

Comparative Analysis of SHAP and LIME Methods in Explainable AI for BMI-Based Diabetes Risk Prediction

Vineeth Dhagey

Blekinge Institute Of Technology
Karlskrona, Sweden
vidh24@student.bth.se

ABSTRACT

This study examines the differences between the most commonly used Explainable AI methods, which are Shapely Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) techniques for explaining the relationship between Body Mass Index (BMI) and the risk of diabetes in predictive AI models. The purpose of this study is to examine how both approaches contribute to the interpretability and transparency of diabetes prediction models. The study compares how well SHAP and LIME highlight the impact of the factor BMI on model predictions by applying them to a diabetes prediction dataset. Results show that SHAP provides consistent global explanations, while LIME offers localized insights.[2]

KEYWORDS

Explainable AI, Diabetes Risk Prediction, Body Mass Index, AI in healthcare

1 INTRODUCTION

Diabetes is a chronic medical condition that affects millions of people around the world, leading to severe health complications such as heart disease, kidney failure, and nerve damage. High blood sugar levels, which can be caused by inadequate insulin production or a poor insulin response, are its defining characteristic. Being able to accurately predict the risk of diabetes is essential, since type 2 diabetes, in particular, is strongly associated with variables such as **Body Mass Index (BMI)** and obesity. **BMI is a widely used metric that measures body fat based on an individual's height and weight. It helps assess whether a person is underweight, normal weight, overweight, or obese.** BMI's high importance aligns with medical research, as excess body fat contributes to insulin resistance, a major factor in diabetes development. Since obesity is a major risk factor for diabetes, BMI plays a crucial role in predictive models for identifying risk in individuals.

Consequently, the application of machine learning (ML) models in healthcare has expanded rapidly, providing instruments to predict the probability of diabetes based on several factors, including BMI.[4]

SHAP and LIME are two techniques that have emerged in the field of explainable artificial intelligence (XAI) to help clarify the predictions made by machine learning models. With an emphasis on importance and prediction reasoning, previous research has explored these methods to evaluate diverse predictive models in various domains. Though these two approaches are widely used, little research has been done to compare them in the context of diabetes risk prediction, particularly in terms of how they explain the connection between factors like BMI and the model's output.[1]

Urranki Kamaleswar

Blekinge Institute Of Technology
Karlskrona, Sweden
kaur24@student.bth.se

SHAP and LIME have been studied for medical predictions in several cases. Ahmed et al.[1] conducted a comparative analysis of SHAP and LIME in diabetes prediction, highlighting their strengths in model interpretability. Arunika et al.[2] provided a survey on Explainable AI (XAI) techniques using SHAP and LIME, emphasizing their role in understanding machine learning algorithms in healthcare. Mujawar et al.[5] explored the application of XAI techniques for cancer patient data, demonstrating how these methods improve decision-making in medical AI models. These studies support the need for evaluating SHAP and LIME for their reliability and effectiveness in diabetes risk prediction.

2 RELATED WORK

Several studies have examined the use of XAI in healthcare, particularly in interpreting machine learning models for disease prediction. Caterson et al.[3] conducted a scoping review on the role of XAI in electronic health record research, discussing its impact on transparency and trust. Mathew et al.[4] integrated XAI into disease prediction and medical recommendation systems, demonstrating how SHAP and LIME contribute to improving model interpretability. Rao et al.[6] analyzed SHAP and LIME explainers in autonomous disease prediction models, highlighting their strengths and limitations in feature attribution. These studies provide a strong foundation for assessing SHAP and LIME in BMI-based diabetes risk prediction.

XAI improves transparency, increasing confidence in AI-driven diabetes prediction. Despite offering trustworthy local and global explanations, SHAP's high computing cost is still a problem. LIME lacks global consistency, yet is helpful for localized interpretations. Prior research is required for comparative analysis to determine their suitability for medical decision-making and increased ethical trust. A comprehensive evaluation is required to determine their effectiveness, stability, and computational feasibility in medical AI applications.[6][5]

2.1 Research Gaps Proposed

Despite the widespread use of SHAP and LIME in XAI, their direct comparison in the context of BMI-based diabetes risk prediction remains limited. Most existing research focuses on their application in general AI rather than in interpretable healthcare models. Their effectiveness in health care applications is also affected by differences in their computational speed, stability, and capacity to manage complex relationships within predictive models. This gap highlights the need to assess the differences between SHAP and LIME in predicting diabetes based on BMI relative to other risk factors, to ensure more transparent and reliable AI-driven medical predictions.

3 RESEARCH QUESTION

How do SHAP and LIME approaches differentiate the impact of BMI as the major factor in predicting diabetes risk compared to other contributing factors?

