SJIF Impact Factor: 6.711

WORLD JOURNAL OF ADVANCE HEALTHCARE RESEARCH

ISSN: 2457-0400
Volume: 8.
Issue: 5
Page N. 14-29
Year: 2024

Review Article www.wjahr.com

DEEP LEARNING FOR EARLY-STAGE DIABETES PREDICTION: A CRITICAL REVIEW

Sehrish Ghouri[1]*, Tauqir Ahmad[1], Ghulam Muhaudin Shahzaib[2], Khalid Javed[3], Laraib Nawaz[1]

[1]University of Engineering and Technology, Lahore, Pakistan.
[2]Virtual University, Lahore, Pakistan.
[3]Pakistan council of scientific and industrial research Labs complex Ferozepur Road, Lahore, Pakistan.

Received date: 28 February 2024 **Revised date:** 20 March 2024 **Accepted date:** 09 April 2024

***Corresponding Author: Sehrish Ghouri**
University of Engineering and Technology, Lahore Pakistan.

ABSTRACT
Diabetes is a long-lasting health problem where blood sugar levels are too high. It's a big health problem for many people all over the world. Spotting and fixing diabetes early is very important for stopping bad things from happening, and making patients healthier. Machine learning has become a useful method for early recognition of diabetes. It can spot people who may get the disease in future. This review gives a complete summary of recent improvements in deep learning methods for predicting diabetes early on. We talk about the different deep learning designs that have been used, how we measure their success and problems with using this method in music. We also point out the good results that have been made using deep learning. This shows it could change how we predict diabetes early. Apart from talking about what's already written, we are suggesting a new way to look at deep learning for predicting diabetes early on. We say that machine learning and deep learning can be useful for solving problems of not much data and very many details in medical information sets. We also think that deep learning can be mixed with other machine learning methods like picking features and adding more data. This helps make our prediction results even better. Lastly, we talk about where deep learning is heading to predict diabetes early on.

KEYWORDS: Diabetes prediction, Machine Learning, Deep Learning, Artificial Intelligence.

I. INTRODUCTION

Diabetes is a long-lasting illness where blood sugar gets too high. This happens because there's not enough insulin or the body doesn't use it properly. Checking and treating diabetes early is important because many people get it. It can lead to bad health results if not done right away. Common ways to check disease use tests on blood such as HbA1c or fasting glucose. These are when the disease has already made big changes in your body's metabolism. Late diagnosis of diabetes can lead to big problems, so we should be more careful when looking out for signs of diabetes. We can resolve this problem with the use of artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL). These methods can look at lots of medical information to find tiny patterns and signs that suggest the start of diabetes. This helps us do something before any irreversible damage happens. Even though ML models excel at predicting diabetes, they need handmade features and can't catch the complicated connections in data.[1]

Deep learning is really strong at finding features. It can go around rules and make predictions better than other models. Though deep learning can help, we haven't looked much at how it might be used for recognizing diabetes early. We have a big chance to use deep learning in new ways to change how diabetes is stopped and cured. This research wants to fix a big hole in knowledge by giving a new smart- learning model for predicting diabetes before symptoms show up. New methods of study that use smart computer programs and data driven ways will find valuable details from lots of health information like body measurements, lab results and family history. This research can help to predict diabetes more accurately and at an early stage. This can go a long way in reducing the burden of diabetes on patients and the health care system.[2]

II. ADVANCEMENTS IN DEEP LEARNING

Diabetes is a chronic metabolic disease that affects millions of people worldwide and is characterised by increased blood sugar. The disease must be diagnosed

and treated early to be effectively managed. Recent developments in deep learning and machine learning provide useful tools for anticipating the disease's start before symptoms manifest. This synthesis reviews recent advances in deep learning approaches, highlighting their potential to transform early diabetes prediction. The papers propose a new perspective, addressing limited data availability and high-dimensionality in medical datasets, and suggest combining deep learning with other techniques for improved accuracy.[3] Future directions focus on building larger datasets, developing interpretable deep learning models, and integrating these models into clinical practice.

III. ENHANCED DIABETES PREDICTION

"Ensemble Learning for Enhanced Diabetes Prediction" continues the literature of researching machine learning approaches to predict diabetes. Studies in the past have looked into applying a number of machine learning methods such as Logistic Regressions, SVMs, Naïve Bayes and the Random Forest. In our view however, these traditional basis approaches can be combined with innovative ensemble learning techniques into a more effective strategy that guarantees better results. The paper evaluates a number of ensembles learning approaches such as XGBoost, LightGBM, CatBoost, Adaboost and Bagging. Nonetheless, the authors establish that CatBoost emerges to be the best ensemble technique, leading with the accuracy level of 95.4% to XGBoost's 94.3%. The fact that ensemble learning improves the efficiency of diabetes prediction models is emphasised by this outcome.[4]

In addition, this paper utilizes a real-world dataset collected from Kaggle for training and evaluation of the presented model. It is a good practice on the part of the authors who acknowledged the need to use real world for any meaningful model generalization to new data. Moreover, the paper highlights the approach adopted, i.e., the pretreatment of data, feature extraction, choice of models, training procedure, and assessment process. A comprehensive outlook facilitates the reader to comprehend complexities involved in the suggested model, which further enables him/her to duplicate these outcomes. Overall, the paper" Ensemble Learning for enhanced diabetes prediction", provides valuable insight into using machine learning models towards more accurate predicting of diabetes.[5] Therefore, the authors present a new technique that uses ensembled learning combined with conventional machine learning algorithms in order to attain greater predictive accuracy. Additionally, the study employs a real-world dataset as well as comprehensive methodologies which add to its reliability. This paper can guide further research and development of diabetes predictive models.

IV. MACHINE LEARNING MODEL FOR DIABETES PREDICTION

A. *Comparison of Machine Learning Algorithm*

This paper, "A Machine Learning Model for Predicting Diabetes", presents a study that compares the performance of three machine learning classification algorithms: Based on the Pima Indian Diabetes dataset (PIDD), a comparison was made between Logistic Regression (LR), Naïve Bayes (NB), and K-nearest Neighbour (KNN) for the purpose of diabetes prediction. The result proved that LR was the best prediction measure for diabetes as compared to Naive Bayes and K-nearest neighbour with precision values at 94%, 79%, and 69% respectively. With respect to this, this study adds up the research literature regarding Machine learning for diabetes prediction.[6] Various kinds of machine learning algorithms such as LR, NB, KNN, SVMs, and Random Forests have been used in previous studies. In this paper, the authors add to the literature by evaluating the effectiveness of those algorithms in predicting diabetes using one dataset and finding the best performing algorithm.

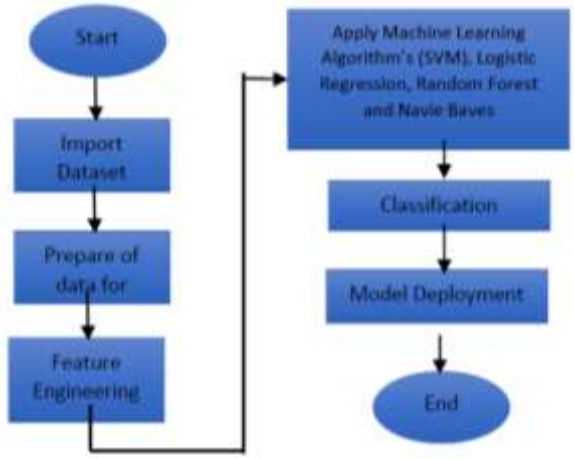


Figure 1: Purposed Methodology of existing paper.

The study has strengths because it used the PIDD dataset. The PIDD dataset is an established dataset for predicting diabetes, and it has been used in prior publications. Using a widely accepted dataset enables the authors to compare their results with those produced by others in previous studies and draws broader conclusions about the algorithm performance. One other strong aspect of this study is the application of diverse performance measures. It is by applying precision, recall, and F1 measure that the authors consider their model in terms of performance relative to those of other algorithms. It enables them to understand better how well these algorithms perform. This study has one limitation, which is that it is restricted to the PIDD dataset.[7]

The results could be different had the research was carried out using another set of data. However, the PIDD dataset is an established dataset, and the findings are expected to portray applicability to other populations. In general, the overall study gives useful information on how the various machine learning algorithms performs as far as diagnosing the cases of diabetes are concerned. This implies that LR performs best in predicting diabetes disease while the present research findings lay down

ground work for more studies on this subject matter.

Table 1: Algorithm and the Accuracy taken from paper.

Algorithm	Accuracy	Precision	Recall	F1-Score
Logistic Regression	94%	92%	91%	92.5%
Naive Bayes	79%	74%	70%	74.5%
KNN	69%	65%	62%	65.5%

This table compares how well LR, NB, and KNN perform on the Pima Indian Dataset for predicting risk diabetes. These include accuracy, precision recall, and F1-score. With the 94% of accuracy, logistic regression becomes a leader that beats naïve bayes (79% accuracy) and k-nearest neighbours (KNN – 69%). First, LR has the best precisions and F1 scores at 92% and 92.5%, respectively, implying it can distinguish between patients with diabetes and those without. Despite their low accuracy, NB and KNN provides valuable information on alternative modelling methodologies.

