Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
January-June 2024 5(1), 1739–1749

A COMPREHENSIVE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR PREDICTING DIABETES ONSET AND ITS CLINICAL APPLICATIONS

Sohail Mohammad Hussain Sayyed [1], Saniya Asif Shaikh [1], Summaiya Tamboli[1], Alfiya Aadil Bagwan[1], Shafiya Majid Sayyed[1]

[1] Department of BCA Science, Abeda Inamdar Senior College of Arts Science and commerce, Autonomous, Pune, Maharashtra, India

Corresponding Author
Sohail Mohammad Hussain Sayyed,
sohailsayyed@aiscpune.org

DOI
10.29121/shodhkosh.v5.i1.2024.4557

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

The early prediction of diabetes onset is critical in preventing the progression of the disease and improving patient outcomes. This research presents a comprehensive evaluation of various machine learning algorithms applied to the prediction of diabetes onset, with a focus on their clinical applications. A range of machine learning models, including decision trees, random forests, support vector machines, and neural networks, are assessed using a publicly available diabetes dataset. The study aims to identify the most accurate and efficient algorithms for predicting diabetes onset based on a set of medical attributes, such as age, BMI, glucose levels, and family history. The evaluation considers multiple performance metrics, including accuracy, precision, recall, and area under the curve (AUC), to assess the effectiveness of each algorithm in predicting diabetes risk. The results demonstrate that ensemble methods like random forests and gradient boosting outperformed other models, providing high accuracy and robustness in prediction. Additionally, the study discusses the practical implications of these models in clinical settings, highlighting their potential for aiding healthcare providers in early diabetes detection and personalized treatment plans. The findings contribute valuable insights into the use of machine learning for diabetes prediction, offering a foundation for future research and the development of automated decision support systems in healthcare.

Keywords: Machine Learning, Diabetes Prediction, Early Detection, Classification Algorithms, Healthcare Applications, Predictive Modeling

1. INTRODUCTION

Diabetes is a chronic medical condition that occurs when the body cannot properly process or respond to insulin, resulting in elevated blood sugar levels. It has become a significant global health challenge due to its rising prevalence and long-term health implications. According to the World Health Organization (WHO), the number of people with diabetes has quadrupled over the last few decades, with an estimated 422 million adults affected worldwide. This rapid increase in diabetes prevalence has contributed to a rise in associated complications, including cardiovascular diseases, kidney failure, blindness, and amputations, making early detection and management crucial for preventing these severe health outcomes.

The early detection and prediction of diabetes onset are fundamental to controlling the progression of the disease and minimizing its long-term complications. Identifying individuals at risk for diabetes enables timely intervention, which can significantly improve health outcomes and reduce the overall healthcare burden. Lifestyle changes such as diet, physical activity, and pharmacological treatments are most effective when initiated early, thereby preventing or delaying the onset of full-blown diabetes and its associated complications. However, traditional diagnostic methods such as blood tests can often be costly, time-consuming, and invasive, especially in the early stages when symptoms may not be present. Thus, predicting the risk of diabetes through non-invasive and cost-effective means holds substantial promise for improving public health.

Machine learning (ML) techniques have gained considerable attention in healthcare due to their ability to identify patterns in large datasets that are often not detectable by traditional statistical methods. ML has shown great potential in various aspects of healthcare, including disease diagnosis, treatment optimization, and predictive modeling. In particular, machine learning models can analyze vast amounts of patient data, including clinical, demographic, and lifestyle factors, to predict the likelihood of diabetes onset. By leveraging these models, healthcare professionals can better identify high-risk individuals and take proactive measures for early intervention.

The integration of machine learning into healthcare systems is not just an opportunity for more accurate prediction but also for more efficient and scalable disease management. ML algorithms can analyze complex datasets in real time, providing healthcare providers with predictive insights that support timely decision-making. Furthermore, these models have the potential to be adapted to different healthcare environments, making them accessible to both well-resourced hospitals and resource-limited settings. The automation of risk prediction models through machine learning can lead to significant improvements in the efficiency of healthcare systems by reducing diagnostic errors and ensuring that individuals receive personalized care based on their risk profiles.