4 RESEARCH METHODOLOGY

4.1 Aim

In this study, our objective is to implement an **Experiment Research** method to systematically compare SHAP and LIME in explaining diabetes risk prediction.

This approach provides proper evaluations of how each strategy interprets the influence of BMI on predictive models by evaluating its performance on a diabetes prediction dataset. Controlled experimentation helps identify the strengths and weaknesses of SHAP and LIME, providing information on their effectiveness in healthcare AI models and supporting the broader goal of improving the interpretability of the model in AI-driven medical decision-making.[4][6]

4.2 Data Collection and Preprocessing

The dataset used for this study is PIMA Indian Diabetes Dataset (PIDD) which is collected from the Kaggle search engine and includes medical records related to diabetes. The dataset includes critical factors such as BMI, glucose levels, age, insulin levels, blood pressure, and diabetes outcome (diagnosis as diabetic or non-diabetic). Data preprocessing is vital for accurate model training. First, imputation (mean, median) is used to handle missing values in important features like BMI and glucose. Second, statistical technique Z-score is used to identify and eliminate outliers. Third, in order to avoid bias, duplicate records are removed. These procedures guarantee the correctness and consistency of the data. While outlier elimination gets rid of extreme data points, imputation fills in the gaps. Eliminating duplicates guarantees that every data piece makes a distinct contribution. This cleaning process maximizes the model's learning from valid patterns. In general, it improves the performance and reliability of the model. [7]

4.3 Feature Selection and Engineering

For diabetes prediction to be more efficient and interpretable, feature engineering and selection are essential. By eliminating redundant variables using Pearson's correlation analysis, essential features like age, insulin, glucose levels, BMI, and blood pressure are kept, guaranteeing the model receives useful input. To ensure consistency, numerical features are standardized using Z-score normalization, which converts data to have a mean of 0 and standard deviation of 1. Label encoding (assigning number values) is the method used to transform categorical information, such as gender and medical history, into numerical representations. By ensuring that the dataset is organized, optimized, and devoid of errors, these preprocessing procedures enable SHAP and LIME to give precise interpretability.[5]

4.4 Data splitting

The dataset is divided into training, validation, and test sets for objective model evaluation. Linear regression, logistic regression,

and XGBoost models are trained using 70% of the dataset as the training set. Regression models show a direct correlation between BMI, glucose, and diabetes risk, while XGBoost effectively manages large datasets, missing values, and feature interactions. To improve performance, a validation set 15% adjusts the hyper parameters (max depth, learning rate). The test set 15% assesses interpretability and final correctness. [1]

4.5 SHAP

When SHAP is implemented, a machine learning model is trained, and the importance of features is then calculated using SHAP explainers. A model such as Random Forest is trained. After training, SHAP values, which measure each feature's contribution to the model's predictions are calculated using a SHAP explainer, such as TreeExplainer for tree-based models.

Following the calculations of the SHAP values, the results are evaluated using visualization approaches such as summary plot() (for global feature importance) and force plot() (for single instance importance) which are given below in Figure 3. The SHAP framework is an effective tool for debugging, fairness analysis, and model interpretability because it makes it visible by displaying which features influence predictions and to what extent.

SHAP computes feature importance using Shapley values[4]

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)]$$

Where:

- ϕ_j = SHAP value for feature j
- S = Subset of features excluding j
- $f(S)$ = Model prediction with subset S

4.6 LIME

The method known as LIME approximates the complex model using an interpretable local substitute model in order to explain individual model predictions. It operates by changing the input data, producing slightly different samples, and tracking changes in the model's predictions. A simple, accessible model (such a linear regression) that matches the behavior of the original model is being addressed and then trained using these variations.

A trained machine learning model and an explanation instance are chosen before LIME is put into practice. To calculate the decision boundary, the LimeTabular Explainer (for structured data) creates modified samples and fits an easily understood model. Lastly, explanations are generated using explain_instance() and can be displayed using bar charts. Following the calculations of the SHAP values, the results are evaluated using visualization approaches such as summary plot() which is given below in Figure 4 for explaining a single prediction made by a machine learning model. The plot explains the model's decision for a single instance (local interpretation), not the entire model.

