International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Predicting Diabetes Mellitus in Healthcare: A Comparative Analysis of Machine Learning Algorithms

Suhag Pandya*

Independent Researcher, India

Received 26 Nov 2023, Accepted 05 Dec 2023, Available online 15 Dec 2023, **Vol.13, No.6 (Nov/Dec 2023)**

Abstract

Hyperglycemia is the underlying cause of diabetes, a long-term health condition. This condition may be diagnosed using a battery of chemical and physical testing. Conversely, the eyes, heart, kidneys, and nerves may all suffer damage or even death from undetected and untreated diabetes. Accordingly, the mortality rate may be decreased by the analysis and early identification of diabetes. The efficacy of using ML and DL models for early illness identification has been discovered in various medical domains lately. This study explores a prediction of Diabetes Mellitus in healthcare using the Pima Indians Diabetes dataset (PIDD), comprising 768 instances and 9 attributes. o address class imbalance, the ADASYN technique generates synthetic data for minority classes. F1-Score, precision, recall, and accuracy are the metrics used to train and assess ML models like LR, RF, and KNN. Outcomes show that LR outperforms RF and KNN, achieving the highest accuracy (92.26%), precision (82%), recall (91%), and F1-Score (86%), demonstrating its effectiveness in diabetes prediction and highlighting its potential in improving healthcare decision-making. Future work in diabetes prediction can focus on several key areas to further enhance model performance and applicability in healthcare.

Keywords: *Diabetes prediction, healthcare, diagnosis, Type-1, Type-2, PIMA dataset, Diabetes mellitus, machine learning*

1. Introduction

Massive amounts of data are produced by the healthcare sector from a variety of sources, such as medical imaging, diagnostic tests, treatment histories, and patient records. This wealth of data holds significant potential to enhance patient care and optimize healthcare systems [1]. However, the true value of healthcare data can only be realized through effective processing, analysis, and interpretation[2]. With the growing complexity and volume of healthcare data, traditional methods of diagnosis and decision-making are becoming increasingly inadequate. In this context, AI and ML have emerged as transformative technologies, enabling healthcare professionals to leverage data for enhanced decision-making, personalized treatment plans, and predictive analytics[3] [4]. These technologies allow for more accurate, timely, and efficient diagnosis, enhancing patient outcomes and decreasing healthcare costs.

The chronic metabolic ailment known as diabetes mellitus (DM) impacts millions of people throughout the globe and is one of the most pressing health issues today [5].

*Corresponding author's ORCID ID: 0000-0000-0000-0000
DOI: https://doi.org/10.14741/ijcet/v.13.6.6

High blood sugar is a defining feature of diabetes, which can be brought on by either insufficient insulin production or ineffective bodily use of the hormone. Type-1,Type-2, pre-diabetes, and gestational diabetes are among the several forms of the disease, and each has unique underlying causes and processes [6]. Type 1 diabetes is an autoimmune disorder that develops when the immune system mistakenly attacks the insulin-producing pancreatic cells [7]. The inverse is correct in the case of type-2 diabetes, a metabolic disorder defined by insulin resistance and subsequently elevated blood glucose levels. The risk of developing Type 2 diabetes is increased in women who have gestational diabetes, a form of the disease that typically disappears after delivering birth. Pre-diabetes are higher blood sugar levels that do not yet fulfil the criteria for Type-2 diabetes[8].

A global prevalence of diabetes is alarming, with the WHO predicting that by 2040, approximately 600 million people worldwide will be affected by the disease [9]. A number of consequences, including cardiovascular disease, renal failure, stroke, and nerve damage, are associated with diabetes, making it a major national health concern. The growing epidemic of diabetes highlights the critical need of identifying and treating the illness at an early stage, as well as intervening promptly to lessen the impact on both people and healthcare systems [10].

Early diagnosis of diabetes is essential to prevent complications and improve long-term health outcomes. Traditionally, diabetes diagnosis has relied on blood tests, including fasting blood glucose, oral glucose tolerance tests (OGTT), and hemoglobin A1c levels. While these methods are effective, it often require time-consuming and invasive procedures, which can delay diagnosis and treatment [11]. Also, these screenings can miss those at risk for diabetes in their early stages, when treatments have the best chance of success [12]. A greater demand for more reliable, practical, and easily available means of early diagnosis of diabetes is becoming acute as the disease's impact on society grows [13].

AI and machine learning have shown great promise in addressing this gap by providing innovative solutions for predicting and diagnosing diabetes[14][15]. A branch of AI, ML [16], entails programming computers to automatically learn from data, spot patterns, and provide predictions[17]. In healthcare, ML models can analyze large datasets, such as EHR, genetic information, and medical imaging, to identify risk factors and predict disease outcomes[18]. Additionally, unlike more conventional statistical approaches, these models are able to spot intricate correlations between variables.

Significance and Contribution paper

Diabetes Mellitus is a chronic disorder that has far-reaching consequences for world health, and this research has the ability to improve our ability to forecast its occurrence. Early diagnosis of diabetes can lead to more effective management and prevention of complications, which is crucial in reducing the healthcare burden. This study contributes to the field of healthcare analytics by offering a comparative analysis of various ML algorithms. The following contribution of study are as:

- Uses the Pima Indians Diabetes dataset, providing a robust basis for diabetes prediction.
- Utilizes ADASYN to generate synthetic data for the minority class, addressing class imbalance and enhancing the model's ability.
- Applies Min-Max Scaling to rescale features, ensuring model performance by preventing skewed learning due to varying feature scales.
- Evaluates several ML algorithms, such as RF, KNN, and LR, to find the best model for diabetes predicted outcomes.
- Evaluates a performance of a models like accuracy, precision, recall, and f1-score.