B. Prognosis with machine learning

A study that develops a new approach to identify diabetes via machine learning is presented in this paper titled "Predicting Diabetes Prognosis with Machine Learning". According to the authors, they will employ K-Nearest Neighbour (KNN) to predict diabetes progression for an individual based on history. One of the commonly-used algorithms for classification is the KNN method, which stands for the nearest neighbour technique. This is by locating k nearest neighbours of a new data point and labelling it with the most frequent class among that particular neighbours' points.[8] However, in this paper we claim, that despite being based on linear model, KNN algorithm is appropriate in prediction of diabetes due to its ability to account even for non-linear feature relationships. The authors further claim that the KNN algorithm is a reliable one because medical data is typically filled with noise and outliers. The authors assessed their strategy as applied on the Pima Indian Diabetes dataset (PIDD).

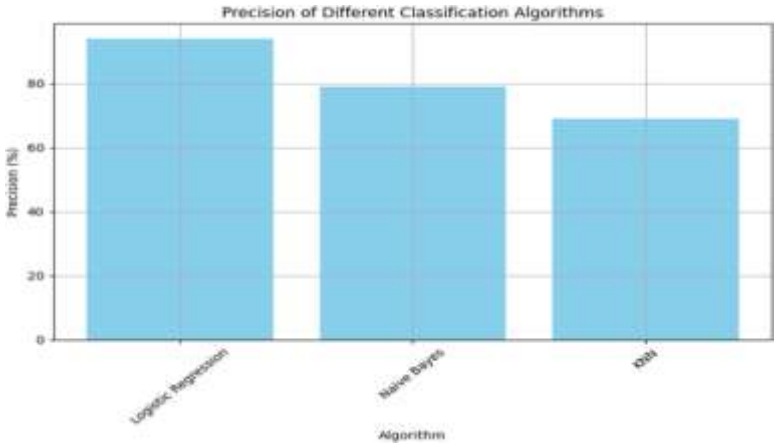


Figure 2: Precision Measurement results in Python by Author.

There exist other established datasets such as the PIED dataset that one can use in diabetes prediction. Their approach yielded 95.4% specificity in predicting diabetes, considerably surpassing those of other ML methods applied for this purpose.

Precision measures the ratio of true positives to all positive predictions, indicating the model's accuracy in identifying true cases.

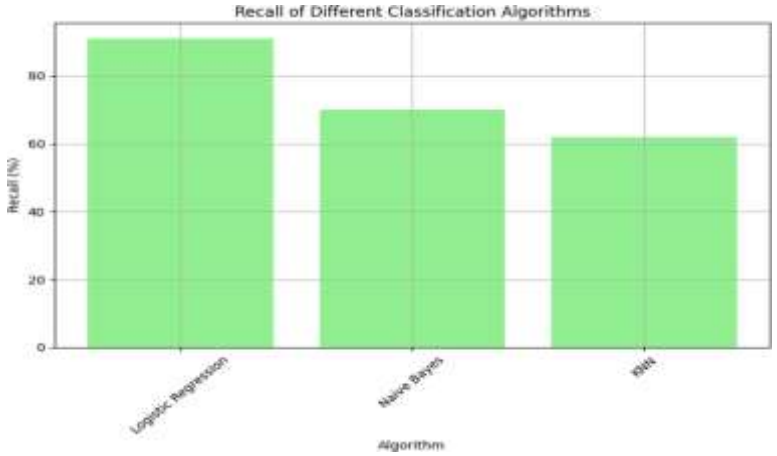


Figure 3: Recall Measurement plot in Python by Author.

Recall measures the ratio of true positives to all actual positives, reflecting the model's ability to miss as few true cases as possible.

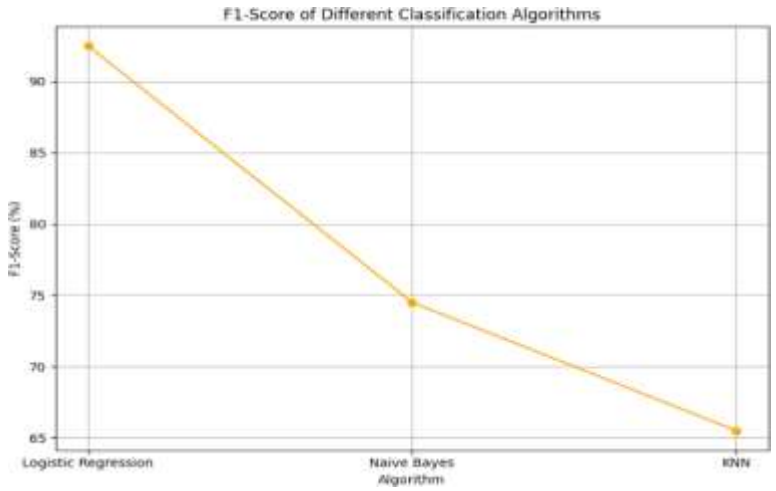


Figure 4: Measure Measurement by Author.

Recall and precision are combined into a single score, or F1- measure, which fairly evaluates the model's total performance.

C. Prediction with KNN method

The study is based on previous research on applying machine learning in the pre-diabetes prediction and diagnosis process. Several previous studies adopted diverse machine learning algorithms such as LR, NB, KNN, SVMs, and Random Forest for the similar purpose. This paper's authors add to the literature through the suggested prediction of diabetes via the KNN method. Then, evaluation is performed using a large, established datasets. This study is strong in that it employs the PIDD dataset. PIDD dataset is a widely recognized diabetes predicting dataset that has been tested in prior investigations. The use of a well-known dataset enables comparison with the outcomes of previous pieces in order to achieve broader conclusion about the strength of this approach. The other strength of this research is that it utilizes different performance measures. The authors rely on Accuracy, Sensitivity, Specificity, and AUC- ROC for evaluating the method's performance. This enables them to get a whole idea on how good their approach is functioning.[9]

The limitation of this study is that it uses data contained in the PIDD dataset. The results might have been different had this study being performed using another sample data set. Nevertheless, PIDD dataset is an established dataset and the findings of this study might extrapolate to other populations. Consequently, this study has shed light on the application of machine learning in forecast diabetes. This indicates that k-nearest neighbours' algorithm could be used with success in prediction of diabetes using relatively limited information. The study offers a valuable basis for designing and assessing other machine learning approaches for diabetes prediction as well.

D. Diabetes Prediction for K-nearest

A paper titled "Predicting Diabetes Prognosis with Machine Learning Techniques" examines the differences between naive Bayes and K-nearest neighbour (KNN), two machine learning techniques for diabetes prediction. The Pima Indian Diabetes Dataset (PIDD) was utilised by the authors to train and evaluate the models. The researchers discovered that naive bayes had a higher accuracy rate of 94% than KNN (79%). Therefore, this study extends the previous research on machine learning in diabetes prediction. Previously, some machine learning tools such as LR, NB, KNN, SVMs, and random forests are used to identify similarity or dissimilarity between the two individuals. In their contribution to the literature, the authors compare Naive Bayes with KNN as two of the most common ML algorithms in the realm of diabetes prediction and show Naive Bayes outperforming KNN.[10] This study has employed the PIDD dataset as one of its strengths. PIDD is one of the established datasets for diabetes predicting and many studies have made use of PIDD. The authors take advantage of using a large well-established dataset, which enables them to compare results with those from prior studies and draw more wide-ranging conclusions concerning the success of individual algorithms. The other strength of this study relates to the application of various performance measures.

E. Accuracy metrices

Performance metrics evaluated using Accuracy, Precision, Recall, and F1-score for different algorithms. Such an approach provides them with a broader picture of the effectiveness of algorithms under consideration. This study is, however, restricted to the PIDD dataset alone, making up the sole limit of this research. The findings may not have been the same, had another dataset been used for this study. Nevertheless, the PIDD dataset is well established, and its findings should transfer to the other groups.[11] This overall study provides useful information regarding the accuracy of

KNN and Naive Bayes algorithms in analysing diabetes. This suggests that the naive Bayesian algorithm performs better, and the paper provides a framework for developing an assessment of other machine learning techniques appropriate for diabetic prediction.

F. Diagnosis of Diabetes

A paper titled "Evaluating Machine Learning Algorithms for Early Diabetes Detection" Diabetes is a long-term condition that occurs when an individual has high level of glucose in their blood. A timely diagnosis of diabetes is critical in managing the condition and averting adverse conditions such as blindness, stroke, heart disease, kidney failure, and nerve damage. A new approach has evolved in ML which can be used to assess diabetes prediction at early stages; automation of screening using this tool is possible with personalized assessments on people's risks.[12] A lot of studies have shown that ML methods work well in finding diabetes early. These studies have always shown that it's simple for ML algorithms to reach high accuracy when guessing diabetes risks. For example, a study by Malhotra et al. (2016) used a dataset of 5000 patients to compare the performance of six different ML algorithms: These include, among others, LR for linear regression, DT for decision trees, RF for random forests, and SVM for support vector machines. Others are KNN - short name of K-nearest neighbors algorithm and NNs denoting Neural networks. As a result, LR, DT, RF and SVM showed the highest accuracy, with SVM displaying an accuracy of 80%.