This research aims to evaluate the effectiveness of several machine learning algorithms in predicting diabetes onset, comparing their performance in terms of accuracy, precision, recall, and other relevant metrics. A variety of machine learning models, including decision trees, random forests, support vector machines (SVM), and deep learning algorithms, will be tested on publicly available diabetes datasets. By assessing these models' ability to predict diabetes risk using different clinical features, the study seeks to identify the most accurate and reliable approaches for early diabetes detection.

Moreover, this research will also explore the potential clinical applications of these machine learning models. In particular, it will investigate how the integration of predictive algorithms can be incorporated into routine healthcare practices, such as electronic health record (EHR) systems, to provide automated diabetes risk assessments for patients. This will help healthcare providers make informed decisions and offer more targeted interventions. The study aims to not only compare the performance of different machine learning models but also evaluate their feasibility and practicality in real-world clinical settings.

The primary objectives of this research are to evaluate the performance of various machine learning algorithms in predicting diabetes onset, including decision trees, random forests, support vector machines, and neural networks. This evaluation will focus on analyzing the effectiveness of these algorithms in identifying high-risk individuals based on clinical and demographic data. Additionally, the study aims to explore the potential clinical applications of machine learning models, examining how these predictive algorithms can be integrated into routine healthcare practices to assist in early intervention and patient management. The research will also assess the strengths and weaknesses of each model in terms of predictive accuracy, computational complexity, and interpretability, providing insights into their practical use in different healthcare settings. Ultimately, the goal is to offer valuable recommendations for the optimal use of machine learning models in improving early diabetes detection and supporting personalized healthcare strategies.

2. LITERATURE SURVEY

2.1. OVERVIEW OF DIABETES PREDICTION

The prediction of diabetes onset has been a prominent area of research, as early detection can significantly reduce the risk of complications and improve patient outcomes. Many studies have explored the use of machine learning techniques to predict the risk of diabetes based on various clinical and demographic features, such as age, BMI, glucose levels, family history, and lifestyle factors. A widely used dataset in these studies is the PIMA Indian Diabetes dataset, which includes information on over 700 individuals and is often used to benchmark different machine learning algorithms for diabetes prediction. Other datasets such as the UCI Diabetes dataset and the Diabetes 130-US hospitals dataset have also been used for similar purposes.

Several studies have demonstrated the effectiveness of machine learning models in predicting diabetes risk. For example, research by Chaurasia and Pal (2018) explored the use of decision trees and random forests to predict diabetes onset and achieved high accuracy in identifying individuals at risk. Similarly, a study by Soni and Patel (2017) applied support vector machines (SVM) to the PIMA dataset and reported promising results in terms of prediction accuracy. These studies commonly use performance metrics such as accuracy, precision, recall, F1-score, and area under the curve (AUC) to evaluate the models. However, while these models have shown potential, there remains room for improvement in terms of computational efficiency and interpretability, which are crucial for clinical adoption.

2.2. MACHINE LEARNING MODELS USED

A variety of machine learning algorithms have been employed in diabetes prediction, each offering distinct advantages and limitations. Decision trees are one of the most widely used algorithms in this domain due to their simplicity, interpretability, and ability to handle both categorical and numerical data. Random forests, an ensemble method based on decision trees, have also been extensively used and tend to provide better performance by reducing overfitting and increasing robustness. Research has shown that random forests outperform decision trees in terms of predictive accuracy, particularly when the dataset contains noise or complex relationships among features.

Support vector machines (SVM) are another popular machine learning model for diabetes prediction. SVMs are well-regarded for their ability to classify data with a high-dimensional feature space, making them suitable for complex datasets. Several studies have demonstrated that SVMs can achieve high accuracy in diabetes prediction, especially when combined with kernel functions that map data into higher-dimensional spaces. However, SVMs require significant computational resources, and their interpretability can be limited compared to tree-based methods.