Structure of paper

How the rest of the paper is structured is as follows. **Section II** delves into an examination of the healthcare industry's diabetic mellitus prediction services. This strategy is laid out thoroughly in **Section III**. Comparing and contrasting the analysis, debate, and findings is done in **Section IV**. In **Section V**, the results and potential avenues for further research are laid forth in detail.

Literature Review

In recent years, researchers have shown a growing interest in the development of Predicting Diabetes Mellitus in Healthcare. Some background studies are provided below:

This study, Alzoubi and Harous, (2022), seeks to delve into the many cutting-edge algorithms that scientists have used to forecast the onset of diabetes at an early stage. In order to understand the present limits of the work and make further improvements, the study focusses on identifying distinct approaches utilised in the literature and the efficacy of those strategies. Consequently, this study demonstrated that the RF and KNN algorithms achieved a 98% accuracy rate in early diabetes prediction, which was higher than other algorithms found in the literature [19].

This development aims Jaiswal and Gupta, (2022) so that individuals may save time and money by presenting a model that can rapidly detect diabetes [20]. The PIDD is one of a most popular options for using while developing the model. Pre-processing the dataset is necessary for removing outliers and filling in missing values. For the purpose of assessment, the ensemble learning methods XGBoost, LightGBM, and CAT Boost are used. This study employed the F1 measure, sensitivity, and correct classification rate to evaluate the outcomes. Compared to modern approaches, the LightGBM model outperforms them with a 96% accuracy rate and a MSE of just 0.04 [21].

In this research Afzal *et al.*, (2022) the dataset presented at WiDS 2021 Datathone has been used to construct two efficient deep learning models: XGBOOST and LGBM. To get a better understanding of linked characteristics, a thorough feature engineering procedure was carried out before deep learning techniques were used. Important participative elements in forecasting anomalous findings were brought to light, including filling missing numbers, class imbalance, and age groupings. The entire dataset was sent to ML algorithms after data segregation based on gender, age, ethnicity, and max glucose did not provide any noticeable variances among the classes. Lastly, the ROC assessment technique was used to evaluate the model, and the accuracy was 0.87 [22].

This research, Awoniran *et al.* (2022) used data science methods to enhance the precision of diabetes mellitus prediction using a dataset. This was accomplished by using principle components analysis to minimise dimensionality and pre-processing the data using dummy categories. The system was further trained using DNNs, a RFC, and a SVM. DNNs achieved an accuracy of 0.89, a SVM an accuracy of 0.76, and a RFC an accuracy of 0.77. Similarly, the most accurate results were produced by deep neural networks. Researchers found that ML algorithms can better anticipate the onset of diabetes mellitus with enhanced pre-processing [23].

This article, Pal, Parija and Panda, (2021), put forth a model for diabetic illness prediction that is based on ML. The three supervised ML algorithms K-NN, LSVM, and RF are responsible for diabetes prediction in order to facilitate early detection. In order to find the AUC and accuracy for these models, they used the PIMA Indian Diabetes dataset that is available in the UCI repository. Results demonstrate that random forest has the best accuracy (78.57) and AUC (95.08) among the three algorithms tested for diabetes risk prediction. The contribution of this article will allow the medical community to improve disease prediction and treatment efficacy. The proposed approach may potentially be able to identify other disorders [24].

This study Estonilo and Festijo, (2021) to build a deep learning-based diabetic mellitus prediction tool for mobile devices. The TensorFlow framework's Sequential function was utilised to construct the diabetes prediction model. After that, the model was converted to a 'tflite' format and used to build a mobile app using the Android Studio IDE that can detect whether someone has diabetes mellitus. A remarkable 93% accuracy was shown by the deep learning model. Furthermore, the program supplies users with crucial directions as well as information about diabetes mellitus. A ground-breaking new tool for the early diagnosis of diabetes mellitus, the created mobile app is based on deep learning. If the prognosis is good, a change in lifestyle may be possible, and a major problem may be averted [25].

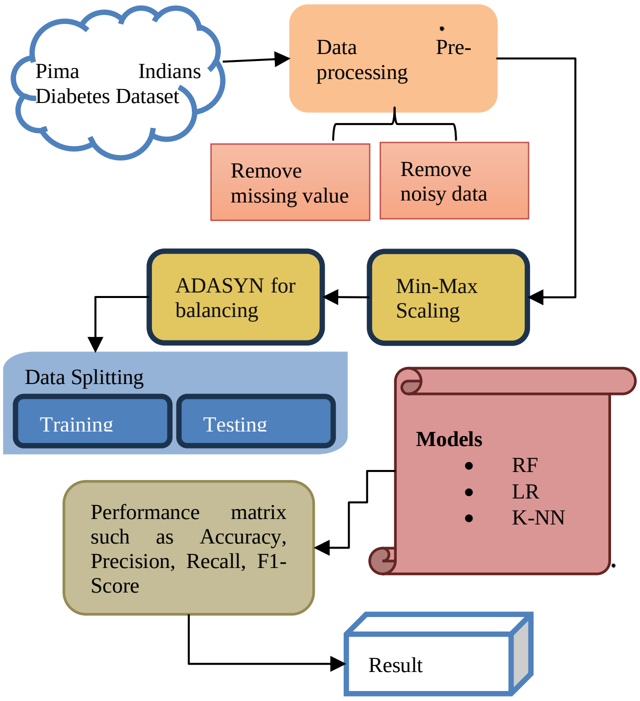
Recent studies on diabetes prediction using machine learning consistently show high accuracy and F1 scores that shows in Table I. Limitations across the studies include the need for more detailed data specifications and broader dataset validation for improved generalizability and model robustness.