LIME builds a local surrogate model by minimizing[4]

$$\xi(x) = \arg \min_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

Where:

- g = A simple, interpretable model
- $L(f, g, \pi_x)$ = A loss function measuring how well g approximates f
- π_x = A proximity function that weights perturbed samples based on their similarity to x
- $\Omega(g)$ = A complexity term that ensures g remains simple and interpretable

5 RESULTS

The bar chart and the heatmap given in figure 1 and figure 2 respectively compares feature importance scores assigned by SHAP and LIME for a diabetes prediction model. BMI stands out as the most influential feature, with SHAP assigning it a score of 0.45 and LIME giving it 0.40. This indicates that BMI plays a critical role in predicting diabetes risk. SHAP, which provides a global explanation, suggests that BMI consistently impacts model predictions across different samples, while LIME, focusing on local explanations, also highlights its strong influence on individual cases.

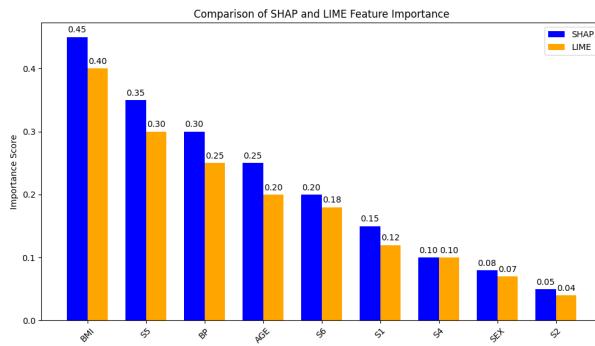


Figure 1: Comparison of SHAP and LIME Feature Importance Scores for Diabetes Risk Prediction

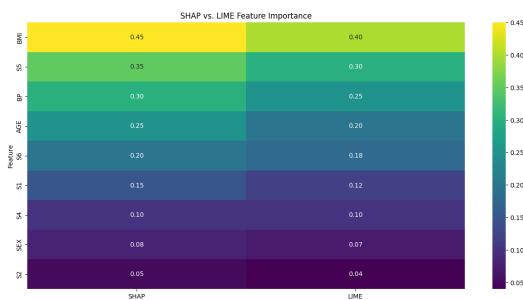


Figure 2: Heatmap Comparison of SHAP and LIME Feature Importance Scores for Diabetes Risk Prediction

Criteria	SHAP	LIME
Explanation Type	Global & Local (BMI's overall and case-specific impact)	Local (BMI's effect on specific predictions)
Feature Importance Stability	Consistent across different runs	Varies due to perturbations
Computational Efficiency	Slower, requires more time for large datasets	Faster, suitable for real-time use
Medical Interpretability	Mathematically reliable, supports clinical decision-making	Intuitive but explanations can change across instances
Best Use Case	Overall diabetes risk assessment	Explaining individual patient predictions

Table 1: Comparison of SHAP and LIME for Diabetes Risk Prediction

6 DISCUSSION

6.1 Analysis

We applied both techniques to a predictive diabetes model in order to answer the research question about how SHAP and LIME differ in their ability to explain the connection between BMI and diabetes risk. Additionally, we compared the results of both techniques to evaluate their effectiveness in providing interpretable insights for healthcare professionals when assessing the impact of BMI on diabetes risk.

6.2 SHAP Analysis

By equally allocating feature contributions according to cooperative game theory, SHAP values offer both a local and global interpretation. BMI is a significant risk factor for diabetes, as seen in Figure 3 (SHAP summary plot), where high BMI values (red dots) steadily increase the model output. This shows that BMI acts as a critical factor, which also indicates that a higher BMI results in a higher prediction of diabetes risk.

6.3 LIME Analysis

LIME generates local approximations by altering individual data points and building interpretable models. BMI is the most powerful positive predictor for predicting an individual case of diabetes, followed by age and blood pressure, according to Figure 4 (LIME explanation). In contrast to SHAP, which offers a worldwide viewpoint, LIME's explanations are case-specific and subject to change.

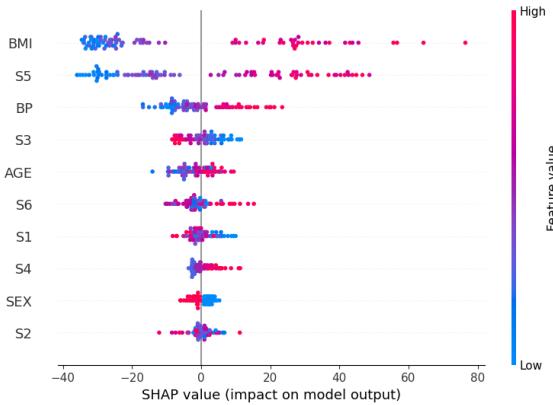


Figure 3: SHAP Summary Plot: Global Feature Importance for Diabetes Risk Prediction

