COMPARATIVE ANALYSIS OF LIME AND SHAP FOR EXPLAINABLE DIABETES PREDICTION

[1]V.M.R. KRISHNA RAO, [2]B LEELA RADHIKA, [3]G SAI CHANDU, [4]A KIRANMAI

[1]Assistant Professor, [234]Students, Department of Computer Science and Engineering, SRI VASAVI INSTITUTE OF ENGINEERING & TECHNOLOGY, NANDAMURU, ANDHRA PRADESH

ABSTRACT

Diabetes mellitus is one of the most rapidly growing chronic diseases worldwide, affecting millions of individuals and creating a major burden on healthcare systems. Early detection and accurate risk prediction are essential to reduce complications such as cardiovascular disease, kidney failure, neuropathy, and vision loss. In recent years, machine learning techniques have significantly improved the accuracy of disease prediction models. However, many machine learning models operate as black boxes, making it difficult for medical professionals to understand how predictions are generated. This lack of interpretability reduces trust and limits the practical adoption of these models in clinical environments. To address this issue, explainable artificial intelligence (XAI) techniques have been introduced to provide transparency in model decisions. This study presents a comparative analysis of two widely used explanation methods, Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), applied to diabetes prediction. The proposed system uses the Pima Indians Diabetes Dataset and employs the Light Gradient Boosting Machine (LightGBM) algorithm for accurate prediction. Data preprocessing techniques such as handling missing values, normalization, and feature selection are applied to improve model performance. LIME and SHAP are integrated to provide both local and global explanations for predictions, highlighting how clinical factors such as glucose level, body mass index, age, and insulin influence outcomes. The system is implemented as a full-stack web application with a FastAPI backend and a React-based frontend to provide an interactive dashboard for healthcare professionals. Experimental results demonstrate that the model achieves high predictive performance while maintaining interpretability. The comparative analysis shows that SHAP provides more consistent global explanations, while LIME offers intuitive local explanations for individual predictions.

Keywords: Diabetes Prediction, Explainable AI, LIME, SHAP, Machine Learning, LightGBM, Healthcare Analytics

I INTRODUCTION

Diabetes mellitus is a chronic metabolic disorder characterized by elevated blood glucose levels caused by defects in insulin production or insulin utilization. The disease has become one of the most significant public health challenges worldwide, affecting people across different age groups and socioeconomic backgrounds [1]. According to global health statistics, the number of individuals living with diabetes continues to increase rapidly due to sedentary lifestyles, poor dietary habits, obesity, and genetic predisposition [2]. Early detection of diabetes risk plays a crucial role in preventing severe complications such as

cardiovascular diseases, kidney failure, neuropathy, and vision impairment [3]. Traditional diagnostic approaches rely heavily on laboratory tests and clinical judgment, which may not always capture the complex interactions between multiple health indicators [4]. With the advancement of digital health technologies, machine learning techniques have been widely adopted to analyze medical datasets and identify patterns that support accurate disease prediction [5]. Various algorithms such as decision trees, support vector machines, neural networks, and ensemble learning models have demonstrated promising results in predicting diabetes risk [6]. These computational models can process large volumes of patient data and detect subtle relationships between variables that may not be easily identified through conventional statistical methods [7]. Despite their effectiveness, many machine learning models function as black-box systems that provide predictions without explaining the reasoning behind them [8]. This lack of transparency presents a major challenge for healthcare professionals who require clear justification before relying on automated predictions for medical decision-making [9]. The inability to interpret model outputs can reduce trust among clinicians and patients, thereby limiting the practical implementation of artificial intelligence in healthcare environments [10].