Table 1 Comparison of Machine Learning Algorithms for Diabetes Prediction

Author	Algorithm	Data	Findings	Limitations/Research Gaps
Alzoubi and Harous (2022)	RF, KNN, SVM, MLP	Healthcare dataset	RF and KNN outperformed other algorithms with 98% accuracy in early diabetes prediction.	Dataset limitations; needs broader testing for general applicability.
Jaiswal and Gupta (2022)	CAT Boost, LightGBM, XGBoost	Pima Indian Diabetes Dataset	LightGBM model achieved 96% accuracy and 0.04 MSE, outperforming other methods.	The model's generalizability to other datasets needs further exploration.
Afzal et al. (2022)	XGBoost, LightGBM	WiDS 2021 Datathon dataset	XGBoost and LightGBM achieved 0.87 accuracy, with feature engineering improving prediction quality.	The impact of specific factors like age, gender, and ethnicity on predictions was not clearly shown.
Awoniran et al. (2022)	SVM, RF, DNN	Diabetes Mellitus Dataset	Deep Neural Networks yielded the highest accuracy of 0.89. Better pre-processing improved accuracy of predictions.	The study did not explore the impact of real-time data or other advanced preprocessing techniques.
Pal, Parija, and Panda (2021)	K-NN, Linear SVM, RF	Pima Indian Diabetes dataset (UCI)	Random Forest showed the best results with 78.57% accuracy and 95.08% AUC.	Could expand to include more diverse datasets for validation of findings.
Estonilo and Festijo (2021)	Deep Learning (TensorFlow)	Healthcare dataset	The deep learning model achieved 93% accuracy, and a mobile app was developed for diabetes prediction.	The study did not discuss the scalability or limitations of the mobile app in varied real-world scenarios.

Research Methodology

The methodology for predicting Diabetes Mellitus in healthcare involves several key stages, starting with the application of the Pima Indians Diabetes dataset, which includes 768 instances and 9 attributes. The following steps and phases of research design are provided in Figure 1 flowchart, diabetes prediction. Data normalisation utilising Min-Max Scaling and the removal of missing values and noisy data are examples of pre-processing processes that improve data quality. Making synthetic data for under-represented classes using the ADASYN method helps the model train better and addresses class imbalance. The dataset is split into training (80%) and testing (20%) subsets. ML models, like RF, KNN, and LR, are trained on the training data. The model performance is evaluated using accuracy metrics, with the confusion matrix providing insights into prediction accuracy and model evaluation. This method improves healthcare decision-making by identifying the best model for diabetes prediction.

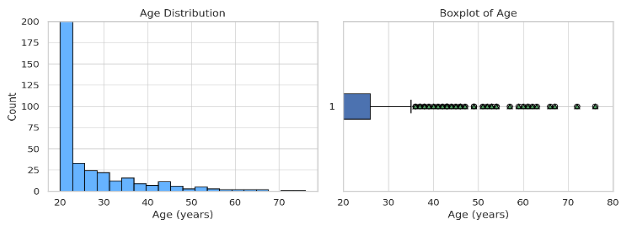


Flowchart for Predicting Diabetes Mellitus in Healthcare

The following is a flow diagram of Figure 1, with short descriptions of each step:

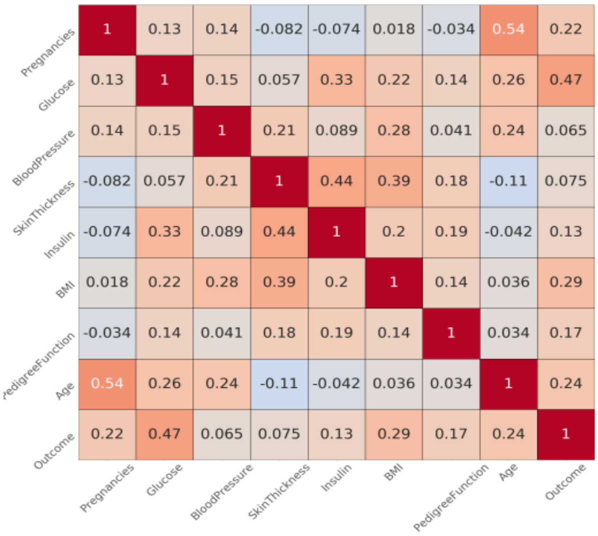
Data Collection

This study made use of the Pima Indians Diabetes Dataset, which is widely utilised and very popular. Discretionary attributes are part of the dataset, which comprises 768 occurrences over 9 columns. The dataset, which was gathered at Pima Indian Heritage, includes information on several patients, all of whom are female and range in age from 21 to 81. Below is a visualisation of the data analysis:



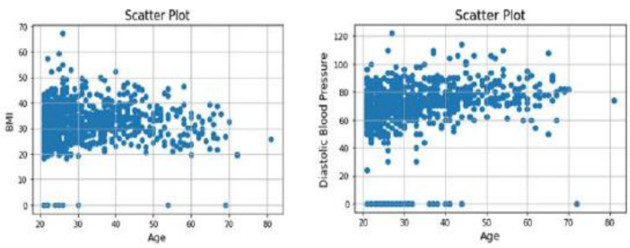
Univariate Data Analysis of Age

The age distribution, as shown in Figure 2, is right-skewed with a long tail extending towards older ages, indicating that the majority of individuals fall into younger age groups while fewer are represented in older age groups. The boxplot further summarizes this distribution, where the interquartile range (IQR) contains 50% of the data, and the median age is approximately 35 years. The relatively small IQR suggests that most individuals are concentrated within a narrow age range. However, several outliers are observed in the older age groups, as represented by dots beyond the whiskers.



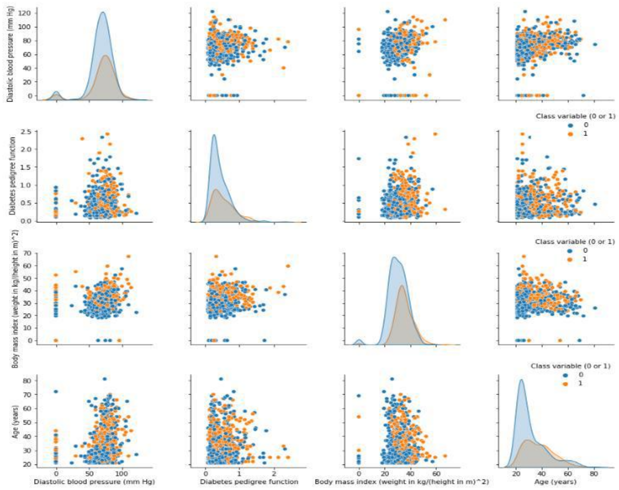
Correlation matrix of PIMA Dataset

he correlation matrix in figure 3 reveals that glucose and BMI have strong positive correlations with diabetes risk, confirming their importance as predictors. Age shows a moderate positive correlation, indicating a slightly higher risk with older age. Additionally, glucose and BMI are strongly correlated with each other, while skin thickness and insulin have weaker associations with the outcome, suggesting their predictive power is limited.



Univariate Analysis of BMI vs Age

Figure 4 present BMI versus age reveals no clear linear relationship between the two variables, as the points are widely scattered across the graph without any discernible pattern. Each dot represents an individual's BMI and age, and while the majority of values are clustered within a typical range, there are a few individuals with very high BMI values, suggesting the presence of outliers. This indicates that BMI is not strongly associated with age in this dataset, and other factors may play a more significant role in determining BMI.



Multivariate Data Analysis

Figure 5 provides insights into the relationships between features in the PIMA dataset, highlighting both the distribution and correlations. Features like "DiastolicBloodPressure" and "BMI" appear normally distributed, while "Diabetes Pedigree Function" is skewed. Positive correlations are observed between "DiastolicBloodPressure" and "BMI," as well as between "Glucose" and "Body Mass Index." A weak negative correlation exists among "Age" and "Diabetes Pedigree Function." The plot also shows better class separation for features like "Glucose" and "Body Mass Index," indicating their potential as strong predictors for diabetes.

Data Preprocessing

The initial stage of all data analysis is data preparation. Data preparation may be done in a myriad of ways. Pre-processing is the first step in creating machine-learning models from unstructured data by making the data more understandable and standardised [26] [27]. To improve the quality of the input data, pre-processing mainly aims to decrease the amount of noise, redundant data, and unnecessary data [28]. Here are some important words related to pre-processing:

Remove missing value: The frequent occurrence of missing values in medical data has emerged as a major concern impacting the accuracy of categorisation results [29].

Remove noisy data: Errors, outliers, missing numbers, or inconsistencies are examples of noisy data that might compromise the accuracy and dependability of analysis [30].

Min-Max Scaling

The features were rescaled from 0 to 1 using the MinMaxScaler method [31]. This method excels at handling outliers because it employs statistical procedures that have no impact on the data's variance (Equation (1)).

$$x' = \frac{x - \min(x)}{\max(x)} - \min(x) \quad (1)$$

Equation (1) shows that although x is an original value, x' is a scaled value [32], a feature value's upper limit is denoted by max, and its lower bound by min[33][34]. MinMaxScaler scaling reduces time by maintaining the sparsity of the input data, even for data with a large number of zero entries.

Data balancing with ADASYN

ADASYN can create data samples for difficult-to-learn minority classes. In addition to reducing the learning bias in the real dataset, the data points created via ADASYN [35] also help to balance the dataset[36][37]. Furthermore, minority samples are unevenly dispersed, with just one example included in each neighbourhood (2).

$$s_i = (x_i + x_{zi} - x_i)\lambda \quad (2)$$

Data Splitting

The dataset was split into two parts for this study: training data and testing data. In this research, the training to testing ratio is 80:20, meaning that the models are trained using 80% of the data and tested using 20%.