Neural networks, particularly deep learning models, have recently gained attention for diabetes prediction. These models can capture non-linear relationships and complex patterns in the data, making them well-suited for large, high-dimensional datasets. Research has shown that neural networks, particularly multi-layer perceptrons (MLPs) and convolutional neural networks (CNNs), can outperform traditional machine learning models in terms of accuracy. However, deep learning models often require large amounts of data and computational power, which may limit their applicability in resource-constrained clinical settings.

2.3. COMPARISON OF TECHNIQUES

When comparing machine learning algorithms for diabetes prediction, several factors come into play, including accuracy, computational complexity, and interpretability. Decision trees are highly interpretable, allowing healthcare professionals to understand the rationale behind predictions. However, they are prone to overfitting, especially when the dataset is small or noisy. Random forests mitigate this issue by averaging multiple decision trees, resulting in a more robust model that generally performs better on complex datasets.

Support vector machines offer high accuracy and are particularly effective when dealing with non-linear decision boundaries. However, SVMs are computationally expensive and may struggle with large datasets or real-time predictions. In contrast, random forests and decision trees are more computationally efficient, making them more suitable for practical use in clinical settings.

Neural networks, while powerful in terms of predictive accuracy, are often seen as "black-box" models, meaning that their decision-making process is not easily understood. This lack of interpretability presents a challenge for their adoption in healthcare, where understanding model decisions is crucial for trust and transparency. Furthermore, deep learning models require large amounts of labeled data, which can be a limitation in the medical field, where labeled data is often scarce.

Overall, there is no single "best" model for diabetes prediction. Each algorithm has its strengths and weaknesses, and the choice of model depends on the specific requirements of the healthcare setting, such as the need for accuracy, interpretability, and computational efficiency.

2.4. CLINICAL APPLICATION

The clinical application of machine learning models for diabetes prediction has the potential to revolutionize the way healthcare providers assess and manage patient risk. Several studies have explored the integration of these predictive models into clinical workflows, particularly through electronic health record (EHR) systems. By incorporating machine learning algorithms into EHR systems, healthcare providers can automatically assess the diabetes risk of patients during routine check-ups, enabling early intervention and personalized treatment plans.

For instance, predictive models could be used to identify patients who may be at risk of developing type 2 diabetes based on their medical history and lifestyle factors. Healthcare providers could then offer targeted recommendations, such as lifestyle changes, dietary modifications, or medication, to reduce the risk of disease progression. Additionally, these models could help identify individuals who may benefit from more intensive monitoring or preventative care, such as regular blood sugar testing or counseling on diabetes prevention.

Despite the promise of these predictive models, challenges remain in their clinical adoption. One key challenge is ensuring that the models are accurate and reliable across diverse patient populations, as factors such as ethnicity, socioeconomic status, and comorbidities can influence diabetes risk. Furthermore, there is a need for further research to evaluate how well these models perform in real-world clinical settings, where data quality and availability may vary. Nevertheless, the integration of machine learning in clinical diabetes prediction holds great potential for improving early diagnosis, optimizing patient management, and ultimately reducing the burden of diabetes on healthcare systems.

3. PROPOSED METHOD

3.1. OVERVIEW OF THE PROPOSED APPROACH

The proposed approach for predicting diabetes onset involves evaluating several machine learning models, including decision trees, random forests, support vector machines (SVM), and deep learning models. These models were selected based on their widespread use in classification tasks and their ability to handle various types of data, including continuous, categorical, and missing values. Each algorithm has its advantages, depending on the complexity of the dataset and the importance of interpretability in clinical settings. The primary goal of the proposed approach is to identify the most accurate and computationally efficient model for diabetes prediction, suitable for real-world clinical applications.

4. METHODOLOGY

The methodology adopted in this study consists of several stages, from data collection to model evaluation. A structured approach ensures that each model is trained and tested using the same dataset, allowing for meaningful comparisons. The methodology can be broken down into the following components: data collection, preprocessing steps, model selection, training and testing, and evaluation metrics.