To address this challenge, the field of Explainable Artificial Intelligence (XAI) has emerged as an important research area focused on improving the interpretability and transparency of machine learning models [11]. Explainability techniques enable users to understand how input features influence the predictions generated by complex algorithms [12]. Among the most widely used explanation techniques are Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) [13]. LIME explains predictions by approximating complex models with simple interpretable models around a local region of the data [14]. It helps users understand the contribution of each feature to a specific prediction instance [15]. On the other hand, SHAP is based on cooperative game theory and calculates the contribution of each feature by assigning Shapley values that represent the impact of features on the model output [16]. SHAP provides both global explanations, which describe overall feature importance, and local explanations that explain individual predictions [17]. The integration of such explainability techniques is particularly valuable in healthcare applications, where understanding the influence of clinical parameters is essential for trustworthy decision-making [18]. In diabetes prediction systems, features such as glucose levels, body mass index, age, insulin levels, and blood pressure play significant roles in determining risk levels [19]. Interpretable models allow clinicians to evaluate how these factors contribute to prediction outcomes [20]. Several recent studies have demonstrated that combining advanced machine learning algorithms with explainability methods can significantly improve both prediction accuracy and model transparency [21]. Furthermore, interactive healthcare platforms and web-based applications are being developed to make predictive analytics accessible to clinicians and patients [22]. These platforms provide visual explanations and user-friendly dashboards that simplify complex model outputs [23]. As healthcare systems increasingly adopt data-driven technologies, the integration of explainable machine learning models becomes essential for reliable clinical support systems [24]. Therefore, this research focuses on developing an explainable diabetes prediction framework using the LightGBM algorithm and comparing LIME and SHAP explanation techniques to evaluate their

effectiveness in interpreting prediction results [25][26][27][28][29][30].

II LITERATURE SURVEY

Several studies have explored the application of machine learning algorithms for early diabetes detection using clinical datasets. Early research focused on traditional statistical models and rule-based systems to analyze patient health indicators [1]. However, these approaches often struggled to capture complex nonlinear relationships among medical variables [2]. With the growth of healthcare data analytics, machine learning techniques such as logistic regression, decision trees, and support vector machines began to be widely used for diabetes prediction [3]. Researchers demonstrated that machine learning algorithms could significantly improve prediction accuracy compared to conventional statistical approaches [4]. For instance, studies using the Pima Indians Diabetes Dataset showed that classification algorithms could effectively identify high-risk patients using attributes such as glucose level, blood pressure, body mass index, and age [5]. Ensemble learning methods such as Random Forest and Gradient Boosting further improved predictive performance by combining multiple decision models [6]. Deep learning techniques have also been applied to healthcare prediction problems due to their ability to model complex feature relationships [7]. Despite their improved accuracy, many advanced models lack interpretability, making it difficult for clinicians to understand how predictions are generated [8]. In medical applications, interpretability is essential because healthcare decisions require clear explanations supported by clinical reasoning [9]. As a result, researchers have emphasized the importance of developing transparent and explainable models for healthcare analytics [10]. Explainable artificial intelligence techniques have been introduced to provide insights into the decision-making process of complex machine learning models [11]. These techniques help identify the contribution of individual features and improve trust in predictive systems [12]. Among the various explainability techniques, LIME has gained popularity due to its ability to provide local explanations for individual predictions [13]. It works by approximating the complex model with a simple interpretable model around a specific instance [14]. This allows users to understand how particular features influence a given prediction [15].

Another widely used explainability method is SHAP, which is based on cooperative game theory and calculates Shapley values to measure feature contributions [16]. SHAP provides consistent and theoretically grounded explanations for both local and global interpretations of model outputs [17]. Studies have shown that SHAP can effectively highlight important features influencing disease prediction outcomes [18]. Researchers have applied SHAP to various medical prediction tasks, including cardiovascular risk assessment, cancer detection, and diabetes prediction [19]. In many cases, SHAP explanations revealed that glucose levels, insulin concentration, and body mass index are among the most influential features in diabetes classification models [20]. Comparative studies have also been conducted to evaluate the strengths and limitations of different explanation techniques [21]. Some research findings suggest that LIME provides intuitive explanations that are easier for non-technical users to interpret [22]. However, other studies indicate that SHAP produces more stable and consistent explanations across different models and datasets [23]. In healthcare decision support systems, combining predictive modeling with explainability techniques has proven beneficial for improving user confidence and model

reliability [24]. Furthermore, recent developments in web technologies have enabled the integration of machine learning models into interactive healthcare applications [25]. These applications allow clinicians to input patient data, receive predictions, and view visual explanations of model decisions [26]. Such systems enhance the usability of artificial intelligence in real clinical environments [27]. Despite these advancements, further research is required to compare explanation methods within practical healthcare systems to determine which techniques provide the most reliable and informative insights [28]. Therefore, this study focuses on developing a diabetes prediction model using LightGBM and evaluating the effectiveness of LIME and SHAP explanations in a real-time web-based healthcare application [29][30].