Logistic Regression (LR)

Binary classification is one application of a machine learning method known as Logistic Regression. The input is given a linear mixture of qualities, and a sigmoid function is applied to simulate the likelihood that the input belongs to a specific class [38][39]. As it trains, it finds the values of the parameters that minimise the log-loss function. Whether your input characteristics are numerical or categorical, LR can handle them all with ease and efficiency [40][41]. An example of a binary classification problem is when there are only two possible answers, like "yes" or "no." One of the most common ML algorithms for this type of problem is logistic regression, which can be used to predict whether a person has diabetes or not [42][43]. Logistic regression is a method for calculating expectations or classifying data using the probability of a categorical dependent variable [44]. Imagine a dataset containing n characteristics (Age, BMI,) and m instances of input parameters (X) that are needed to create a prediction. Then, think of Y as a matrix of values. As for Y, it's a vector containing m instances that represent the sample results from X [45]. The goal is to teach the LRML model to identify the class to which the upcoming values belong.

$$Y = b_0 + b_1 * X \quad (3)$$

When the sigmoid function is introduced to Eq. 1 for Linear Regression, the result is eq.4.

$$P = \frac{1}{1+e^{-y}} \quad (4)$$

$$Ln\left(\frac{p}{1-p}\right) = b_0 + b_1 * X \quad (5)$$

Finally, to obtain the logistic regression function, substitute the classification dependent variable Y with the aforementioned probability sigmoid function and plug it into equation 1.

Evaluation metrics

It is crucial to evaluate the ML model in order to assess the algorithm's capabilities. To do this, one may determine the ML model's accuracy by using the confusion matrix. Rounded to the nearest whole number, it represents the ratio of accurate forecasts to total predictions. The 2x2 matrix that summarises the classification algorithm's prediction outcomes is called the Accuracy = Confusion matrix.

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

Classes of Confusion Metrics

Here class1: positive and class2: negative Definition of terms: If an observation is positive and predicted to be positive, it is marked as a true positive (TP); False negatives (FNs) occur when positive observations are expected to be negative. while an observation turns out to be negative while a positive result was expected, it's called a false positive (FP). When an observation is both predicted to be negative and negative in reality, we say that it is a true negative (TN).

Accuracy: It measures how many predictions were accurate relative to a total number of samples used for training. This is the formula: (6)-

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \quad (6)$$

Precision: The precision rate is the percentage of predicted outcomes that were actually categorised correctly. It is expressed as (7)-

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

Recall: It is computed as the number of correct positive results divided by the total number of relevant samples. A formula for it in mathematics is (8)-

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

F1-Score: For a classifier to be considered precise, its anticipated number of positive outcomes must be divided by the actual number of positive results. It is calculated by Equation (9):

$$F1 - Score = \frac{2*Precision*Recall}{(Precision+Recall)} (9)$$

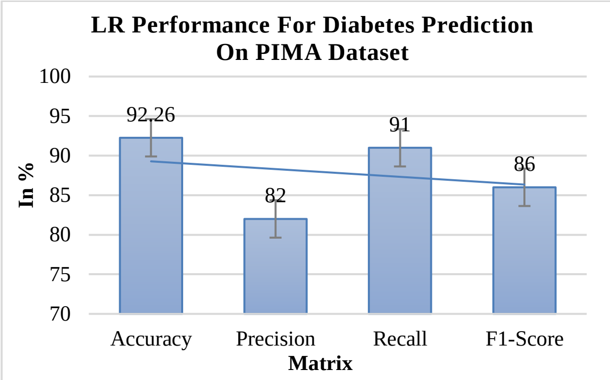
These measures, when taken as a whole, show how well the model predicts the important variable.

Results And Discussion

The models' experimental outcomes are detailed in this section. The following outcomes are evaluated using f1-score, recall, accuracy, and precision metrics. Accuracy graphs, confusion matrices, loss matrices, and other visual statistics showing the LR model's performance are shown in Table II. The following LR models are compare (see in table III) with existing models like RF[46], and KNN [47] with the same parameter on the PIMA dataset.

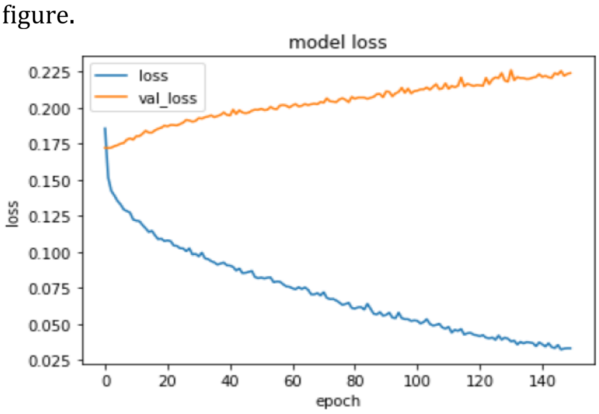
Table 2 Results of LR model efficiency for diabetes Mellitus prediction

Measures	Logistic Regression (LR)
Accuracy	92.26
Precision	82
Recall	91
F1-Score	86



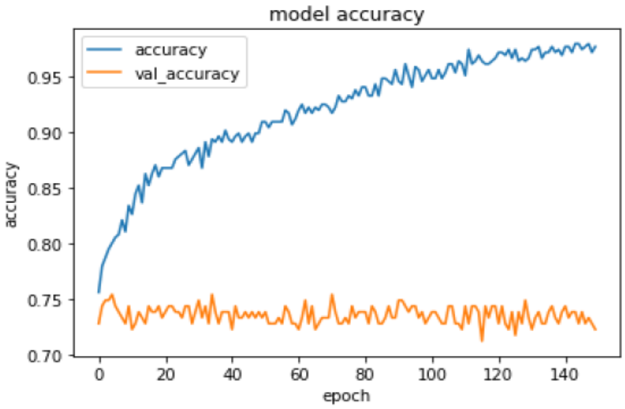
Bar graph for LR model performance

Figure 7 displays the results of the LR model for healthcare diabetes mellitus prediction. Impressive overall performance and accurate categorisation of positive examples are shown by LR's 92.26% accuracy, 82% precision, 91% recall, and 86% fi-score in this figure.