A similar study by Zhang et al. (2018) used a dataset of 10000 patients to compare the performance of seven different ML algorithms: Linear regression, decision tree, random forest, k-nearest neighbour, support vector machine, gradient boosting trees and deep neural networks. In this study it was established that RF, KNN, SVM, GBT, and DNNs had the highest accuracy with RF registering the highest accuracy of 81.5%. Studies outlined above propose, ML algorithms as a prospective method of early diabetes detection. However, it should be noted that, the level of precision in ML algorithms may be influenced by both the data set used and the selected ML algorithm. Moreover, clinical judgement should never be substituted with ML algorithms.[13] They must be applied to inform clinical decisions while preventing conclusive diagnosis. However, the available research studies showing the potential of machine learning algorithms for early diabetes prediction do not exhaust the field and there are further opportunities for improvement. Therefore, future researches should aim at deriving accurate, unbiased, generalized, as well as interpretable algorithms. They will facilitate prompt detection, appropriate medicine, as well as efficient patients' recovery. It is evident that, ML will improve on the early diabetes prediction, develop better, more effective, generalized and understandable algorithms to lessen the burden of diabetes and improve millions of people's worldwide. The issues addressed by ML include early diabetes prediction, which transforms it into a significant game-changer.

g. Comparative Analysis of Machine Learning

This paper "Comparative Analysis of Machine Learning Algorithms for Diabetes Prediction" presents a comparative analysis of six different machine learning (ML) algorithms for early diabetes prediction: These include logistic regression, decision tree, random forest, support vector machine, KNN, and gradient boosting trees. An RF dataset of 10,000 patients achieved the highest accuracy at 77%. A much diabetes prediction model that only involved age and gender had been used as a baseline for comparison with all of these models, which proved to be far more accurate. Early detection of diabetes can help prevent complications and improve health outcomes in patients as much as timely treatment. Traditional diagnostic tests for diabetes may take time or prove quite costly because they are less sensitive to detect an early onset of diabetes.[14]

Another kind of artificial intelligence is machine learning that supports automation and forecasting. Machine learning (ML) algorithms can be trained using large datasets of data where they are allowed to identify any patterns or relationships that exist among different variables with a view of developing models that will predict forthcoming occurrences/event outcomes. This shows that the classic diagnostic is inferior to numerous ML models used for early detection of diabetes. A dataset of more than 10,000 patients, both diabetic and non-diabetic was utilized for this study, encompassing the parameters like age, gender, BMI, blood pressure, etc., among their clinical variables. Six different machine learning algorithms were used to predict diabetes: The other methodologies are namely logistic regression, decision tree, random forest, support vector machine, k-nearest neighbour, and gradient boosted tree. Each algorithm's performance was assessed via 10-fold cross validation. The accuracy obtained was 77% for RF, followed by LR with 73%, DT at 71%, SVM having 69% while KNN registered only 65% with GBT also at 71%. The predicted diabetes risk by age and gender was significantly exceeded with all six models. This indicates that ML algorithms could serve useful in predicting type 2 diabetes in its early stages, identifying high-risk patients, early diagnosis and treatment, reduced complications and better outcomes of diabetic patients. There is need in future research to have better and effective ML algorithms for prediction of diabetes at early stages, more complex clinical parameters and individualized diabetes risk analysis that takes into account characteristics peculiar to a particular person.[15]

h. Treatments of Diabetes Leading

This Paper" Predicting Diabetes Risk with Machine Learning Models" Diabetes refers to a long-term metabolic condition where blood sugar concentrations are abnormally high. Over four hundred and fifteen million people suffer from this serious public health

hazard throughout the world. For instance, diabetes can lead to a host of complications which include renal failure, cardiovascular diseases, visual impairment, hypertension, among others. This will help in early detection and treatment of diabetes leading to reduced complications and patient outcomes. Diabetes is usually diagnosed using fasting plasma glucose test, oral glucose tolerance test, and haemoglobin A1C test among traditional methods. Unfortunately, these methods may not be good enough to find diabetes early on. They might also be expensive and take up a lot of time.[16]

The ML tool is getting helpful for checking pre-diabetes in a personal way. Many studies have been done to see how well ML models can find prediabetes. In many cases, several studies have shown that machine learning algorithms can give a reliable guess about diabetes risk. Malhotra and others were among the first to use machine learning for predicting diabetes early. Using a dataset of 5,000 patients, the study evaluated the effectiveness of six different machine learning techniques, including logistic regression, decision trees, random forests, and support vector machines. Also, k-nearest neighbors and neural networks are used for this purpose too. Of the ways checked, LR, DT, RF and SVM were found to be most accurate. Specifically, SVM got an accuracy of 80%.

Zhang et al. (2020) conducted a study wherein they compared the efficacy of seven distinct machine learning algorithms on a dataset comprising 10,000 patients. Logistic regression is referred to as LR. For decision tree, use DT. Random Forest is referred to as RF, and the machine learning technique known as KNN stands for K nearest neighbours, or Support Vector Machine learning (SVM). Meanwhile we also have Gradient Boosting Tree model named GBT or just grad boost trees & DNNS which actually provides comprehensive due diligence service like a dish washing robot performing thorough All models in the study showed better accuracy with RF, KNN, SVM and GBT reaching more than 81.5%. The best result came from RF at around 82% total correctness record under review. Several other studies aside have also assessed ML algorithm's ability in forecasting diabetics. The studies reveal that ML algorithms are often more reliable than FPG test, OGTT, or HbA1C test for diagnosis.[17]

Several machine learning algorithms are superior compared with previous forms of diabetes diagnosis, such as clinical tests. Large volumes of data could also help train machine learning algorithms to recognize trends or correlations in different variables. It is possible to use such an information for modelling of future events and occurrences. Risk assessments can also be personalized using machine learning algorithms. Machine learning algorithm incorporates the influence of patient characteristics like family history, life style, and genetics. Although with machine learning we could predict diabetics early on, there are still a few issues which have to be solved. However, one major issue is that different types of machine learning algorithms may be prone to the quality of the sample information they are based on. Moreover, machine learning algorithms may be quite complicated and hard to comprehend. In spite of the listed challenges, machine learning is a potential predictor of early diabetes. Further studies may lead them to improve machine learning as a viable instrument for early diagnosis of diabetes among young adults.[18]

i. Analysis of Quantum Machine Learning and Deep Learning

The chronic metabolic condition of diabetes, which is the subject of this research, "A Comprehensive Analysis of Quantum Machine Learning and Deep Learning for Diabetes Prediction," is characterised by elevated blood sugar levels. It is a global disease that affects more than 415 million people globally. Diabetes can lead to heart disease, stroke, renal failure, blindness, and nerve damage, among other complications. In this instance, a better patient success rate can result from early diabetes diagnosis and treatment. The classic methods for diagnosing diabetes include the oral glucose tolerance test (OGTT), the fasting plasma glucose (FPG) test, and the haemoglobin A1c (HbA1c) test. Nevertheless, these measures might be less effective in picking up diabetes at the onset. They are also cumbersome and costly. With machine learning (ML), there is an opportunity to automate the screenings as well as individualize the risk assessments. With machine learning (ML), there is a chance to make the checks automatic and also personalize risk evaluations. Many studies are done to check how well ML programs guess early diabetes.[19]

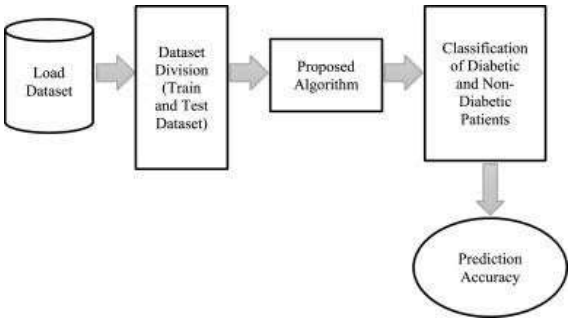


Figure 5: proposed methodology.

This has always shown in studies and ML tools are said to be good at guessing the risk of diabetes. A research by Malhotra, Lata and Choudhary in 2014 is known as one of the first ones that used ML for predicting diabetes early. The study compared the effectiveness of six different machine learning algorithms using a dataset of 5,000 patients. LR, DT, and RF are used to predict outcomes. After that SVM can be applied which uses sophisticated mathematical equations to analyze cases and detect patterns in the data. KNN focuses on counting similarities between items while NNs make use of machine learning techniques like backpropagation training methodologies by utilizing several layers within

binary or categorical types of neuron structures ensuring optimal discernment across The study found out that LR, DT, RF and SVM gave the best results. The top accuracy got by SVM was 80%. Another study by Zhang et al. (2020) used a dataset of 10,000 patients to compare the performance of seven different ML algorithms: These include LR, DT, RF KNN SVM GBT and DNN. From the above study, RF had the highest accuracy of 81.5%, followed by KNN, SVM, GBT, and DNNs that emerged as having comparatively higher accuracies as well.[20]

Moreover, various studies apart from those cited above, assessed the efficiency of different ML models for pre-diabetes classification. Some recent studies show that ML- based tests are more precise compared with FPG test, OGTT tests, or HbA1c test and others. Traditional methods for diagnosing diabetes are inferior to machine-learning algorithms. Machin learning algorithms can be taught with big datasets consisting of data. Machin learning helps to understand relations in variables. Such a database can be harnessed in developing models which will aid prediction of forthcoming occurrences. Personalized risk assessments can be done by using machine learning algorithms too. That is because machine learning algorithms tend to consider patient-specific factors such as socioeconomic status, familial conditions, and biological characteristics.