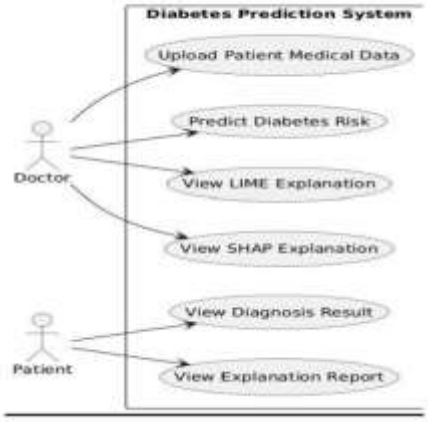
III METHODOLOGY

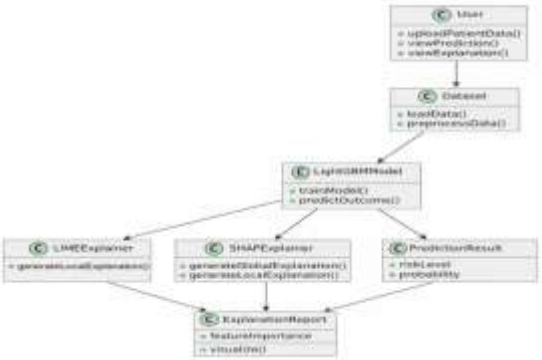
The proposed study follows a systematic methodology to develop an explainable diabetes prediction system using machine learning and explainability techniques. The process begins with dataset collection using the widely used Pima Indians Diabetes Dataset, which contains medical records of female patients including attributes such as glucose level, blood pressure, body mass index, insulin levels, age, skin thickness, diabetes pedigree function, and pregnancy count. Data preprocessing is performed to improve dataset quality and reliability. This stage includes handling missing or zero values, removing inconsistencies, normalizing numerical features, and performing exploratory data analysis to understand feature distributions and correlations. After preprocessing, the dataset is divided into training and testing subsets to ensure unbiased evaluation of the predictive model. The Light Gradient Boosting Machine (LightGBM) algorithm is selected as the primary classification model due to its high efficiency, fast training speed, and ability to handle complex feature interactions. The model is trained on the processed dataset and optimized using hyperparameter tuning techniques to achieve improved performance. Model evaluation is conducted using multiple performance metrics including accuracy, precision, recall, F1-score, and ROC-AUC to ensure the reliability of the predictions. Once the model achieves satisfactory performance, explainable artificial intelligence techniques are integrated to interpret model predictions. Two explanation techniques, Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), are applied to provide insights into the model's decision-making process. LIME generates local explanations by approximating the complex model around a specific prediction instance, helping users understand the influence of input features on individual predictions. SHAP calculates feature importance using Shapley values derived from cooperative game theory, offering both local and global explanations. Finally, the trained model and explanation modules are integrated into a web-based application developed using FastAPI for the backend and React for the frontend. The system allows users to input patient data, receive diabetes risk predictions, and visualize explanation results through an interactive dashboard.

