Amity Business Journal
2026, Vol. IX, No. 1 (Print ISSN: 2278-0904)

Diabetes Prediction using SHAP and LIME Analysis: An Explainable AI Approach

Dr. Rajesh Kumar K V[1] **Konala Nikhil**[1*] **Goda Shruthi**[1]
M. Akhil Gupta[1] **Nida Shaik**[1] **Sri Priya Upadhyayula**[1]
Saad[1]

ABSTRACT

Diabetes mellitus has emerged as a major health concern globally, especially in the context of initial diagnosis in preventing complications. The work being described here outlines a framework for explainable artificial intelligence (XAI) in forecasting cases of diabetes using the PIMA Indians Diabetes dataset. The logistic regression classifier was used in training a model for the classification of patients as diabetic or non-diabetic, with the development of the interpretable model being used in explaining the importance of features for the global nature of the model using Shapley Additive explanations (SHAP) and local, instance, and case reasoning for each individual case using Local Interpretable Model-Agnostic Explanations (LIME). The technical proof of concept framework for forecasting cases of diabetes was implemented through a streamlet web application that allows users to input patient information. The model was able to perform with the top three features being glucose level, body mass index (BMI), and age. The use of dual-layered explainability in integrating predictive modelling with the application of ML demonstrates the application of interpretable ML in creating trust, transparency, and application in calculations in a clinical setting.

Keywords: Diabetes Prediction, Explainable Artificial Intelligence (XAI), SHAP, LIME.

1. Introduction

Diabetes exists in numerous forms, which include Type 1,2 & gestational diabetes; here, Type 2 is predominant. Since early detection is significant, if diabetes not treated properly may lead to complications like renal failure and vision loss.

The International Diabetes Federation (IDF) states that 537 million adults have diabetes at present. By the year 2030, this figure may rise to 640 million. India is expected to have 77 million cases, which makes it one of the countries having higher rate of diabetes in the world. A lot of people with diabetes still aren't recognised or are detected late, which can lead to heart disease, renal failure, and/or blindness. Early identification of diabetes is crucial for preventing future consequences. According to IDF, millions of adults had remained undiagnosed and in India, the burden of diabetes is substantial, affecting both expenditure on healthcare and productivity. Tools like telemedicine and wearables have given opportunity for AI tools, but such predictions must be explainable, interpretable and must be aligned with clinical practices. This underscores need for such framework which gives transparent reasoning by making use of SHAP and LIME approaches (International Diabetes Federation [IDF], 2021).

By recognising patterns from existing patient data, ML models can surpass previous methods of disease detection, such as for diabetes. In the field of health care, interpretability is extremely important because predictions impact treatment decisions directly. Clinicians depend on knowledge of the input variable's contribution to the predicted outcome to demonstrate alignment between predictions and medical knowledge. SHapley Additive explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) are two of the most common XAI technologies. SHAP follows a mathematical approach using cooperative game theories as a foundation to determine not only features contributions to the models predictive performance (global view) as well as how features collectively contribute to an individual prediction (local view). LIME builds simple interpretable models around single predictions (i.e., for individual patients) and estimates the reasoning behind the prediction.

2. Methodology

2.1 Dataset Description

This study has used the PIMA Indians Diabetes Dataset, which is an extensively used dataset that came from female patients of Pima Indian

1. *Woxsen University, Hyderabad, 502345, Telangana, India.*

descent located in Phoenix Arizona, United States (Smith et al., 1988). The PIMA Indians Diabetes Dataset has established itself as the benchmark dataset for research into medical machine learning because it is a relatively balanced-sized dataset, it also has a rich, diverse number of clinical parameters, and was also relevant in the prediction of diabetes mellitus. Here, the important thing to be noted is that clinical variables contain some zero entries, mainly due to physiological limitations like insulin and skin thickness. During the time of imputation, median values ensured to give model stability, and this may give it bias.

The dataset provided information on 768 records of patients serviced with one record described by eight independent clinical features in addition to one binary dependent variable. The independent variables were described as follows:

Pregnancies: the number of pregnancies the patient has had.

Glucose: The glucose level in the plasma 2 hours after the oral glucose tolerance test.

Diastolic Blood Pressure: The blood pressure, which is measured in mm Hg.

Skin Thickness: The thickness of the triceps skin fold measured in millimetres.

Insulin: serum insulin for 2 hours(μU/mL).