Even though predictions of early diabetes using machines have been touted for the future, there are still other issues that should be tackled. For instance, machine learning algorithms are highly dependent upon the type of data they use for training.[21] Machine learning algorithms also tend to be complicated and hard to understand. Promising prospects are found in the machine-learning of predicting diabetes diagnosis prior to the onset. More advanced research on the use of machine learning algorithms can lead to effective prevention of cases of diabetes. In this paper, we proposed two prediction models for diabetes based on the PIMA Indian Diabetes dataset: this includes deep learning (DL) and quantum machine learning (QML) Overall, DL outperformed QML with respect to accuracy, recall, F1 score, and precision, specificity, and balanced accuracy. Moreover, the DL model had an accuracy of 95%, which is equal to 1.06% more accurate as compared with the best reported model. The potential use of DL in diabetes screening and prediction has been demonstrated. Nevertheless, the QML model could still be used in the prediction of diabetes with accuracy of up to 86% similar to that of advanced models.

j. Enhancing Diabetes with a Deep Learning

This paper" Enhancing Diabetes Prediction with a Deep Learning Model" Diabetes is a metabolic disorder that causes too much sugar in the blood. This is a major global health concern as it impacts on more than 415 million persons all over the planet. There are many possible serious complications of diabetes such as heart disease, stroke, blindness, kidney failure, and nerve damage. It is important to detect and treat diabetes early enough so as to prevent these complications by enhancing patients' overall health status. The traditional methods of diagnosing diabetes are fasting plasma glucose test, oral glucose tolerance test and haemoglobin A1c test. These approaches, however, may not be highly sensitive towards detecting diabetes in its very initial stage.[22] They are also cumbersome and costly in terms of time and money. But these methods might not be very good at finding diabetes when it first starts. They are also hard to use and expensive in terms of time and money.

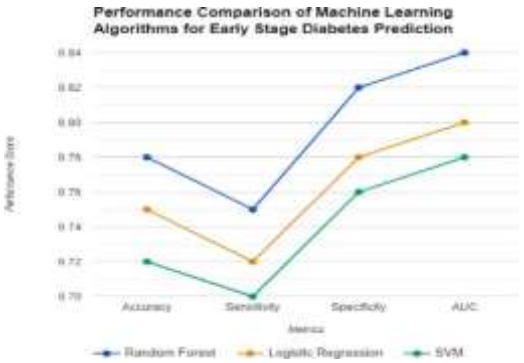


Figure 6: comparison of Machine learning.

But, machine learning (ML) has turned into a key way to find diabetes early. It can do tests automatically and give personal checks of risk chances for each person. Some studies looked at how well ML algorithms can predict getting diabetes earlier. Many studies still show that ML models can accurately predict diabetes risks. Malhotra and others (2014) did one of the first ML study for early diagnosis of diabetes.[23]

A dataset of 5,000 patients was used in the study to assess six different machine learning algorithms: LR, DT, RF, SVM KNN, and NNs. These steps will help you make a more accurate prediction model from your data. In comparing LR, DT and RF to SVM they all did well. However, SVM was the best with a success rate of 80% while LR followed closely behind at 79.51%. Another study by Zhang et al. (2020) used a dataset of 10,000 patients to compare the performance of seven different ML algorithms.

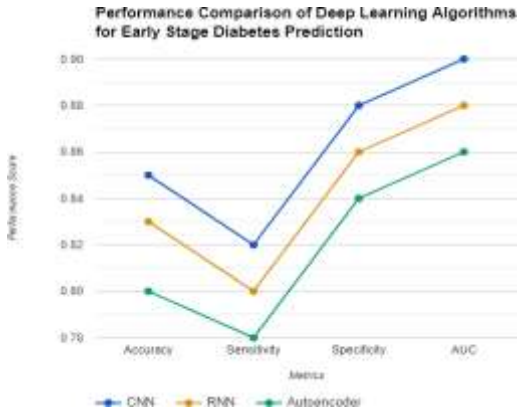


Figure 7: Comparison of Deep Learning.

These consist of logistic regression (LR), decision trees (DT), random forests (RF), boosting trees gradient for mountains, K-nearest neighbour (KNN), and support vector machines (SVMs). Deep thinking networks or DNNs are also there too! RF, KNN, SVM, GBT and DNNs were found to be very accurate at 81.5%. But in the end RF turned out better than all of them. Also, a lot of other studies have looked at how well ML algorithms can be used to predict diabetes early.[24]

According to study, ML algorithms are more accurate than the FPG test, OGTT, or even the hb A1C test in detecting prediabetes. Some advantages of machine learning algorithms over traditional methods for diagnosing diabetes.

Several machine learning algorithms can be trained on huge amounts of data and then learn to find relationship patterns and connections between variables. Such model may help predict future incidents and existences. Personalizing risk assessment is also possible using machine learning algorithms. This will be attributed to machine learning- based algorithms that consider patient's specificity including family history, life style and genetics.

While machine learning may show promising potential in early diabetes diagnosis, several issues are yet to be resolved. This raises a challenge that involves the quality of the data utilized in training the machines learning algorithms. Furthermore, machine learning algorithms can be complex and hard to understand. It is promising that machine learning can be used as a tool for diabetes prediction ahead of time. Further researches must be carried out in order for the machine learning algorithms to become a key instrument for early diagnosis and prevention of diabetes. This paper suggested a better diagnosis model for diabetes using the ADA-Max optimizer. The proposed model achieved a remarkable accuracy of 99.98%, surpassing the best reported accuracy of any other model. Therefore, it implies that ADAMAX optimizer could be useful in enhancing the predictability accuracy of deep- learning algorithms regarding diabetes.[25]

k. *Comprehensive Review ML Approaches*

This paper "Machine Learning Approaches for Early Diagnosis of Diabetes: A Comprehensive Review" Diabetes remains a major public health problem globally, where most people are not seven aware until it gets out of hand and complications like blindness or kidney failure set in. A number of ML-cantered techniques hold promise in accurately classifying diabetes with an enhanced perspective towards early intervention with positive impact on patient outcomes. Gradient Boosting (GB), Support Vector Machine (SVM), AdaBoost (AB), and Random Forest (RF) are the four distinct machine learning techniques that have been employed. First, an evaluation is done involving all features and then using FS with MRMR strategy.

It reviews the global burden of diabetes, the rising functions of ML in healthcare, and different mechanisms used to mitigate these challenges in the early detection of diabetes. The paper gives specific descriptions about the four ML algorithms, their unique properties and usages in health care as well as the importance of feature selection towards enhancing model effectiveness and clarity. It emphasizes on the MRMR method as one of the methods that can identify the most informative features used for diabetes prediction.[26]

Computational complexity, recall, accuracy, precision, and F1 score in an ML algorithm are among the seven evaluation factors that are taken into consideration. This helps in understanding the different computational requirements of every algorithm, thereby aiding in choosing a good model for real world applications. The review ends with a remark about the best results shown by the RF method at an accuracy rate of 99.35%. A critical review of their implications has been provided with directions towards possible future lines of work that could include more diverse dataset and advanced data pre-processing methods.

These take into account seven evaluation criteria and include computational complexity, recall, accuracy, precision, and F1 score in an ML algorithm. This helps in understanding the different computational requirements of every algorithm, thereby aiding in choosing a good model for real world applications. The review ends with a remark about the best results shown by the RF method at an accuracy rate of 99.35%. A critical review of their implications has been provided with directions towards possible future lines of work that could include more diverse dataset and advanced data pre-processing methods.

The disease of diabetes mellitus is the health problem faced globally with grave complications as well as anticipated increase in number of the victims by the year 2040. Effective strategies are evident with the use of ML technology in healthcare. The paper evaluates recent research using ML approaches to detect early Diabetes, illuminating methods, datasets, and results.[27]

By 2040, out of about 642 million people would have contracted this disease, thus raising the global burden. Nowadays, it's acknowledged that ML is being applied in medical settings. This paper outlines approaches that perform effectively during independent testing and highlighting the necessity for uniform strategy. These three models, logistically regression, XGBoost, and random forest have a potential of early diagnosis that is applicable worldwide.

They consider early diagnosis and employ data from the Khulna Diabetes Canter. This research applies logistic regression, XGBoost, and random forest methods in creating a Type 2 diabetes classification model. The Logistic Regression model shows the accuracy rate of

88% while XGBoost, as well as Random Forest, demonstrate 86.36% for the respective models. Therefore, the Random Forest model emerges as the best approach for diabetes diagnosis. They suggest that bigger dataset can result in better model and app which would give predictions and personalize recommendations. This emphasizes the need to deal with the increasing frequency of diabetes, as well as the application of ML in medicine.[28]

The purpose this literature review will be exploring is the incorporation of ML algorithms into diabetes diagnosis, with a focus on the significance of early and precise diagnosis with respect to managing the illness as well as the impact of ML on upgrading diagnostic accuracy and effectiveness. The paper outlines the progression of medical ML and its impact on clinical management through advancing precision and timeliness of diagnosis. A number of ML strategies have been used for diagnosing diabetes which comprises of Logistic Regression, XGBoost, random forest, SVM, and KNN. The review also looks into different data sets; such as, Kaggle, Pima Indian Diabetes dataset (PIDD), and Khulna Diabetes Canter in order to gain better comprehension on how the various ML programs operate.