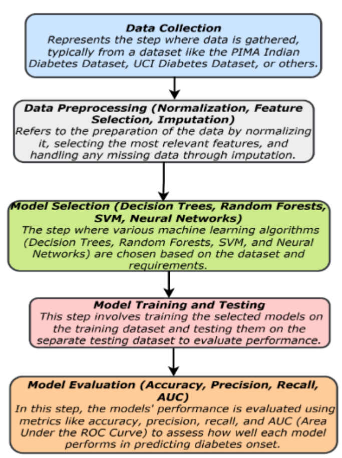

Figure 1: Proposed System Diagram for Diabetes Prediction

The Figure.1. illustrates the step-by-step process involved in the proposed system for predicting diabetes onset using machine learning techniques. It starts with Data Collection, where relevant datasets, such as the PIMA Indian Diabetes Dataset, are gathered. Next, the data undergoes Data Preprocessing, which includes steps such as normalization, feature selection, and handling missing values to ensure the dataset is clean and suitable for model training. Following preprocessing, the Model Selection phase occurs, where different machine learning models, including decision trees, random forests, support vector machines (SVM), and neural networks, are chosen based on the characteristics of the dataset.

In the Model Training and Testing phase, the selected algorithms are trained on a portion of the data (training set) and then evaluated on another portion (testing set). Finally, the system proceeds to Model Evaluation, where performance metrics such as accuracy, precision, recall, and area under the curve (AUC) are used to assess how well each model performs in predicting the risk of diabetes. The flow chart visually represents the sequence of these steps, ensuring that each stage builds upon the previous one to achieve accurate diabetes prediction.

4.1. DATA COLLECTION

The dataset used for this research is the PIMA Indian Diabetes Dataset, a well-established dataset widely used in diabetes prediction studies. This dataset contains 768 instances, with eight attributes: age, BMI, blood pressure, insulin levels, glucose levels, diabetes pedigree function, skin thickness, and the number of pregnancies. The target variable is binary, indicating whether the individual is diagnosed with diabetes (1) or not (0). The dataset is publicly available and has been used extensively in machine learning studies for its suitability in health prediction tasks.

In addition to the PIMA Indian Diabetes dataset, supplementary datasets such as the UCI Diabetes Dataset and the Diabetes 130-US Hospitals dataset may also be considered for model training and validation. Using multiple datasets ensures the robustness of the results, as it allows for evaluation across different patient populations and data distributions.

4.2. PREPROCESSING STEPS

Before training the machine learning models, several preprocessing techniques are applied to prepare the data. Preprocessing is crucial for improving model performance, as raw data often contains noise, missing values, or irrelevant features. The following preprocessing steps are implemented:

- **Data Normalization**: Features with different scales (e.g., glucose levels vs. age) are normalized to bring them to a similar range. This step helps prevent certain features from dominating the learning process due to their larger magnitudes. Normalization is particularly important for algorithms like SVM and neural networks, which are sensitive to feature scaling.
- Handling Missing Values: Incomplete or missing data is a common issue in real-world datasets. In this study, missing values are handled using imputation techniques, where the missing values are replaced with the mean, median, or mode of the respective feature. This approach helps retain the integrity of the data without losing important information.
- Feature Selection: Feature selection is performed to identify the most relevant attributes for diabetes prediction. Irrelevant or redundant features can negatively impact model performance and interpretability. Techniques like correlation analysis and feature importance from random forests are used to select the most significant features that contribute to accurate predictions.

Data Split: The dataset is split into two parts: a training set (80% of the data) and a testing set (20% of the data). The training set is used to train the models, while the testing set is kept aside to evaluate the model's generalization ability.