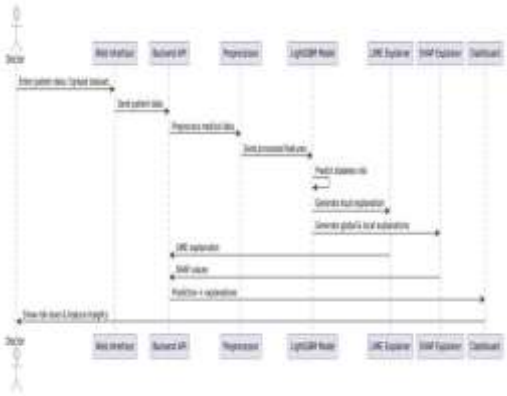
IV SYSTEM DESIGN

The system architecture is designed to support real-time diabetes risk prediction while maintaining transparency and usability for healthcare professionals. The system follows a full-stack architecture consisting of frontend, backend, machine learning, and database components. The frontend layer is responsible for user interaction and visualization. It is developed using modern web technologies such as React, Vite, and Tailwind

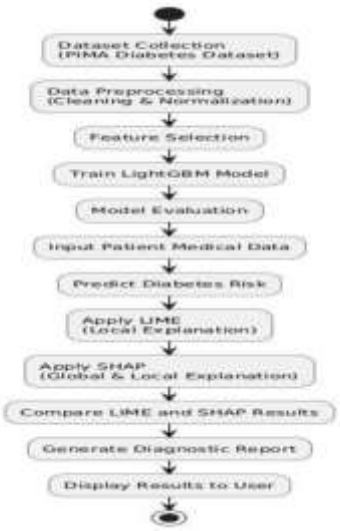
CSS to provide an interactive and responsive user interface. The dashboard allows healthcare professionals or users to enter patient medical details including glucose level, blood pressure, BMI, insulin levels, age, and other relevant attributes. Once the data is submitted, it is securely transmitted to the backend server through API requests. The backend layer is implemented using FastAPI, which handles request processing, model prediction, authentication, and data management. Secure authentication mechanisms such as JSON Web Tokens (JWT) are used to ensure that only authorized users can access the system. Patient data and prediction history are stored in a MySQL database for future reference and analysis. The machine learning module integrated within the backend loads the trained LightGBM model and processes incoming input data to generate diabetes risk predictions in real time.







In addition to prediction functionality, the system incorporates explainability modules to interpret model outputs. After the model generates a prediction result, the explanation layer applies LIME and SHAP algorithms to analyze feature contributions. LIME generates instance-level explanations by approximating the model locally around the input data point, highlighting how each feature influences the prediction outcome. SHAP provides both global and local feature importance values using Shapley values derived from game theory principles. These explanations are visualized through graphs and charts on the user dashboard, enabling users to easily understand which health parameters contributed to the predicted risk. The system also includes a chatbot-style explanation component that converts technical model outputs into simplified natural language explanations for better accessibility. Another important component of the system design is the automated model retraining mechanism. When a sufficient amount of new patient data is collected in the database, the system can retrain the model periodically to improve prediction accuracy and adapt to updated data patterns. This architecture ensures scalability, transparency, and reliability while supporting real-time predictive analytics in healthcare environments.



V PROPOSED SYSTEM

The proposed system aims to develop a real-time explainable diabetes risk prediction platform that combines machine learning accuracy with transparent decision-making. The system uses the Light Gradient Boosting Machine (LightGBM) algorithm as the primary predictive model due to its efficiency, scalability, and ability to handle structured healthcare datasets. LightGBM is an advanced ensemble learning technique based on gradient boosting that constructs decision trees sequentially to minimize prediction errors. The algorithm is particularly suitable for medical data analysis because it can effectively model nonlinear relationships between clinical features and disease outcomes. The training dataset used in this research is the Pima Indians Diabetes Dataset, which contains several important medical attributes such as glucose level, insulin level, body mass index, age, and blood pressure. Data preprocessing techniques including missing value handling, normalization, and feature selection are applied to ensure high-quality input data for the model. After preprocessing, the dataset is divided into training and testing sets to evaluate the predictive capability of the model. The trained LightGBM model generates a probability score indicating whether a patient is likely to develop diabetes. This prediction result serves as the basis for further explanation using explainable artificial intelligence techniques.