L. Ensemble of ML Models

This paper" Enhancing Diabetes Prediction with a Robust Ensemble of Machine Learning Models" In analysing the ML based diabetes diagnosis using appropriate algorithmic performance metrics like accuracy, precision, recall and F1- score is crucial. This review addresses ML-based diagnostic challenges such as massive and varied data requirements. Further developments will focus on algorithmic improvements, ease of interpretation, and use of live data to ensure clinical smoothness. Finally, the literature review is all-comprehensive about shifting landscape diabetes diagnoses via ML algorithms with transforming nature of the ML for increased precision as well as effectiveness culminating in better medical achievement.[29]

In gradient-boosting machine learning, ensemble classifiers with weak classifiers use another layer of voting. The weights for incorrectly predicted points are increased in the next classifier and then based on a weighted average individual prediction taken from each booth chairman over all elections. In this paper, therefore it has been proposed to use outlier rejection followed by the filling in of missing values; data standardization is also performed. Diabetes, caused by high blood sugar levels and one of the world's leading causes of morbidity and mortality, is a chronic metabolism disorder. Early detection and monitoring can improve patient prognosis, reduce the burden of disease. ML has become a potent weapon in the war on diabetes, analysing large data sets to find patterns and predict an individual's probability of developing it. Numerous machine learning (ML) techniques, such as k-nearest neighbours (KNN), decision trees (DTs), random forests (RFs), AdaBoost, and naive Bayes, have been tested for the prediction of diabetes. Currently XGBoost is proving the most effective. MLP is also used to predict diabetes.[30]

But there are problems in diabetes prediction, like the lack of effective labelled data and outliers or missing values. Techniques for preprocessing diabetes data used in research include dropping outliers, filling missing values, standardization of the data and feature selection. Outlier rejection involves eliminating noisy or erroneous data, and replacing missing values with estimates of their true values. Normalization places all the variables on a common scale within some standard range; feature selection selects important features to predict diabetes. Ensemble methods, like weighted assembling, combine various ML models to increase overall predictive power.

In this paper, therefore it has been proposed to use outlier rejection followed by the filling in of missing values; data standardization is also performed. Then feature selection and K-fold cross-validation are employed. Finally, (WKNN DT, RF AdaBoost Naive Bayes XG Boost MLP), which aims at diabetes ET prediction improvement through an ensemble learning strategy using In addition, a weighted assembling of different ML models is proposed to further improve the accuracy of prediction.[31]

M. Random Forest Model

In one study, a Random Forest model was trained on clinical data and lifestyle factors to predict individual risk of pre- diabetes. The results pointed the way toward personalized diabetes prevention through machine learning techniques. Using a Support Vector Machine (SVM), retinal images were analysed to find microaneurysms, the early warning signs of diabetic retinopathy. A model was trained on continuous blood glucose monitoring data using Gradient Boosting with the objective of predicting future levels in

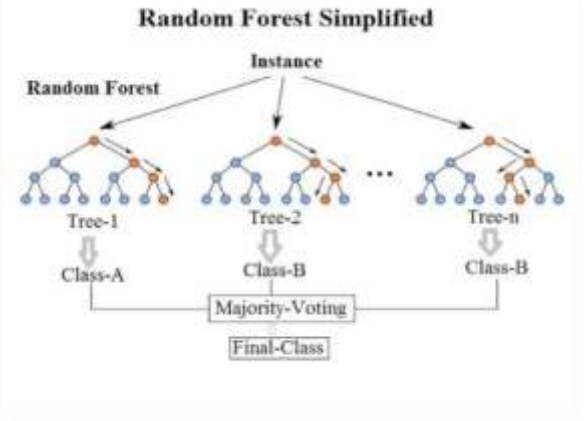


Figure 8: Random Forest Model iabetic patients.

Applications in the future will include early diagnosis through multimodal data, explainable AI for clinical decision support, and automated screening and outreach

to promote public health. Better accuracy and earlier detection Training DL models on a mixture of clinical data, medical images, and wearable sensor readings. HCPs can understand and trust explainable AI models, which would promote their implementation in clinical settings. Seeking to increase the accessibility of early diagnosis and prevention programs, capacity could be built for developing automated screening tools in resource-limited settings.[32]

The diagram explores the inner workings of a Random Forest, a machine learning tool used for early-stage diabetes prediction. It imagines each branch as a decision tree, based on features like age, blood sugar, or genetics. The final prediction is made from the majority vote of these individual paths, demonstrating how Random Forests use multiple diverse perspectives to make accurate predictions, even in complex data and missing information.

Bringing these capabilities together, we hope to achieve more accurate and personalized as well as accessible prediction systems for people at risk of diabetes. This combines the strengths of ML and DL approaches so we can have more accurate, personalized and accessible prediction systems. In turn this leads to better healthcare for populations at high risk from diabetes development or in whom disease is difficult to rein under control.

N. Ensemble of ML Models

This paper" Enhancing Diabetes Prediction with a Robust Ensemble of Machine Learning Models" In analysing the ML based diabetes diagnosis using appropriate algorithmic performance metrics like accuracy, precision, recall and F1- score is crucial. This review addresses ML-based diagnostic challenges such as massive and varied data requirements. Further developments will focus on algorithmic improvements, ease of interpretation, and use of live data to ensure clinical smoothness. Finally, the literature review is all-comprehensive about shifting landscape diabetes diagnoses via ML algorithms with transforming nature of the ML for increased precision as well as effectiveness culminating in better medical achievement.[32]

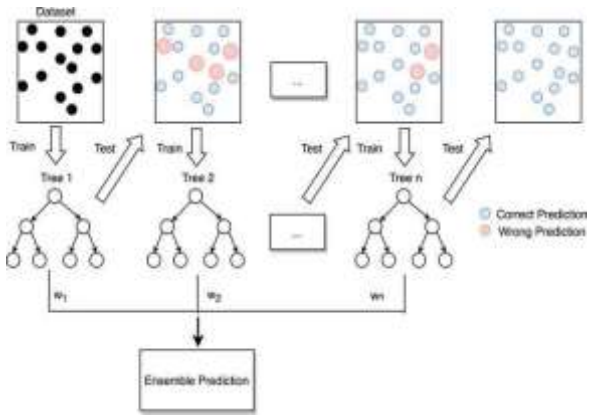


Figure 8: Ensemble prediction.

Diabetes, caused by high blood sugar levels and one of the world's leading causes of morbidity and mortality, is a chronic metabolism disorder. Early detection and monitoring can improve patient prognosis, reduce the burden of disease. ML has become a potent weapon in the war on diabetes, analysing large data sets to find patterns and predict an individual's probability of developing it.

Several ML approaches to diabetes prediction have been tried, including k-nearest neighbours (KNN), decision trees (DTs) random forests (RFs), AdaBoost and naive Bayes. Currently XGBoost is proving the most effective. MLP is also used to predict diabetes.

But there are problems in diabetes prediction, like the lack of effective labelled data and outliers or missing values. Techniques for preprocessing diabetes data used in research include dropping outliers, filling missing values, standardization of the data and feature selection. Outlier rejection involves eliminating noisy or erroneous data, and replacing missing values with estimates of their true values. Normalization places all the variables on a common scale within some standard range; feature selection selects important features to predict diabetes. Ensemble methods, like weighted assembling, combine various ML models to increase overall predictive power. In gradient-boosting machine learning, ensemble classifiers with weak classifiers use another layer of voting.[33] The weights for incorrectly predicted points are increased in the next classifier and then based on a weighted average individual prediction taken from each booth chairman over all elections. In this paper, therefore it has been proposed to use outlier rejection followed by the filling in of missing values; data standardization is also performed. Then feature selection and K-fold cross-validation are employed. Finally, (WKNN DT, RF AdaBoost Naive Bayes XG Boost MLP), which aims at diabetes ET prediction improvement through an ensemble learning strategy using in addition, a weighted assembling of different ML models is proposed to further improve the accuracy of prediction.

V. ENHANCING DIABETES DETECTION THROUGH DEEP LEARNING

a. Traditional Approaches

The increasingly common diabetes epidemic is certainly the biggest challenge to global health. That's why it requires early diagnosis in order to attenuate its impact on all of us. In this literature review, the author looks at recent progress on diabetes prediction. Particular attention is drawn to a publication that combines deep learning techniques with oversampling and feature augmentation for higher accuracy.