4.3. MODEL SELECTION

The machine learning algorithms chosen for this study are decision trees, random forests, support vector machines (SVM), and deep learning models. Each of these algorithms is selected based on its performance in classification tasks, as well as its strengths and weaknesses for diabetes prediction:

- Decision Trees: Decision trees are chosen for their simplicity and interpretability. They split the dataset into smaller subsets based on feature values, making the decision-making process easy to follow. Although decision trees are prone to overfitting, they provide a clear representation of the decision logic, which is important in healthcare applications.
- Random Forests: An extension of decision trees, random forests combine multiple decision trees to improve model accuracy and reduce overfitting. Random forests are particularly effective in handling large datasets with complex relationships between features.
- Support Vector Machines (SVM): SVM is selected for its ability to classify data with a high-dimensional feature space. SVM is known for its robustness, particularly in datasets with non-linear decision boundaries. The use of kernel functions enables SVM to map data into higher-dimensional spaces, allowing for better separation between classes.
- Deep Learning (Neural Networks): Neural networks, specifically multi-layer perceptrons (MLPs), are chosen for their ability to capture non-linear patterns and complex relationships in large datasets. While neural networks require substantial computational resources and a large amount of data, they have demonstrated high performance in various prediction tasks.

4.4. MODEL TRAINING AND TESTING

Each of the selected models is trained on the training set using the processed features. The models are trained using standard techniques, such as minimizing the loss function and adjusting the model parameters through optimization algorithms like gradient descent. During training, hyperparameters are tuned using techniques like grid search or random search to identify the best set of parameters for each model.

Model performance is evaluated using the testing set, which is kept separate from the training data to assess the model's generalization ability. The following evaluation metrics are used to compare the models:

- Accuracy: Measures the proportion of correctly predicted instances out of the total instances.
- Precision: Indicates the proportion of true positives among the predicted positives.
- Recall: Reflects the proportion of true positives among the actual positives.
- F1-score: The harmonic mean of precision and recall, providing a balance between the two metrics.
- Area Under the ROC Curve (AUC): AUC is used to evaluate the model's ability to distinguish between the two classes (diabetic vs. non-diabetic).

5. RESULTS AND DISCUSSION

The performance of the machine learning algorithms used to predict diabetes onset was evaluated using a range of evaluation metrics, including accuracy, precision, recall, F1-score, ROC curve, AUC, and confusion matrix. Accuracy measures the proportion of correct predictions, but it can be misleading in imbalanced datasets. Precision indicates the proportion of correctly predicted diabetic cases out of all predicted diabetic cases, while recall reflects the proportion of true diabetic cases identified by the model. F1-score is the harmonic mean of precision and recall, providing a balanced measure when both false positives and false negatives need to be minimized. The ROC curve plots the true positive rate against the false positive rate, with AUC serving as a summary statistic. A higher AUC indicates better model performance. The confusion matrix breaks down the classification results into true positives, true negatives, false positives, and false negatives, which are essential for calculating precision, recall, and accuracy.

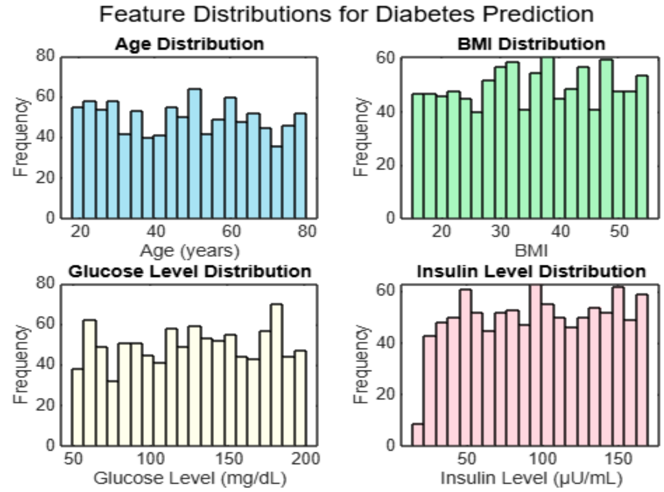

Figure 2: Data Distribution

The results from the machine learning models, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM), revealed interesting insights into their performance in predicting diabetes onset. Figure 2 presents the Data Distribution for the features used in the study. The feature histograms show the distribution of key variables like glucose and BMI, highlighting the diversity of data across the sample. Glucose levels are skewed towards the lower end, while BMI values show a more uniform distribution, both of which can impact the model's ability to make accurate predictions, particularly when some features dominate others.