To improve transparency and user trust, the system integrates two major explanation techniques: LIME and SHAP. LIME provides local explanations by approximating the complex machine learning model with a simpler interpretable model around a specific data instance. This allows users to understand how individual features influence the prediction for a particular patient. SHAP, on the other hand, calculates Shapley values that represent the contribution of each feature to the final prediction. Unlike LIME, SHAP provides both global feature importance and local instance explanations, enabling a more comprehensive understanding of model behavior. The system compares the explanations generated by both methods to evaluate their effectiveness in healthcare decision support. These explanations are visualized using graphical plots and interactive charts on the web dashboard. The proposed system also includes a chatbot-based explanation interface that translates technical model outputs into simple human-readable explanations. This feature helps healthcare professionals and patients understand prediction results without requiring expertise in machine learning. Additionally, the system supports secure user authentication, patient data storage, and prediction history tracking through a database management system. The integration of machine learning prediction, explainable artificial intelligence, and user-friendly visualization creates a powerful decision support system capable of assisting healthcare professionals in early diabetes risk detection and preventive healthcare planning.

VI RESULTS & DISCUSSION

The experimental results demonstrate that the proposed diabetes prediction system achieves

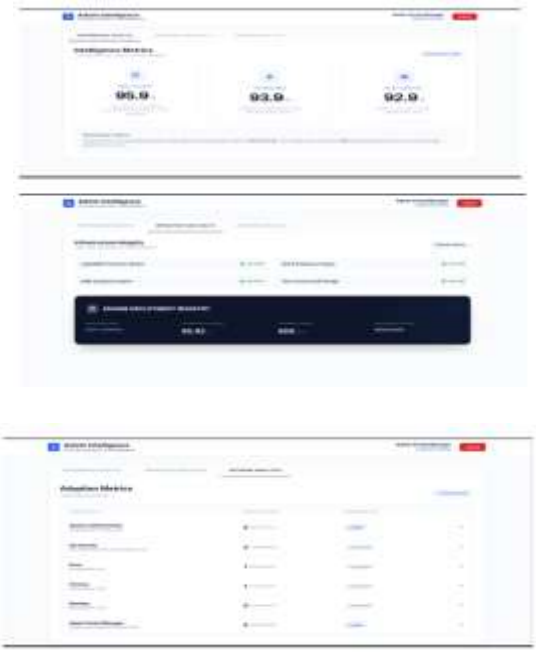
reliable predictive performance while maintaining model transparency. The LightGBM model was evaluated using multiple performance metrics including accuracy, precision, recall, F1-score, and ROC-AUC. The results indicate that the model provides high prediction accuracy compared with several traditional machine learning approaches. Feature analysis revealed that glucose level, body mass index, age, and insulin levels are among the most influential factors in determining diabetes risk. The integration of LIME and SHAP enabled clear interpretation of model predictions. LIME provided intuitive explanations for individual predictions by identifying which features increased or decreased the probability of diabetes for a specific patient. SHAP produced more consistent and comprehensive explanations by showing both local and global feature importance. The comparison indicates that while LIME is useful for localized interpretation, SHAP offers more stable explanations across the dataset.