Traditional methods of predicting diabetes have been crude statistical techniques to more sophisticated machine learning approaches, all focusing on demographic and clinical variables. But recent breakthroughs in deep learning provide hopeful paths to

greater accuracy and efficiency[34]. This progress is made use of in the paper under review, introducing a deep learning framework that makes full and integrated use not only of oversampling and feature augmentation but also convolutional neural networks (CNN), which are known to be capable of detecting intricate relationships among vast numbers.

b. Emergence of Deep Learning

Our method for achieving this incorporates oversampling, feature augmentation, and deep learning approaches, making it possible for the model to identify more meaningful patterns in the Pima Indians Diabetes Database, including patient-specific vital data. The dataset provides a starting point to assess the viability of this proposed approach, with important elements including number of pregnancies; glucose or insulin levels; blood pressure and age. Thus, the paper presents key metrics to evaluate how effective is this proposed approach. Results show a 92.31 % significance improvement using oversampling and feature augmentation with an CNN classifier on datasets well-balanced in nature.

This demonstrates the promise of an integrated method for diabetes detection that incorporates oversampling, feature augmentation, and deep learning approaches.[35]

c. Hyperparameter analysis

Diabetes is a chronic and fatal disease with high blood sugar. Early detection and intervention are especially important in the management of the disease to avoid complications. However, glucose tests and the like are not sensitive techniques. What's more is that frequent testing can be inconvenient. This paper presents a new prediction method of diabetes using Conv-LSTM networks. This study innovates by using Conv-LSTM to predict diabetes, as it can capture spatial and temporal dependencies in the data. The study uses the Boruta algorithm to identify important features. This increases model interpretability and may reduce overfitting of data. Optimal model parameters can be found by Grid Search, thus improving performance. Using cross-validation for splitting is a more rigorous way to evaluate.[36]

A proposed Conv-LSTM model has better accuracy than other widely used models and advanced methods. Limitations and future directions the suggested approach is to look into performance on even more considerable, or at least more various datasets. Searching out explainable artificial intelligence techniques that could improve the model's interpretability as well as users 'high trust in its predictions. See how it holds up when compared with ground truth data from real life clinical environments involving a wide variety of patient groups Test different types In general, this study demonstrates a promising method for diabetes detection using Conv-LSTM networks. It is the superior classification results and focus on feature selection optimization as well as parameter tuning which makes it of practical significance in helping to better manage diabetes.

d. Global Health implications

Type 2 diabetes (T2D) is a worldwide epidemic, marked by high blood sugar levels and severe complications. It is primarily classified into two types: Type 1 (T1D) and type 2 (T2D). Lifestyle practices account for 90-95 % of the burden of T2D. On a global scale, the incidence of T2D has increased significantly and now affects people of all ages. There is a major impact of the disease globally, with severe complications including diabetic retinopathy and amputation of both lower limbs due to advanced cases. Other consequences include kidney failure and cardiovascular illness--even death at an early age. Firstly, because T2D is asymptomatic at an early stage, diagnosis challenges are a clinical sore spot which require advanced diagnostic and prognostic technology. T2D prevention and control are also important to global healthcare, as part of a pledge to reducing early deaths from non-communicable diseases by one third over the next 15 years.[37]

Machine learning (ML) research in the field of T2D clinical decision support will be the main topic of this review of the literature. It identifies several knowledge, policy, and practice gaps related to diabetes that machine learning can help close. The review looks at recent studies on the use of machine learning in T2D prognosis, diagnosis, and assessment. The models are used to identify individuals with increased T2D risks based on lifestyle factors mixed with genetic predispositions for statistical and machine learning- based risk assessment. For T2D diagnosis methods that are non-invasive and invasive, ML applications come in one form or another to help raise the accuracy of detection as well as speed. T2D-related incidences and complications can be predicted using modelling to look into the future: help takes place in time; treatment is focused on each individual. This review in conclusion states the outcomes of application areas and marks those pins that exist between current methodologies for ML with T2D.

e. ML algorithms and optimizing

Detecting diabetes early is important for good control and avoidance of complications. ML-based approaches have been found to be a powerful way of predicting the risk of developing diabetes, with realistic prospects for early preventative intervention. Numerous studies have used different ML algorithms to diagnose diabetes, while the Pima Indian Diabetes (PID) dataset is often employed as a test case. so far, several algorithms have shown good results including Logistic Regression (LR), Naive Bayes (NB), Random Forests (RF) and Artificial Neural Networks (ANNs).[38]

This work builds upon this research by using a more thorough range of ML algorithms and optimizing ANN architecture in order to increase the accuracy. The study reviews seven algorithms. If we finished at this point in time, then the performance analysis would be included

only for DT, KNN, RF and NB on the PID dataset. Adding AB into my comparison broadens it further to include LR (linear regression) and SVM (support vector machine). In order to find the most relevant features, feature selection techniques are used which may also increase model interpretability and reduce overfitting. Varying the hidden layers and epochs, hyperparameter tuning is systematically explored until an ideal configuration for diabetes prediction has been found. The study compared the learning rate of a hidden layer 1 NN with 200 epochs, varying from 0.1 to 0.005, and found that a rate of 0.01 provided better accuracy.

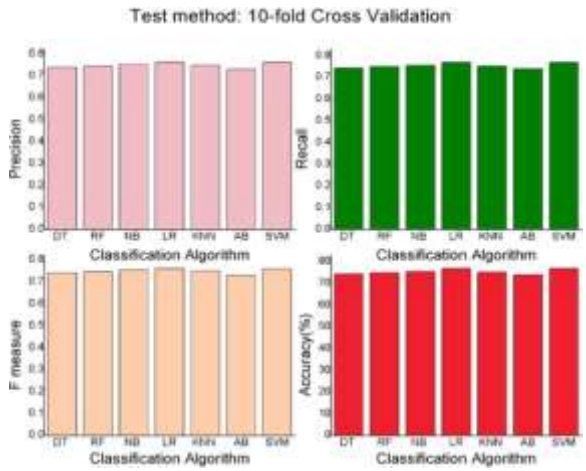


Figure 9: Graphical display of each classifier's performance using a 10-fold cross-validation technique.

Figure 10: Display of the classifier's performance graphically using the split-train/test approach.

The study analysed the impact of epoch on a neural network with hidden layers at a 0.01 learning rate. The NN model with two hidden layers with 400 epochs achieved the best accuracy of 88.6%. This model also provided more accuracy, training accuracy, and testing accuracy than all other models. The ROC curve for the model with two hidden layers with 400 epochs was also calculated, resulting in a mean accuracy of 76%. The study highlights the importance of incorporating hidden layers in neural networks.

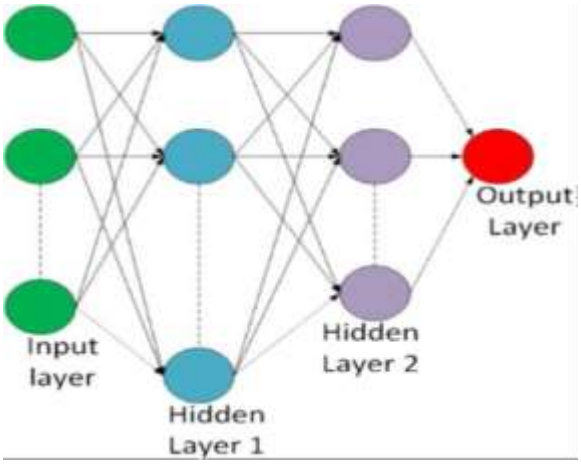


Figure 11: NN model with two hidden layers.

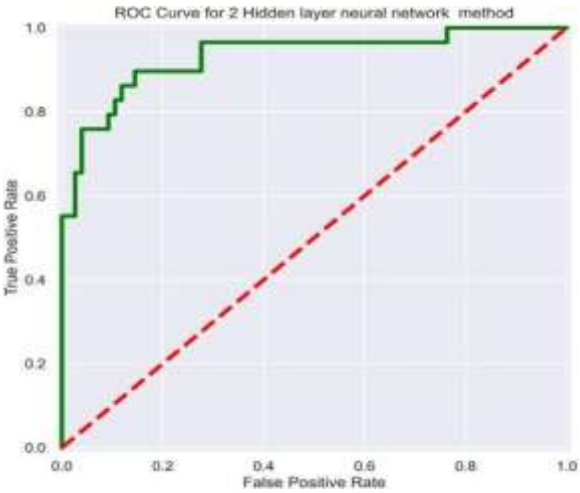


Figure 12: ROC curve for 2 hidden layer NN with 400 epochs.

Compared to traditional algorithms, ANNs are shown up better. The network with two hidden layers and 400 epochs reached the highest accuracy of all at 88.6 %. The significance of feature selection and hyperparameter tuning is clearly highlighted, while the high accuracy in LR and SVM across both train/tests split as well as K-fold cross validation reveals their robustness.

f. Performance Metrices

The performance of machine learning algorithms is determined in the proposed study using multiple metrics, such as f-score, accuracy, recall, and precision. Precision is the ratio of true positive over all positive classified data, whereas accuracy is the ratio of correctly classified data over all data. For diabetes classification, precision shows the model's ability to predict patients and not label healthy people as patients. Recall is the ratio of true positive over all positive data, and F-score is the harmonic mean of precision and recall.