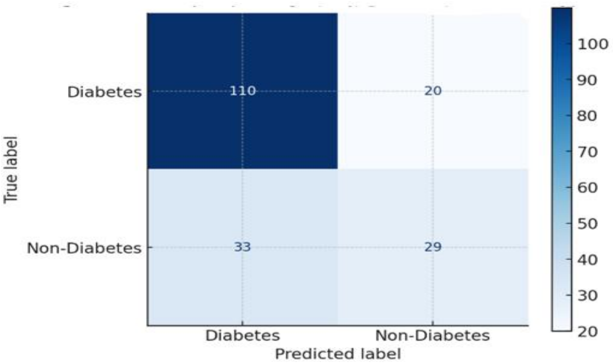
Training and validation loss graph for LR

Blue represents training loss and orange represents validation loss; both lines show the LR's performance across 140 epochs in figure 8. The fact that they begin high and then drop shows that the model is learning. It seems that the model is becoming better without overfitting since the training loss drops at a little quicker rate, but eventually both lines flatten.



Training and validation accuracy graph for LR

Accuracy in training and validation are shown by the blue and orange lines, respectively, in Figure 9, which covers 140 epochs of an LR. Over a course of the training and validation epochs, both lines rise, signifying better performance; this indicates that the model is learning well from both datasets.



Confusion matrix for LR

Figure 10 displays a confusion matrix that assesses how well a diabetes prediction classification model performs. It consists of four key values: 110 TP (Diabetes correctly predicted as Diabetes), 20 FN (Diabetes incorrectly predicted as non-diabetes), 33 FP (Non-Diabetes incorrectly predicted as Diabetes), and 29 TN (Non-Diabetes correctly predicted as non-diabetes). The matrix shows how well the model can identify diabetes patients and where it has made mistakes. The accuracy and precision of a model may be enhanced if the number of FP and FN is relatively large.

Table 3 Comparison between LR and existing model performance on PIMA dataset

Model	Accuracy	Precision	Recall	F1-Score
RF	87.1	80.6	85.4	83
KNN	75.7	79	88	83
LR	92.26	82	91	86

Table III compares the performance of Logistic Regression (LR) with existing models, RF and KNN, on the PIMA dataset. LR demonstrates the highest accuracy at 92.26%, outperforming RF at 87.1% and KNN at 75.7%. Additionally, LR achieves a precision of 82% and a recall of 91%, striking a balance between predicting positive cases accurately and minimizing false negatives. In comparison, RF achieves a precision of 80.6% and a recall of 85.4%, while KNN shows a higher recall of 88% but a comparatively lower precision of 79%. LR also yields the highest F1-Score of 86, surpassing RF 83 and KNN 83, showcasing its superior overall performance and effectiveness for diabetes prediction on the PIMA dataset.

References

[1] B. Patel, V. K. Yarlagadda, N. Dhameliya, K. Mullangi, and S. C. R. Vennapusa, “Advancements in 5G Technology: Enhancing Connectivity and Performance in Communication Engineering,” Eng. Int., vol. 10, no. 2, pp. 117–130, 2022, doi: 10.18034/ei.v10i2.715.

[2] M. S. Arora, Rajeev, Sheetal Gera, “Impact of Cloud Computing Services and Application in Healthcare Sector and to provide improved quality patient care,” no. October 2021, 2021.

[3] K. V. V. Sandeep Gupta Jubin Thomas , Piyush Patidar, “An Analysis of Predictive Maintenance Strategies in Supply Chain Management,” Int. J. Sci. Res. Arch., vol. 6, no. 01, pp. 308–317, 2022.

[4] S. R. Bauskar, “Evaluation of deep learning for the diagnosis of leukemia blood cancer,” Int. J. Adv. Res. Eng. Technol., vol. 11, no. 3, pp. 661–672, 2020.

[5] S. Bauskar, “An Analysis: Early Diagnosis and Classification of Parkinson’s Disease Using Machine Learning Techniques,” Int. J. Comput. Eng. Technol., pp. 54–66, 2021, doi: 10.5281/zenodo.13836264.

[6] Rajesh Goyal, “The role of business analysts in information management projects,” Int. J. Core Eng. Manag., vol. 6, no. 9, 2020.

[7] V. K. Yarlagadda, S. S. Maddula, D. K. Sachani, and ..., “Unlocking Business Insights with XBRL: Leveraging Digital Tools for Financial Transparency and Efficiency,” Asian Account. ..., no. December, 2020, [Online]. Available: https://www.researchgate.net/profile/Vamsi-Krishna-Yarlagadda/publication/381922694_Unlocking_Business_Insights_with_XBRL_Leveraging_Digital_Tools_for_Financial_Transparency_and_Efficiency/links/6684c2270a25e27fbc1fbd14/Unlocking-Business-Insights-with-XB

[8] U. A. Zia and N. Khan, “Predicting Diabetes in Medical Datasets Using Machine Learning Techniques,” Int. J. Sci. Res. Eng. Trends, 2019.