BMI: Body Mass Index

Diabetes Pedigree Function: Hereditary impact / Family history of having diabetes.

Age: The age of the patient.

Outcome will be a binary variable, where 1 indicates that the patient is a diabetic and 0 indicates that the patient is non-diabetic.

2.2 Data Preprocessing
The team established an effective preprocessing procedure because they wanted the model to achieve both dependability and optimal performance. The steps we have taken are as follows:

2.3 Exploratory Data Analysis (EDA) The project started with a series of fundamental tests. We used boxplots and histograms and correlation heatmaps to study both data distribution patterns and how different features relate to each other. The diabetes outcome showed strong ties to both BMI and glucose measurements while skin thickness and insulin levels showed multiple exceptions.

2.4 Treatment of Implausible Values
We identified zeros in clinical feature values as implausible values which we replaced with the data's median value. The median value provides better protection against extreme data points when compared to the average value.

2.5 Outlier Detection and Treatment
The Interquartile Range (IQR) method served as our chosen technique for selecting outlier data points. Insulin and other clinical features displayed considerable levels of skewness. We applied logarithmic transformations to address this skewness through our initial preprocessing methods.

2.6 Feature ScalingWe assessed the effect of two distinct clinical feature weight scaling methods on our classification results:- Min-Max Scaler: We used this to normalize all feature values to be within the [0,1] interval.

StandardScaler: We used this to normalize feature values to have a mean of 0 and a standard deviation of 1.

2.7 Train-test Split
The researchers conducted their study by dividing the data into two sections which they used for training and testing their experimental models. The research team implemented stratified random sampling to maintain the original distribution of outcomes throughout their study. The team executed the stratification process by utilizing a random number generator. Our team selected a specific random seed value to achieve reproducibility.

3. Proposed Method

The proposed framework incorporates interpretability into the extendable application of artificial intelligence (XAI) to achieve both trustworthy predictions and trustworthy decisions, which the authors describe as "Explainable AI". Using a logistic regression classifier, the model predicts patients as diabetic or non-diabetic. The authors employ two complementary post-hoc XAI methods to assist with the models interpretability: SHapley Additive exPlanations (SHAP) for global and local feature attributions, and Local Interpretable Model-agnostic Explanations (LIME) for per-instance interpretability. The model and

explanations are packaged in a Streamlit web app providing an interactive interface for user-based data input, prediction, and visualization.

3.1 Algorithms for Machine Learning:

Machine Learning is been used widely, especially in case of medical industry, to change the data but also to make the predictive models which help doctors to make accurate and efficient decisions. We have used here the 6 machine learning classifiers for testing the PIMA Indians Diabetes dataset: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and Extreme Gradient Boosting (XGBoost).

Logistic Regression (LR): A statistical procedure that models the probability of an outcome, such as diabetes, as a series of clinical attributes. LR achieved 84% accuracy and 91% recall with reasonable predictive capacity and transparency. While LR is not superior for prediction, it was chosen as the final model for two reasons: it is interpretable and can be used with SHAP and LIME.

Decision Tree (DT): Rule based classifier/hierarchical splits. Accuracy level was 78%.

Prone to overfitting, and generalization was inferior to the other models.

Random Forest (RF): A forest of decision trees that imparted some strength, accuracy was enhanced to 85%. More complexity took away from interpretability, less interpretable than LR.

Support Vector Machine (SVM): A very strong machine learning technique that can have classes with non-linear boundaries. The accuracy rate was 82%. Lower degree of interpretability with SVM took away from the potential for new uses in a clinical environment.

Naïve Bayes (NB): Probabilistic classifier that assumes the features are independent, which is not completely valid for medical data. The accuracy level was 80%, making NB a solid baseline model.

XGBoost: A gradient boosting method that achieved the highest accuracy level (86%), but was a black-box type model that limits transparency for a healthcare space.

3.2 Explainable Artificial Intelligence (XAI)

Although it is important to develop ML algorithms with high predictive accuracy, the interpretability of complex models remains a hindrance to ML implementation in healthcare. Clinicians and patients require interpretability and transparency to trust algorithms. XAI seeks to address this challenge by explaining how and why a model has arrived at a prediction. In this study, two conventional post-hoc interpretability methods (SHAP and LIME) were employed after the final logistic regression analysis.