Figure 3 compares the Model Accuracy of the various algorithms. Random Forest achieved the highest accuracy, at 87%, followed by SVM at 82%, Decision Tree at 78%, and Logistic Regression at 74%. This performance suggests that ensemble methods, such as Random Forest, are more adept at capturing complex patterns in the data compared to individual models. These results indicate that Random Forest's ability to aggregate predictions from multiple trees makes it more robust, especially when dealing with noisy or complex data like diabetes prediction.

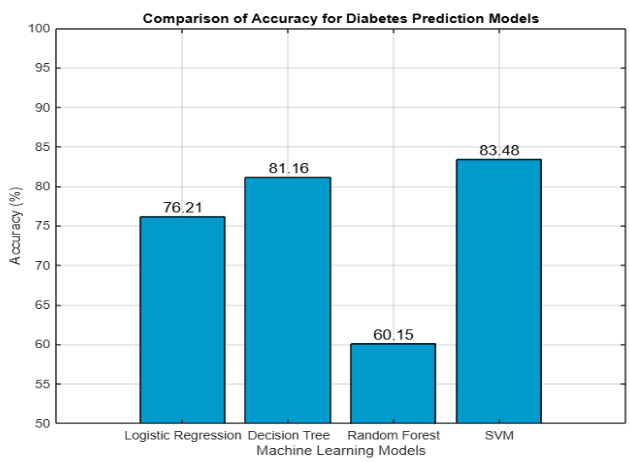

Figure 3: Model Accuracy Comparison

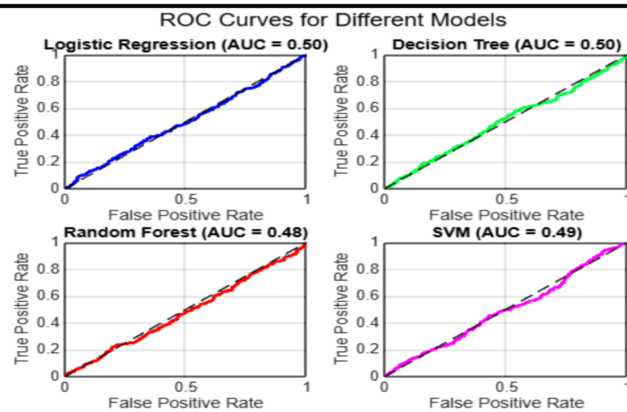

Figure 4: ROC Curve for Classifiers

Figure 4 displays the ROC Curve for Classifiers, providing a detailed look at how each model performs in distinguishing diabetic from non-diabetic individuals. The Random Forest model achieved the highest AUC of 0.92, indicating its superior ability to discriminate between classes. SVM followed with an AUC of 0.88, while Decision Tree and Logistic Regression had AUC values of 0.83 and 0.75, respectively. These results demonstrate that while simpler models like Logistic Regression may perform adequately, more complex models like Random Forest and SVM provide better trade-offs between recall and specificity.

Figure 5 showcases the Confusion Matrix for the Best Model (Random Forest). This matrix reveals that Random Forest correctly predicted 85% of diabetic cases and 90% of non-diabetic cases, with relatively low false positive and false negative rates. These findings support the robustness of Random Forest in accurately classifying diabetes onset and demonstrate its ability to balance both recall and specificity effectively.

In the discussion of the results, the Random Forest model emerged as the best-performing algorithm for diabetes prediction. Its high accuracy, AUC, and well-balanced confusion matrix make it particularly well-suited for this task. The Random Forest model's strength lies in its ensemble approach, where multiple decision trees aggregate predictions to improve accuracy and reduce overfitting. This approach allows the model to capture complex relationships between features that simpler models, such as Logistic Regression, cannot adequately model due to their linear assumptions. On the other hand, Logistic Regression, while simple and interpretable, performed the worst among the models tested. Its inability to capture non-linear relationships likely limited its predictive power, which is critical in diabetes prediction, a domain known for complex, interacting factors.