[9] S. G. Jubin Thomas, Kirti Vinod Vedi, “The Effect and Challenges of the Internet of Things (IoT) on the Management of Supply Chains,” Int. J. Res. Anal. Rev, vol. 8, no. 3, pp. 874–878, 2021.

[10] P. Prabhu and S. Selvabharathi, “Deep Belief Neural Network Model for Prediction of Diabetes Mellitus,” in 2019 3rd International Conference on Imaging, Signal Processing and Communication, ICISPC 2019, 2019. doi: 10.1109/ICISPC.2019.8935838.

[11] R. Goyal, “The role of requirement gathering in agile software development: strategies for success and challenges,” Int. J. Core Eng. Manag., vol. 6, no. 12, pp. 142–152, 2021.

[12] M. R. Kishore Mullangi, Vamsi Krishna Yarlagadda, Niravkumar Dhameliya, “Integrating AI and Reciprocal Symmetry in Financial Management: A Pathway to Enhanced Decision-Making,” Int. J. Reciprocal Symmetry Theor. Phys., vol. 5, no. 1, pp. 42–52, 2018.

[13] R. R., Y. Y., R. M., and N. S., "Comparative evaluation of IADPSG criteria with ADA and WHO criteria for diagnosis of gestational diabetes mellitus," Journal, Indian Acad. Clin. Med., 2015.
[14] R. A. Anoop Kumar, Ramakrishna Garine, Arpita Soni, "Leveraging AI for E-Commerce Personalization : Insights and Challenges from 2020," pp. 1–6, 2020.
[15] S. Arora and A. Tewari, "AI-Driven Resilience : Enhancing Critical Infrastructure with Edge Computing," Int. J. Curr. Eng. Technol., vol. 12, no. 2, pp. 151–157, 2022, doi: https://doi.org/10.14741/ijcet/v.12.2.9.
[16] R. Bishukarma, "The Role of AI in Automated Testing and Monitoring in SaaS Environments," Int. J. Res. Anal. Rev., vol. 8, no. 2, pp. 846–851, 2021.
[17] S. K. R. Anumandla, V. K. Yarlagadda, S. C. R. Vennapusa, and K. R. V Kothapalli, "Unveiling the Influence of Artificial Intelligence on Resource Management and Sustainable Development: A Comprehensive Investigation," Technol. \& Manag. Rev., vol. 5, no. 1, pp. 45–65, 2020.
[18] J. Bradley and S. Rajendran, "Developing predictive models for early detection of intervertebral disc degeneration risk," Healthc. Anal., 2022, doi: 10.1016/j.health.2022.100054.
[19] R. Alzoubi and S. Harous, "Machine Learning Algorithms for Early Prediction of Diabetes: A Mini-Review," in 2022 International Conference on Electrical and Computing Technologies and Applications, ICECTA 2022, 2022. doi: 10.1109/ICECTA57148.2022.9990240.
[20] S. S. Pranav Khare, "The Impact of AI on Product Management : A Systematic Review and Future Trends," Int. J. Res. Anal. Rev. (IJRAR, vol. 9, no. 4, 2022.
[21] S. Jaiswal and P. Gupta, "Ensemble Approach: XGBoost, CATBoost, and LightGBM for Diabetes Mellitus Risk Prediction," in 2022 2nd International Conference on Computer Science, Engineering and Applications, ICCSEA 2022, 2022. doi: 10.1109/ICCSEA54677.2022.9936130.
[22] E. Afzal, T. Saba, K. Javed, H. Ali, and A. Karim, "Diagnosis and Prognosis of Diabetes Mellitus with Deep Learning," in Proceedings - 2022 5th International Conference of Women in Data Science at Prince Sultan University, WiDS-PSU 2022, 2022. doi: 10.1109/WiDS-PSU54548.2022.00031.
[23] O. M. Awoniran, M. O. Oyelami, R. N. Ikono, R. F. Famutimi, and T. I. Famutimi, "A Machine Learning Technique for Detection of Diabetes Mellitus," in Proceedings of the 5th International Conference on Information Technology for Education and Development: Changing the Narratives Through Building a Secure Society with Disruptive Technologies, ITED 2022, 2022. doi: 10.1109/ITED56637.2022.10051439.
[24] M. Pal, S. Parija, and G. Panda, "Improved prediction of diabetes mellitus using machine learning based approach," in 2nd International Conference on Range Technology, ICORT 2021, 2021. doi: 10.1109/ICORT52730.2021.9581774.
[25] C. G. Estonilo and E. D. Festijo, "Development of Deep Learning-Based Mobile Application for Predicting Diabetes Mellitus," in Proceedings - 2021 4th International Conference on Computer and Informatics Engineering: IT-Based Digital Industrial Innovation for the Welfare of Society, IC2IE 2021, 2021. doi: 10.1109/IC2IE53219.2021.9649235.
[26] S. Vijayarani, M. . Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining-An Overview Privacy Preserving Data Mining View project," Int. J. Comput. Sci. Commun. Networks, 2015.