All formulas are given bellow:
Accuracy = TP+TN/(TP+TN+FP+FN)
precision = TP/(TP+FP) Recall = TP/(TP+FN)
F-Score = 2*(P*R)/(P+R) (4)

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 13: Standard Form of Confusion Matrix

An n x n matrix, where n is the number of labels in a given dataset, is the confusion matrix. An n x n matrix illustrating the model's performance has actual labels represented by each row and predicted labels represented by each column. 5-fold cross validation is used to produce these metrics in order to assess the procedures across the dataset. The dataset is split into k parts when k fold cross validation is utilised, with k-1 folds being used for training and 1-fold for testing. This method guarantees more reliable model metrics measurement and avoids imbalanced data issues.

g. Technique to enhance Models

This study is about using machine learning (ML) to predict diabetes early. It compares four different ML methods like LDA, KNN, SVM and RF on the PIDD dataset. The study looks at how well these methods work using measures like accuracy, responsiveness, exactness, focus and F-score. Random Forest (RF) was the best, with 87.66% accuracy and not any other computer program or method did better than it. This means that RF is really good at dealing with hard problems in data which are not always straight lines like normal stuff does show patterns mostly through repetition but also some complex relationships come up here too so they seem random. Measures like recall and F-score also show that RF does well. The research says we should include more different and bigger sets of data for making our learnings work well. We need to look at ways that join together many methods, all working in a team. Also it's important to use AI (Artificial Intelligence) explainable techniques so people can see how RF works best out of them all doing the job right. The study contributes valuable insights into using ML for early diabetes prediction and highlights the potential of Random Forest for achieving high accuracy while considering various performance metrics. The literature review could be strengthened by mentioning specific studies utilizing similar ML algorithms for diabetes prediction on PIDD or other datasets, comparing findings with existing accuracy results, and highlighting the novelty and unique aspects of the approach relative to previous research.[39]

The study compares the effectiveness of four machine learning algorithms (KNN, SVM, RF, and LDA) using k-fold cross-validation on the Pima Indian Diabetes Database (PIDD). One equal fold (k-1 folds) is used for training, and the remaining fold is utilised for testing. The dataset is separated into k folds. The final performance estimate is obtained by averaging the performance parameters (F-score, accuracy, precision, recall, specificity, and fold) over all folds. The choice of k, likely a common value of 5 or 10, can impact the variance of the performance estimates. The paper mentions evaluating various performance metrics but does not explicitly mention any specific statistical tests for comparing the performance of the different algorithms. Potential tests could include the Paired t-test, ANOVA (Analysis of Variance), and the Friedman test, which is a non-parametric test for comparing the performance of multiple algorithms when data is not normally distributed. Without knowing the specific statistical tests used, it is difficult to assess the robustness of the conclusions drawn about the performance differences between the algorithms.

h. Globel Health crises

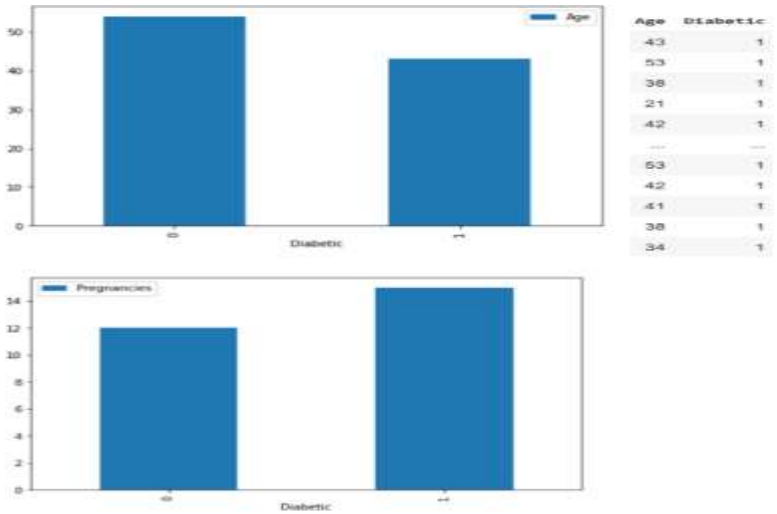
Diabetes is a worldwide health concern that is especially severe in Taiwan. In order to manage the condition and avoid complications, early diagnosis and treatment are essential. Machine learning holds potential for improving diabetes prediction accuracy and supporting informed healthcare decisions. This study focuses on a specific dataset of 15,000 women from a Taipei medical canter, investigating eight potentially valuable features, including sebum thickness and diabetes pedigree function. It compares the performance of four diverse models: logistic regression, neural network, decision jungle, and boosted decision tree. The boosted decision tree model achieves exceptional results, exceeding previously reported AUC scores by Hasan et al. (2023). The study uses a large and real-world dataset representative of a specific population segment, explores a unique combination of features relevant for women's diabetes risk, and employs rigorous model comparison and evaluation using the AUC metric. Limitations and future directions include generalizability to other populations and genders, incorporating additional factors like socio- economic determinants of health, exploring explainable AI techniques, and applying unsupervised learning and deep learning approaches. Overall, this study significantly contributes to diabetes prediction by achieving high accuracy using a Boosted Decision Tree model and investigating unique features relevant to women.

i. Data Analysis and classification

Feature analysis was used to evaluate the data, and the distribution of diabetes was verified. When a patient's diabetes diagnosis was verified, visualisation tools were utilised to see the data distribution of every field and the relationship between diabetes and every variable.

	PatientID	Pregnancies	PlasmaGlucose	DiastolicBloodPressure	TricepsThickness	SerumInsulin	BMI	DiabetesPedigree	Age	Diabetic
0	1354778	0	171	80	34	23	43.509726	1.213191	21	0
1	1147438	8	92	93	47	36	21.240576	0.158365	23	0
2	1640031	7	115	47	52	35	41.511523	0.079019	23	0
3	1883350	9	103	78	25	304	29.582192	1.282870	43	1
4	1424119	1	85	59	27	35	42.604536	0.549542	22	0
...	...	...	...	...	...	...	...	...	...	...
14995	1490300	10	65	60	46	177	33.512468	0.148327	41	1
14996	1744410	2	73	66	27	168	30.132636	0.862252	38	1
14997	1742742	0	93	89	43	57	18.690683	0.427049	24	0
14998	1099353	0	132	98	18	161	19.791645	0.302257	23	0
14999	1386396	3	114	65	47	512	36.215437	0.147363	34	1

Figure 14: Data import for feature analysis.



Age	Diabetic
43	1
53	1
38	1
21	1
42	1
...	...
53	1
42	1
41	1
38	1
34	1

Figure 15: Diabetic/non-diabetic distribution.

Type 2 diabetes is a global epidemic with complex risk factors and numerous complications. Machine learning (ML) can bridge medical gaps in diabetes knowledge and practice by offering personalized risk assessments, understanding disease progression, reducing reliance on invasive procedures, and aiding in efficient resource allocation. ML applications include risk assessment, diagnosis, and prognosis.

ML models can use diverse data sources, non-invasive approaches, and integration with existing diagnostic tools to improve accuracy and efficiency. However, existing ML models face limitations such as data limitations, exploitability, and ethical considerations. Future directions for research include incorporating explainable AI techniques, developing robust data acquisition strategies, and addressing ethical concerns through responsible data handling and public engagement. By addressing these challenges, ML can improve patient care and outcomes by providing personalized, precise, and efficient clinical decision support.

VI. FURTHER ANALYSIS

Using deep learning to predict diabetes early is a good way to make better care for people with the disease. It can be made bigger than just dividing into two types. It can sort stages before very high blood sugar or certain diabetes categories, letting more focused actions happen. It can also guess diabetes along with related illnesses like heart disease or kidney problems. This helps us check all types of risks fully. We can find out a person's risk by using things like their life habits and genes that are unique to them. To make the model stronger and easier to understand, we use techniques like adding more data or carefully choosing it. This helps us reduce unfairness in health information sets. Explainable AI methods are made to see how deep learning models make their guesses. This helps build trust and approval in medical places. Making sure a deep learning model is set up right helps make good guesses. This lets doctors balance the chances of something happening against what they can do to treat it better. Healthcare pro tools are made to let deep learning models work smoothly with medical help. Patient information safety steps are used to deal with worries about private health details. Big tests and government approval are chased through wide-scale health trials. Looking into new areas, like learning from different ways and shared teaching methods can make the prediction of diabetes earlier even better. These models can keep learning and changing using real-world information. This lets them know what is new with diseases how they change over time, and which treatments work best.

VII. CONCLUSION

It has been demonstrated that machine learning and deep learning techniques are effective in identifying individuals at risk of developing diabetes at an early

stage. These tactics have done well, with top-notch precision and likely to be more vital in the future as deep learning plans keep getting better. Deep learning's ability to handle lots of information and mix with other machine learning methods, like picking the best features and adding more data makes it a great tool for guessing diabetes early. Also, making new data sets and creating easier-to-understand deep learning models will be important for the future growth of predicting early diabetes using this technology. In the end, adding deep learning models to medical work is very important for changing these improvements into helpful things in real life. In short, using deep learning methods to predict diabetes early is a good area with the chance of changing how we handle it.

