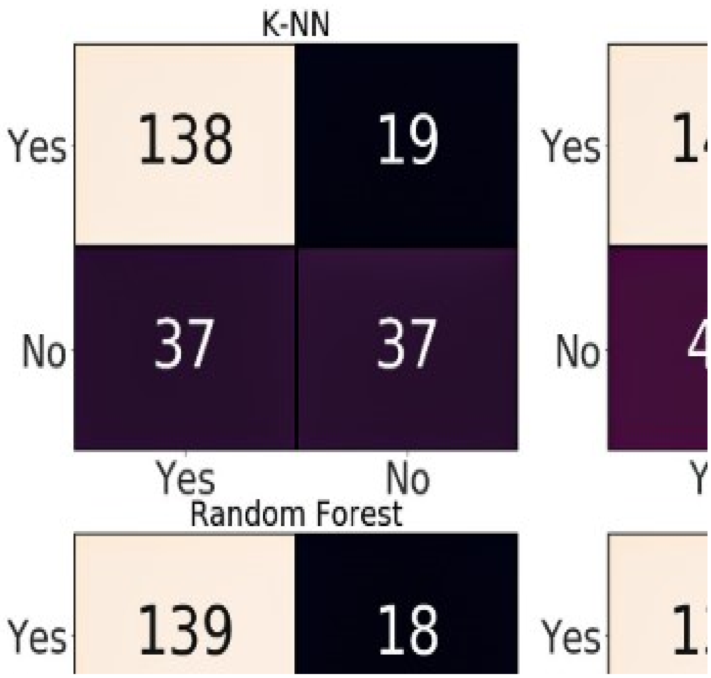

Figure 15: The KNN, DT, RF, and NB algorithms all have a confusion matrix.

E. Boosting algorithms

This was built to match the data of risk in GDM from Tianchi Quality Medical Contest and Artification intellect using the prediction algorithms LightGBM, XGboost, and Random Forrest for comparative study. According to the findings, LightGBM accounts for 84.87 percent of the AUC. In comparison to other greater than the linear regression.

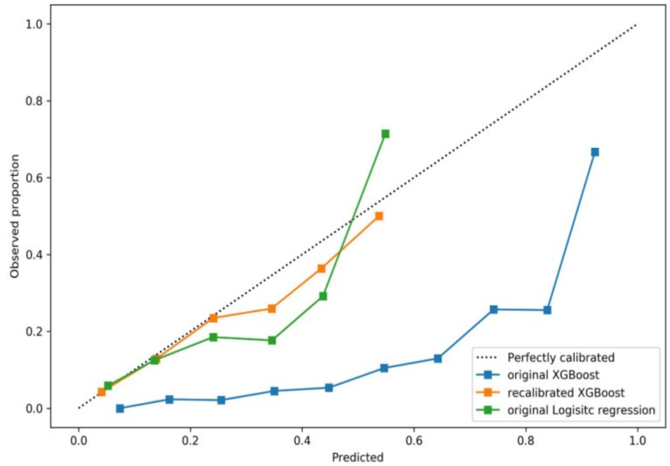

Figure 17: Using a modified XGBoost method, the AUC was improved.

III. CONCLUSIONS

This article provided a comprehensive summary of automatic diabetic detection as well as diagnostic procedures. This study examines each research project from four angles: databases, deep learning-based categorisation, AI-based smart companions for diabetic patients, and effectiveness measures. The DNN and SVM were found to have better classification outcomes in most studies, followed by RF and Ensemble Classifier. The CNN was discovered to learn to automatically retrieve and categorize DM data in a significant way. Many academics have developed smart aides like chatbots and bots to assist patients with their daily DM management tasks, including as food control, insulin administration, and so on. As measures, the majority of the scientific community employed reliability, accuracy, specificity, and AUC.

REFERENCES

1. Ibrahim, Ibrahim, and Adnan Abdulazeez. "The role of deep learning algorithms for diagnosing diseases." *Journal of Applied Science and Technology Trends* 2.01 (2021): 10-19.

2. Subramani, Prabu, and Parameshachari BD. "Prediction of muscular paralysis disease based on hybrid feature extraction with deep learning technique for COVID-19 and post-COVID-19 patients." *Personal and ubiquitous computing* (2021): 1-14.
3. Chen, Joy IongZong, and P. Hengjinda. "Early prediction of coronary artery disease (CAD) by deep learning method-a comparative study." *Journal of Artificial Intelligence* 3.01 (2021): 17-33.
4. Ghosh, Pronab, et al. "Use of efficient deep learning techniques in the identification of patients with heart diseases." *2021 the 5th International Conference on Information System and Data Mining*. 2021.
5. Kwekha-Rashid, Ameer Sardar, Heamn N. Abduljabbar, and Bilal Alhayani. "Coronavirus disease (COVID-19) cases analysis using machine-learning applications." *Applied Nanoscience* (2021): 1-13.
6. Ahmed, ShaymaaTaha, and SuhadMalallahKadhem. "Using Deep learning via Deep Learning Algorithms to Diagnose the Lung Disease Based on Chest Imaging: A Survey." *International Journal of Interactive Mobile Technologies* 15.16 (2021).
7. Bari Antor, Morshedul, et al. "A comparative analysis of deep learning algorithms to predict alzheimer’s disease." *Journal of Healthcare Engineering* 2021 (2021).
8. Jindal, Harshit, et al. "Heart disease prediction using deep learning algorithms." *IOP conference series: materials science and engineering*. Vol. 1022. No. 1. IOP Publishing, 2021.
9. Mei, Jie, Christian Desrosiers, and Johannes Frasnelli. "Deep learning for the diagnosis of Parkinson's disease: a review of literature." *Frontiers in aging neuroscience* 13 (2021): 633752.
10. Akter, Shamima, Hossain Uddin Shekhar, and Sharif Akhteruzzaman. "Application of biochemical tests and deep learning techniques to diagnose and evaluate liver disease." *Advances in Bioscience and Biotechnology* 12.6 (2021): 154-172.
11. Nadeem, Muhammad Waqas, et al. "Fusion-based deep learning architecture for heart disease prediction." (2021).
12. Kishor, Amit, and Wilson Jeberson. "Diagnosis of heart disease using internet of things and deep learning algorithms." *Proceedings of second international conference on computing, communications, and cyber-security*. Springer, Singapore, 2021.
13. Lamba, Rohit, et al. "A hybrid system for Parkinson’s disease diagnosis using deep learning techniques." *International Journal of Speech Technology* (2021): 1-11.
14. Ahsan, MdManjurul, et al. "Effect of data scaling methods on deep learning algorithms and model performance." *Technologies* 9.3 (2021): 52.
15. Arumugam, K., et al. "Multiple disease prediction using Deep learning algorithms." *Materials Today: Proceedings* (2021).