3.3 SHapley Additive exPlanations (SHAP)

SHAP uses rooted cooperative game theory to assign each feature a Shapley value to quantify its contribution to the model output. SHAP explains information to the model globally – ranking

features (i.e., glucose, BMI, age, etc.) over the entire data set where glucose, BMI, and age were the top three features - as well as locally, which indicates which features moved a particular prediction to a specific model class. The prediction results identified patients as diabetic when glucose levels exceeded 150 mg/dL during both the global analysis and local analysis of the study.

3.4 Local Interpretable Model-Agnostic Explanations (LIME)

LIME creates local surrogate models to explain and predict the original model's output through simplified models which use interpretable components such as linear models. LIME establishes the original model's functioning through its method of generating input data variations which it employs to evaluate output results. LIME shows which features most affected the medical decision that was made for this patient case. LIME determined that the diabetic classification for the patient used glucose BMI and age as classifiers with glucose having a weight of 0.29 and BMI having a weight of 0.12 and age having a weight of 0.08.

3.5 Evaluation Metrics

The trained logistic regression model for binary classification used standard evaluation metrics which included accuracy and precision and recall and F1-score and ROC-AUC as evaluation metrics for its performance assessment. Each predictive performance measure provides different information about how well the model predicts outcomes. The model achieved an accuracy rate of 84 percent which correctly predicted test outcomes 84 percent of the time. The model demonstrates strong prediction abilities because its results show high accuracy across different situations. The model predicted

86 percent of cases correctly as true positive results after the actual diabetic patients received their diagnosis. The model achieved 91 percent accuracy when predicting diabetic patients because it successfully identified most actual diabetes cases. The model successfully identified all patients who presented with high-risk conditions within this specific context. The F1 score establishes a performance index which connects both precision and recall results through an arithmetic relationship that assesses the model's performance across two dimensions. The area under the ROC curve indicates that the model is reasonably accurate in separating the diabetic from non-diabetic cases.

4. Global Feature Importance (SHAP Analysis)

For ease of interpretability, SHAP (SHapley Additive exPlanations) was used for the logistic regression model. From the SHAP summary plot we can see that Glucose, BMI, and Age are the most important features in the model for predicting diabetes, followed by Diabetes Pedigree Function and Insulin. High glucose and BMI especially contributed positively for predicting the diabetic class and lower glucose and insulin predicted the non-diabetic class, supporting what medical literature suggests, which improves the clinical credibility of the model.

On the instance level, SHAP force plots visualised the contribution of each feature to individual predictions. For example, with Glucose = 180, BMI = 33, and Age = 45 there was positive contribution towards the diabetic class prediction, while moderate blood pressure contributed negatively in terms of overall risk. The individual level basis for making a prediction is important for clinician to understand why a particular prediction was made.

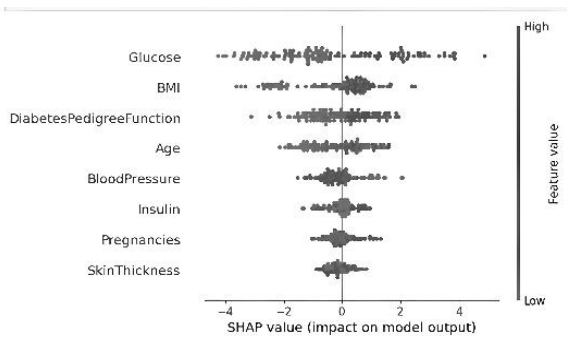

Figure 1: Shap Analysis

5. Local Explanations (LIME Analysis)

In addition to SHAP, we also included LIME (Local Interpretable Model-Agnostic Explanations) to provide intuitive explanations for specific instances. LIME perturbs input data around a selected patient case and trains a local surrogate model to show which features are most important.

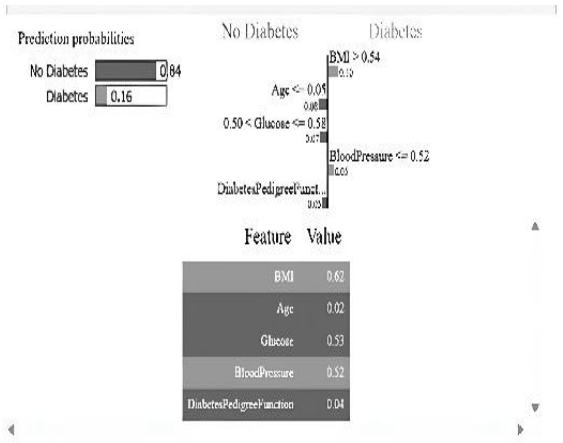

Figure 2: Lime Analysis

6. Comparative Insights: SHAP vs. LIME

Although both SHAP and LIME are intended for explainability, they are complementary in providing that in different ways. SHAP provides consistent, model-wide insights into contributions of features, allowing for reproducibility and fairness. LIME, on the other hand, is more appropriate for interpretability case-by-case as it provides clinicians with a relatively simple way to explain the mechanism behind the single prediction.