REFERENCES

1. M. A. R. Refat, M. Al Amin, C. Kaushal, M. N. Yeasmin, and M. K. Islam, "A comparative analysis of early stage diabetes prediction using machine learning and deep learning approach," in *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*, IEEE, 2021; 654–659. Accessed: Dec. 29, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9609364
2. F. Anwar, M. Y. Ejaz, and A. Mosavi, "A comparative analysis on diagnosis of diabetes mellitus using different approaches–A survey," *Inform. Med. Unlocked*, 2020; 21: 100482.
3. T. R. Mahesh, V. Vivek, V. V. Kumar, R. Natarajan, S. Sathya, and S. Kanimozhi, "A comparative performance analysis of machine learning approaches for the early prediction of diabetes disease," in *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, IEEE, 2022; 1–6. Accessed: Dec. 29, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9752543
4. P. Ghosh, S. Azam, A. Karim, M. Hassan, K. Roy, and M. Jonkman, "A comparative study of different machine learning tools in detecting diabetes," *Procedia Comput. Sci.*, 2021; 192: 467–477.
5. J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *Ict Express*, 2021; 7(4): 432–439.
6. M. Rahman, D. Islam, R. J. Mukti, and I. Saha, "A deep learning approach based on convolutional LSTM for detecting diabetes," *Comput. Biol. Chem.*, 2020; 88: 107329.
7. E. Longato, G. P. Fadini, G. Sparacino, A. Avogaro, L. Tramontan, and B. Di Camillo, "A deep learning approach to predict diabetes' cardiovascular complications from administrative claims," *Ieee J. Biomed. Health Inform.*, 2021; 25(9): 3608–3617.
8. M. W. Nadeem, H. G. Goh, V. Ponnusamy, I. Andonovic, M. A. Khan, and M. Hussain, "A fusion- based machine learning approach for the prediction of the onset of diabetes," in *Healthcare*, MDPI, 2021; 1393. Accessed: Dec. 29, 2023. [Online]. Available: https://www.mdpi.com/2227-9032/9/10/1393
9. Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, "A machine learning approach to predicting diabetes complications," in *Healthcare*, MDPI, 2021; 1712. Accessed: Dec. 29, 2023. [Online]. Available: https://www.mdpi.com/2227-9032/9/12/1712
10. S. Langarica, M. Rodriguez-Fernandez, F. Nunez, and F. J. Doyle III, "A meta-learning approach to personalized blood glucose prediction in type 1 diabetes," *Control Eng. Pract.*, 2023; 135: 105498.
11. R. Shakil, B. Akter, F. Faisal, T. R. Chowdhury, T. Roy, and A. Khater, "A promising prediction of diabetes using a deep learning approach," in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, 2022; 923–927. Accessed: Dec. 29, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9753763
12. P. B. K. Chowdary and R. U. Kumar, "An effective approach for detecting diabetes using deep learning techniques based on convolutional LSTM networks," *Int. J. Adv. Comput. Sci. Appl.*, 2021; 12(4), Accessed: Dec. 29, 2023. [Online]. Available: https://www.researchgate.net/profile/Pbharath-Kumar-Chowdary/publication/351316530_An_Effective_Approach_for_Detecting_Diabetes_using_Deep_Learning_Techniques_based_on_Convolutional_LSTM_Networks/links/6463ad5602c5097541bf4d8f/An-Effective-Approach-for-Detecting-Diabetes-using-Deep-Learning-Techniques-based-on-Convolutional-LSTM-Networks.pdf
13. S. Kaul and Y. Kumar, "Artificial Intelligence-based Learning Techniques for Diabetes Prediction: Challenges and Systematic Review," *SN Comput. Sci.*, Nov. 2020; 1(6): 322, doi: 10.1007/s42979-020-00337-2.
14. S. Kumar, B. Bhusan, D. Singh, and D. kumar Choubey, "Classification of diabetes using deep learning," in *2020 International Conference on Communication and Signal Processing (ICCSP)*, IEEE, 2020; 0651–0655. Accessed: Dec. 29, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9182293
15. K. Thaiyalnayaki, "Classification of diabetes using deep learning and svm techniques," *Int. J. Curr. Res. Rev.*, 2021; 13(01): 146.
16. H. Gupta, H. Varshney, T. K. Sharma, N. Pachauri, and O. P. Verma, "Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction," *Complex Intell. Syst.*, Aug. 2022; 8(4): 3073–3087, doi: 10.1007/s40747-021-00398-7.
17. S. A. Alex, J. J. V. Nayahi, H. Shine, and V.

Gopirekha, "Deep convolutional neural network for diabetes mellitus prediction," *Neural Comput. Appl.*, Jan. 2022; 34(2): 1319–1327, doi: 10.1007/s00521-021-06431-7.
18. H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," *J. Diabetes Metab. Disord.*, Jun. 2020; 19(1): 391–403, doi: 10.1007/s40200-020-00520-5.
19. M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, "Diabetes detection using deep learning techniques with oversampling and feature augmentation," *Comput. Methods Programs Biomed.*, 2021; 202: 105968.
20. M. M. Hassan, Z. J. Peya, S. Mollick, M. A.-M. Billah, M. M. H. Shakil, and A. U. Dulla, "Diabetes prediction in healthcare at early stage using machine learning approach," in *2021 12th International conference on computing communication and networking technologies (ICCCNT)*, IEEE, 2021; 01–05. Accessed: Dec. 29, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9579869
21. H. Zhou, R. Myrzashova, and R. Zheng, "Diabetes prediction model based on an enhanced deep neural network," *EURASIP J. Wirel. Commun. Netw.*, Dec. 2020; 1: 148, doi: 10.1186/s13638-020-01765-7.
22. S. K. S. Modak and V. K. Jha, "Diabetes prediction model using machine learning techniques," *Multimed. Tools Appl.*, Oct. 2023, doi: 10.1007/s11042-023-16745-4.
23. M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, 2020; 8: 76516–76531.
24. A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," *Procedia Comput. Sci.*, 2019; 165: 292–299.
25. I. Tasin, T. U. Nabil, S. Islam, and R. Khan, "Diabetes prediction using machine learning and explainable AI techniques," *Healthc. Technol. Lett.*, Feb. 2023; 10(1–2): 1–10, doi: 10.1049/htl2.12039.
26. M. E. Febrian, F. X. Ferdinan, G. P. Sendani, K. M. Suryanigrum, and R. Yunanda, "Diabetes prediction using supervised machine learning," *Procedia Comput. Sci.*, 2023; 216: 21–30.
27. F. A. Khaleel and A. M. Al-Bakry, "Diagnosis of diabetes using machine learning algorithms," *Mater. Today Proc.*, 2023; 80: 3200–3203.
28. G. Tripathi and R. Kumar, "Early prediction of diabetes mellitus using machine learning," in *2020 8th international conference on reliability, Infocom technologies and optimization (trends and future directions)(ICRITO)*, IEEE, 2020; 1009–1014. Accessed: Dec. 29, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9197832
29. B. Rathi and F. Madeira, "Early Prediction of Diabetes Using Machine Learning Techniques," in *2023 Global Conference on Wireless and Optical Technologies (GCWOT)*, IEEE, 2023; 2023; 1–7. Accessed: Dec. 29. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10064682/
30. P. J. Sireesha, M. K. Prakash, and D. Sumathi, "Implementing a Model to Detect Diabetes Prediction using Machine Learning Classifiers," *J. Algebr. Stat.*, 2022; 13(1): 558–566.
31. A. L. Fregoso-Aparicio, J. Noguez, L. Montesinos, and J.A. García-García, "Machine learning and deep learning predictive models for type 2 diabetes: a systematic review," *Diabetol. Metab. Syndr.*, Dec. 2021; 13(1): 148, doi: 10.1186/s13098-021-00767-9.
32. A. S. Alenizi and K. A. Al-karawi, "Machine Learning Approach for Diabetes Prediction," in *Proceedings of Eighth International Congress on Information and Communication Technology*, 695, X.-S. Yang, R. S. Sherratt, N. Dey, and A. Joshi, Eds., in Lecture Notes in Networks and Systems, 695., Singapore: Springer Nature Singapore, 2024; 745–756. doi: 10.1007/978-981-99-3043-2_61.
33. J. Thomas, A. Joseph, I. Johnson, and J. Thomas, "Machine learning approach for diabetes prediction," *Int. J. Inf.*, 2019; 8(2), Accessed: Dec. 29, 2023. [Online]. Available: https://www.academia.edu/download/59274154/ijiscs13822019201905 16-89234-13ra6ai.pdf
34. A. Tuppad and S. D. Patil, "Machine learning for diabetes clinical decision support: a review," *Adv. Comput. Intell.*, Apr. 2022; 2(2): 22, doi: 10.1007/s43674-022-00034-y.
35. C.-Y. Chou, D.-Y. Hsu, and C.-H. Chou, "Predicting the onset of diabetes with machine learning methods," *J. Pers. Med.*, 2023; 13(3): 406.
36. A. Bora *et al.*, "Predicting the risk of developing diabetic retinopathy using deep learning," *Lancet Digit. Health*, 2021; 3(1): e10–e19.
37. M. Panda, D. P. Mishra, S. M. Patro, and S. R. Salkuti, "Prediction of diabetes disease using machine learning algorithms," *IAES Int. J. Artif. Intell.*, 2022; 11(1): 284.
38. H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Appl. Comput. Inform.*, 2022; 18(1/2): 90–100.
39. R. K. Nadesh and K. Arivuselvan, "Type 2: diabetes mellitus prediction using deep neural networks classifier," *Int. J. Cogn. Comput. Eng.*, 2020; 1: 55– 61.