[27] N. Abid, "A Climbing Artificial Intelligence for Threat Identification in Critical Infrastructure Cyber Security," Int. J. Res. Anal. Rev., vol. 9, no. 4, 2022.
[28] K. Patel, "Quality Assurance In The Age Of Data Analytics: Innovations And Challenges," Int. J. Creat. Res. Thoughts, vol. 9, no. 12, pp. f573–f578, 2021.
[29] Vasudhar Sai Thokala, "Efficient Data Modeling and Storage Solutions with SQL and NoSQL Databases in Web Applications," Int. J. Adv. Res. Sci. Commun. Technol., pp. 470–482, Apr. 2022, doi: 10.48175/IJARSCT-3861B.
[30] M. S. Rajeev Arora, "Applications of Cloud Based ERP Application and how to address Security and Data Privacy Issues in Cloud application," 2022.
[31] Mani Gopalsamy, "An Optimal Artificial Intelligence (AI) technique for cybersecurity threat detection in IoT Networks," Int. J. Sci. Res. Arch., vol. 7, no. 2, pp. 661–671, Dec. 2022, doi: 10.30574/ijsra.2022.7.2.0235.
[32] V. S. Thokala, "A Comparative Study of Data Integrity and Redundancy in Distributed Databases for Web Applications," Int. J. Res. Anal. Rev., vol. 8, no. 4, pp. 383–389, 2021.
[33] M. Z. Hasan, R. Fink, M. R. Suyambu, and M. K. Baskaran, "Assessment and improvement of intelligent controllers for elevator energy efficiency," in IEEE International Conference on Electro Information Technology, 2012. doi: 10.1109/EIT.2012.6220727.
[34] V. V. Kumar, A. Sahoo, and F. W. Liou, "Cyber-enabled Product Lifecycle Management: A Multi-agent Framework," Procedia Manuf., vol. 39, pp. 123–131, 2019, doi: 10.1016/j.promfg.2020.01.247.
[35] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in Proceedings of the International Joint Conference on Neural Networks, 2008. doi: 10.1109/IJCNN.2008.4633969.
[36] M. Gopalsamy, "Artificial Intelligence (AI) Based Internet-ofThings (IoT)-Botnet Attacks Identification Techniques to Enhance Cyber security," Int. J. Res. Anal. Rev., vol. 7, no. 4, pp. 414–420, 2020.
[37] M. Z. Hasan, R. Fink, M. R. Suyambu, M. K. Baskaran, D. James, and J. Gamboa, "Performance evaluation of energy efficient intelligent elevator controllers," in IEEE International Conference on Electro Information Technology, 2015. doi: 10.1109/EIT.2015.7293320.
[38] V. V. Kumar, F. W. Liou, S. N. Balakrishnan, and V. Kumar, "Economical impact of RFID implementation in remanufacturing: a Chaos-based Interactive Artificial Bee Colony approach," J. Intell. Manuf., vol. 26, no. 4, pp. 815–830, Aug. 2015, doi: 10.1007/s10845-013-0836-9.
[39] S. Pandya, "Predictive Analytics in Smart Grids : Leveraging Machine Learning for Renewable Energy Sources," Int. J. Curr. Eng. Technol., vol. 11, no. 6, pp. 677–683, 2021, doi: https://doi.org/10.14741/ijcet/v.11.6.12.
[40] A. and P. Khare, "Cloud Security Challenges : Implementing Best Practices for Secure SaaS Application Development," Int. J. Curr. Eng. Technol., vol. 11, no. 6, pp. 669–676, 2021, doi: https://doi.org/10.14741/ijcet/v.11.6.11.
[41] N. G. Singh, Abhinav Parashar A, "Streamlining Purchase Requisitions and Orders : A Guide to Effective Goods Receipt Management," J. Emerg. Technol. Innov. Res., vol. 8, no. 5, 2021.
[42] J. Thomas and V. Vedi, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," J. Emerg. Technol. Innov. Res., vol. 8, no. 9, 2021.
[43] V. V. Kumar, S. R. Yadav, F. W. Liou, and S. N. Balakrishnan, "A Digital Interface for the Part Designers and the Fixture Designers for a Reconfigurable Assembly System," Math. Probl. Eng., vol. 2013, pp. 1–13, 2013, doi: 10.1155/2013/943702.

[44] S. A. and A. Tewari, “Security Vulnerabilities in Edge Computing: A Comprehensive Review,” Int. J. Res. Anal. Rev., vol. 9, no. 4, pp. 936–941, 2022.

[45] V. V. Kumar, F. T. S. Chan, N. Mishra, and V. Kumar, “Environmental integrated closed loop logistics model: An artificial bee colony approach,” in SCMIS 2010 - Proceedings of 2010 8th International Conference on Supply Chain Management and Information Systems: Logistics Systems and Engineering, 2010.

[46] Q. Wang, W. Cao, J. Guo, J. Ren, Y. Cheng, and D. N. Davis, “DMP_MI: An effective diabetes mellitus classification algorithm on imbalanced data with missing values,” IEEE Access, 2019, doi: 10.1109/ACCESS.2019.2929866.

[47] B. Pranto, S. M. Mehnaz, E. B. Mahid, I. M. Sadman, A. Rahman, and S. Momen, “Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh,” Inf., 2020, doi: 10.3390/INFO11080374.