7. Deployment and Usability

The model and its explanations were more accessible by deploying it through a Streamlit application. The app's interface allows the user to input values for all clinical features using interactive sliders, and receive predicted values with SHAP and LIME explanations. The app also includes:

Visual outputs of the SHAP summary plots, incrementally increasing SHAP force plots and LIME features contribution graphs.

Finally, the user-based design considers the need to provide not just accurate predictions, but also interpretable and actionable predictions for clinicians, patients, and policymakers on the whole.

8. Findings

The findings of the current study suggest that adding explainable artificial intelligence (XAI) model methodology for diabetes prediction improves accuracy and explanation. In particular, a logistic regression predictive model was effective at identifying high-risk patients, which is reflected in the performance metrics of accuracy = 84%/precision = 86%/recall = 91% and F1= 0.89. These descriptive metrics provide assurance that the model has identified patients across the spectrum of the high-risk subtype of diabetes.

Importantly, SHAP analysis revealed glucose, BMI, age, followed closely by the diabetes pedigree function and insulin were the most important features influencing the outcomes of a predictive model. The results of visual analysis substantiated the SHAP findings, where diabetic patients had glucose (higher), BMI (higher), and number of patients over the age of 40 (greater prevalence). These measures rise with age. The correlation heatmap and radar charts also supported SHAP findings in that glucose and BMI had the strongest contextual relationships, for the outcomes of diabetes.

9. Discussion

This study demonstrates that the implementation of explainable artificial intelligence (XAI) techniques such as SHAP and LIME to the prediction of diabetes provides accurate predictions and an appropriate degree of interpretability. Although the machine learning models Random Forest, Decision Tree, and Naïve Bayes displayed comparable performance, Logistic Regression was found to be the best model for this work, balancing predictive accuracy (84%) and transparency. The two-layered explainability was valuable, as we captured a global understanding of diabetes prediction (SHAP), as well as reasoning behind individual cases (LIME). This is particularly valuable in healthcare environments where medical doctors require not only predictions but also justification for predictions from their modelling before taking further actions.

SHAP study also showed that Glucose, BMI, and Age were the best predictors of diabetes categorisation. LIME also made things more trustworthy by breaking down individual predictions, which led to explanations that were more relevant to each patient and far more useful. The study introduced several limitations: the small size of the dataset (PIMA Indians Diabetes dataset) and its predominantly population-specific nature restrict generalisability. Moreover, it is important to note that Logistic

10. Implications

The findings from this study have significant implications for the healthcare use of artificial intelligence. This approach to the explainable model upholds clinical transparency by offering the ability to understand the reasoning behind simple predictive procedures rather than relying on a "black-box" system. This type of explanation and interpretability means a clinician can substantiate the reported results and communicate that clearly to the patient, thereby improving accountability and establishing better trust with patients.

Therefore, in order to provide quick and precise diabetes risk assessments, healthcare organizations could incorporate such models into diagnostic tools or telemedicine platforms.

11. Conclusion

What we have presented in this research is a machine learning framework for diabetes prediction that contains an extra layer of explainability by introducing SHAP and LIME explanations in the methods pipeline. A Logistic Regression model fit on the PIMA Indians Diabetes dataset had a strong predictive performance (84% accuracy and a high recall value equal to 91%) showing it is a good candidate for the early screening of diabetes. Though accuracy is important, the greatest contribution of this work is in the details of explainability: SHAP provided insight into the global feature importance across the dataset while LIME provided more intended and meaningful local explanations for the individual patient. Integrating these two explainability methods creates a meaningful relationship between a strong predictive performance and the usability in clinical practice as it creates a foundation of transparency, accountability, and trust for AI-based healthcare solutions.

References

Federation, I. D. (2021). IDF Diabetes Atlas Brussels. *Belgium: international diabetes federation*.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model

predictions. *Advances in neural information processing systems*, *30*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., & Johannes, R. S. (1988, November). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the annual symposium on computer application in medical care* (p. 261).

***