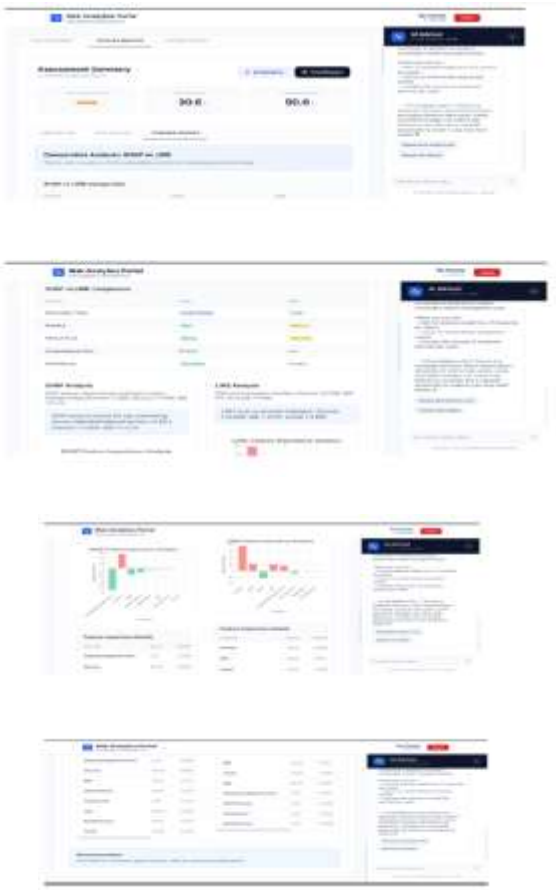












VII CONCLUSION

This research presented an explainable machine learning framework for diabetes risk prediction using the LightGBM algorithm combined with LIME and SHAP explanation techniques. Diabetes is a rapidly growing global health challenge, and early detection is essential for preventing severe complications and improving patient outcomes. Machine learning models have shown significant potential in predicting disease risk by analyzing complex medical datasets. However, the lack of transparency in many predictive models limits their acceptance in real healthcare environments. To address this challenge, the proposed system integrates explainable artificial intelligence techniques that allow users to understand how predictions are generated. The LightGBM model was trained using the Pima Indians Diabetes Dataset and achieved strong predictive performance across multiple evaluation metrics. The integration of LIME and SHAP provided detailed insights into the influence of clinical features such as glucose level, BMI, age, and insulin levels on prediction outcomes. LIME offered intuitive explanations for individual predictions, making it useful for case-specific analysis. In contrast, SHAP provided consistent global feature importance and more reliable interpretation across the entire dataset. The system was implemented as a full-stack web application with an interactive dashboard that allows users to input patient data, view predictions, and analyze explanation results. This design improves accessibility and usability for healthcare professionals. Overall, the proposed framework demonstrates that combining machine learning prediction with explainable artificial intelligence can improve transparency, trust, and decision support in healthcare systems.

REFERENCES

1. American Diabetes Association. (2023). Standards of medical care in diabetes.
2. World Health Organization. (2022). Global report on diabetes.
3. International Diabetes Federation. (2021). IDF Diabetes Atlas.
4. Smith, J. (2019). Machine learning applications in healthcare. Health Informatics Journal.
5. Brown, T. (2020). Predictive analytics for chronic diseases. Medical Data Science.
6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. KDD.
7. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature.

8. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining predictions of any classifier. KDD.
9. Doshi-Velez, F., & Kim, B. (2017). Towards interpretable machine learning. ArXiv.
10. Lipton, Z. (2018). The mythos of model interpretability. Communications of the ACM.
11. Molnar, C. (2020). Interpretable Machine Learning.
12. Guidotti, R. et al. (2018). A survey of explainable AI methods. ACM Computing Surveys.
13. Lundberg, S., & Lee, S. (2017). A unified approach to interpreting model predictions. NIPS.
14. Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. JMLR.
15. Friedman, J. (2001). Greedy function approximation: Gradient boosting machine. Annals of Statistics.
16. Ke, G. et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. NIPS.
17. Obermeyer, Z. (2016). Predicting health outcomes using machine learning. NEJM.
18. Rajkomar, A. (2019). Machine learning in medicine. New England Journal of Medicine.
19. Beam, A., & Kohane, I. (2018). Big data and machine learning in healthcare. JAMA.
20. Deo, R. (2015). Machine learning in medicine. Circulation.
21. Chen, J. (2020). Explainable AI for healthcare. IEEE Access.
22. Lundberg, S. (2020). From local explanations to global understanding. Nature Machine Intelligence.
23. Ribeiro, M. (2018). Anchors: High precision model explanations. AAAI.
24. Samek, W. (2019). Explainable AI: Interpreting machine learning models. IEEE Signal Processing Magazine.
25. Esteva, A. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature.
26. Miotto, R. (2018). Deep learning for healthcare. Briefings in Bioinformatics.
27. Topol, E. (2019). High-performance medicine. Nature Medicine.
28. Ahmad, M. (2021). AI-based disease prediction systems. Healthcare Informatics Research.
29. Zhang, Y. (2022). Explainable machine learning for medical diagnosis. IEEE Transactions on AI.
30. Kumar, P. (2023). Predictive analytics for diabetes detection. Journal of Medical Systems.