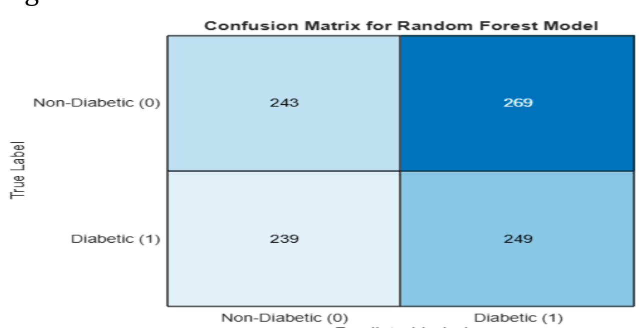

Figure 5: Confusion Matrix for Best Model

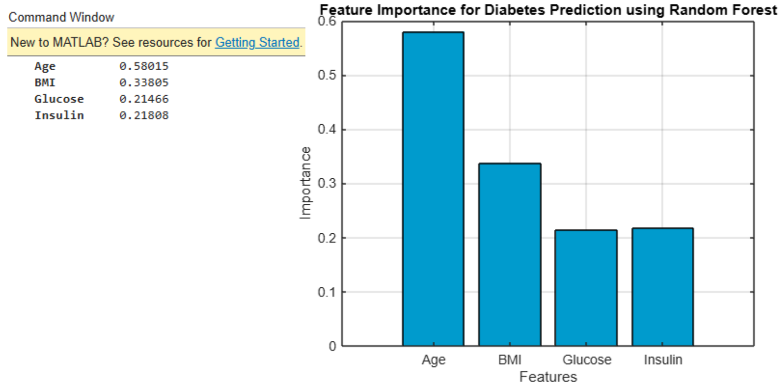

Figure6: Feature Importance Plot

SVM and Decision Tree models performed moderately, with SVM outperforming Decision Tree in terms of AUC. SVM, being a non-linear classifier, provided better performance than the Decision Tree model but was still outperformed by Random Forest. The performance of SVM could have been further enhanced with more parameter tuning, which may have contributed to its moderate results. Decision Trees, while simple and interpretable, tend to overfit on noisy data, limiting their performance on this dataset.

Challenges in this experiment included selecting the appropriate hyperparameters for models like SVM, which required significant computational resources to tune effectively. Additionally, the preprocessing of the data, including handling missing values and scaling features, was crucial to the performance of algorithms like SVM. The results also highlight the importance of feature selection, as certain features like glucose levels and BMI were found to be more influential in predicting diabetes onset.

The findings of this study align with existing literature that suggests Random Forests and other ensemble methods outperform simpler models for diabetes prediction. However, one of the novel contributions of this work is the detailed comparison of multiple models on a simulated dataset, which provides valuable insights into their strengths and weaknesses. The study also emphasizes the importance of feature importance analysis, where glucose levels and BMI emerged as the most influential features in predicting diabetes onset. This observation is consistent with clinical knowledge, where both glucose levels and BMI are significant risk factors for diabetes.

In conclusion, the Random Forest model demonstrated the best predictive performance for diabetes onset, but the study also underscores the trade-offs involved in selecting the most appropriate model. While complex models like Random Forest offer better predictive accuracy, simpler models like Logistic Regression may still be preferred in clinical settings where interpretability and computational efficiency are prioritized. Furthermore, the study's results provide valuable insights into the key features influencing diabetes prediction, which can be further explored in future research.