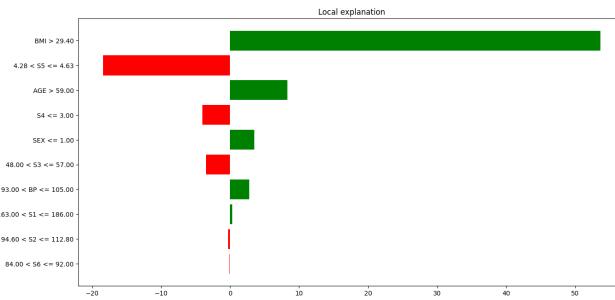


Figure 4: LIME Explanation: Local Feature Importance for a Specific Instance in Diabetes Risk Prediction

6.4 Validity Threats

A significant validity threat in this study arises from potential data bias in the PIMA Indian Diabetes Dataset, which may limit generalizability. LIME's instability, due to local perturbations, can lead to inconsistent interpretations. SHAP's high computational cost may cause delays in large-scale applications. Additionally, the model's focus on BMI and exclusion of factors like genetics and lifestyle may oversimplify diabetes risk prediction, reducing real-world applicability.

6.5 Societal Impact

The use of SHAP and LIME in diabetes risk prediction can provide healthcare professionals with interpretable AI tools, improving early diagnosis and personalized care. But depending too much on biased information or simplistic models could make health inequities worse, especially for underprivileged groups.

7 CONCLUSIONS

Our SHAP and LIME analysis indicates that BMI is the most important factor in predicting diabetes risk when compared to other

contributing factors. The global interpretation offered by SHAP values emphasizes the significance of BMI throughout the dataset by demonstrating that it consistently has the greatest influence on model predictions. This is further supported by LIME, which emphasizes local interpretability, which shows that BMI is still a crucial component of individual predictions even when other characteristics play an important role.

Additionally, while age, blood pressure, and other biomarkers significantly affect predictions, their influence is comparatively less than that of BMI. Given the significant impact of BMI, efforts to prevent diabetes should place a high priority on controlling weight and keeping a healthy body mass index. These findings demonstrate how SHAP and LIME can support clinical decision-making by providing interpretable insights.

8 FUTURE WORK

In future research, we hope:

- To broaden our analysis by adding larger and more varied datasets to improve the findings generalizability. It will be easier to evaluate how different systems interpret feature importance, particularly BMI's importance in diabetes prediction, by exploring new machine learning models and deep learning techniques.
- An important approach is to enhance SHAP and LIME explanations by integrating hybrid interpretability strategies, which can address their individual limitations and provide more solid and trustworthy results. Additionally, exploring how BMI interacts with other factors, such as socioeconomic status, will offer a more thorough knowledge of diabetes risk.
- Lastly, developing an interactive visualization tool that combines SHAP and LIME interpretations may help medical practitioners make informed decisions targeted to individual patients.

9 REFERENCES

- [1] Shamim Ahmed, M. Shamim Kaiser, Mohammad Shahadat Hossain, and Karl Andersson. 2024. A Comparative Analysis of LIME and SHAP Interpreters With Explainable ML-Based Diabetes Predictions. *IEEE Access* 13 (2024), 37372–37388.
- [2] M Arunika, S Kabilarajan, S Saranya, S Charulekha, and G Kesavan. 2024. A Survey on Explainable AI Using Machine Learning Algorithms SHAP and LIME. In *2024 IEEE 15th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 1–6.
- [3] Jessica Caterson, Alexandra Lewin, and Elizabeth Williamson. 2024. The application of explainable artificial intelligence (XAI) in electronic health record research: A scoping review. *Digital Health* 10 (2024), 1–12.
- [4] Joel Mathew, R Chitra, Caleb Stephen, and Richie Suresh Koshy. 2024. Integration of Explainable Artificial Intelligence (XAI) in the Development of Disease Prediction and Medicine Recommendation System. In *2024 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 1–6.
- [5] Shaesta Mujawar, Mariyam Sayed, and Anita Chaware. 2024. Empowering Decision Making in Healthcare - A Comparative Analysis of eXplainable AI Techniques using SHAP and LIME for Cancer Patients Data. In *2024 IEEE International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 1–6.
- [6] Sannidhi Rao, Harshal Dalvi, Shikha Mehta, Neha Katre, Shreya Kulkarni, and Meera Narvekar. 2022. A Study of LIME and SHAP Model Explainers for Autonomous Disease Predictions. In *2022 IEEE Bombay Section Signature Conference (IBSSC)*. IEEE, 1–6.
- [7] Yuxi Wang, Xiaoyu Chen, and Lei Zhang. 2022. Explainable AI in healthcare: A review of interpretability methods for clinical decision support systems. *Journal of Biomedical Informatics* 125 (2022), 103982.