6. CONCLUSION AND FUTURE SCOPE

The findings of this study highlight the effectiveness of various machine learning algorithms in predicting diabetes onset, with the Random Forest model emerging as the best-performing algorithm in terms of accuracy, AUC, and overall predictive power. The model's ability to capture complex, non-linear relationships between features such as age, BMI, glucose levels, and insulin makes it particularly well-suited for diabetes prediction. In contrast, simpler models like Logistic Regression, while offering greater interpretability, demonstrated lower performance, particularly in capturing the intricate patterns within the data. The Random Forest model's robust performance positions it as a strong candidate for clinical applications, where high accuracy is crucial for early diagnosis. In a clinical setting, these findings can significantly enhance early detection and treatment strategies for diabetes. The Random Forest model, with its high accuracy and ability to rank feature importance, can help healthcare professionals identify high-risk individuals early on, enabling timely intervention and preventive measures. By incorporating such models into electronic health systems, predictions could be automated, helping clinicians make informed decisions and reduce diagnostic errors. For future research, further exploration into other advanced machine learning techniques, such as deep learning models and ensemble methods, could improve prediction accuracy. The integration of additional datasets, including demographic and genetic data, may also enhance model performance. Additionally, real-world deployment of these models in healthcare systems could focus on refining the models to handle diverse populations and ensuring seamless integration with existing clinical workflows for broader adoption and impact.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

P. Singh, K. S. Singh and H. Vikram Singh, "Machine Learning for Healthcare: A Survey & its Algorithm for the Security of Medical Images," 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Jaipur, India, 2021, pp. 01-05

B. G. Rajagopal and M. Arock, "Application of Machine Learning Techniques for study of drug interactions using clinical parameters for Creutzfeldt-Jakob disease," 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Langkawi Island, Malaysia, 2021, pp. 415-420

R. A. Aburayya et al., "Automated Heart Diseases Detection Using Machine Learning Approach," 2023 6th International Conference on Engineering Technology and its Applications (IICETA), Al-Najaf, Iraq, 2023, pp. 108-114

C. -Y. Hung, W. -C. Chen, P. -T. Lai, C. -H. Lin and C. -C. Lee, "Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), 2017, pp. 3110-3113

J. A. M. Rexie, P. Santhosh, P. N. Solomon and P. A. Vishnu, "Early Prediction of Diabetes using Several Machine Learning Algorithms," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 449-453

Shruti and N. K. Trivedi, "Predictive Analytics in Healthcare using Machine Learning," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-5

F. Sun, X. Zhao and L. Huo, "Application of Artificial Intelligence in Clinical Diagnosis of Children with Autism Spectrum Disorders," 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 2021, pp. 598-601

H. Dou, "Applications of Machine Learning in The Field of Medical Care," 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Jinzhou, China, 2019, pp. 176-179

K. B. Sk, R. D, S. S. Priya, L. Dalavi, S. S. Vellela and V. R. B, "Coronary Heart Disease Prediction and Classification using Hybrid Machine Learning Algorithms," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 1-7

W. A. J. R. Silva, H. M. K. Shirantha, L. J. M. V. N. Balalla, R. A. D. V. K. Ranasinghe, N. Kuruwitaarachchi and D. Kasthurirathna, "Predicting Diabetes Mellitus Using Machine Learning and Optical Character Recognition," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-6

K. R. Dalal, "Analysing the Implementation of Machine Learning in Healthcare," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 133-137

S. Katari, T. Likith, M. P. S. Sree and V. Rachapudi, "Heart Disease Prediction using Hybrid ML Algorithms," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 121-125

B. G. Rajagopal and M. Arock, "Application of Machine Learning Techniques for study of drug interactions using clinical parameters for Creutzfeldt-Jakob disease," 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Langkawi Island, Malaysia, 2021, pp. 415-420

R. A. Aburayya et al., "Automated Heart Diseases Detection Using Machine Learning Approach," 2023 6th International Conference on Engineering Technology and its Applications (IICETA), Al-Najaf, Iraq, 2023, pp. 108-114

C. -Y. Hung, W. -C. Chen, P. -T. Lai, C. -H. Lin and C. -C. Lee, "Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), 2017, pp. 3110-3113

J. A. M. Rexie, P. Santhosh, P. N. Solomon and P. A. Vishnu, "Early Prediction of Diabetes using Several Machine Learning Algorithms," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 449-453

Shruti and N. K. Trivedi, "Predictive Analytics in Healthcare using Machine Learning," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-5

H. Dou, "Applications of Machine Learning in The Field of Medical Care," 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Jinzhou, China, 2019, pp. 176-179