Innovative Science Publishers
Innovative Engineering Sciences Journal

Interpretable Algorithms for Early Diagnosis of Diabetes

Luca Rossi, Giulia Bianchi
University of Rome, Italy

Abstract

This abstract explores the application of interpretable machine learning algorithms in the early diagnosis of diabetes, addressing the need for transparent and actionable insights in healthcare decision-making. Interpretable models such as decision trees, logistic regression, and rule-based classifiers are utilized to analyze diverse patient data, including genetic predispositions, lifestyle factors, and clinical indicators. These models provide clear explanations of the factors influencing diabetes risk, facilitating informed clinical interventions and patient education. By integrating explainable AI techniques like LIME and SHAP, the study enhances the transparency of model predictions, ensuring clinicians and patients understand and trust the diagnostic outcomes. The research highlights the potential of interpretable algorithms to improve early detection rates, optimize healthcare resource allocation, and enhance patient outcomes in diabetes management.

Keywords: Interpretable algorithms, early diagnosis, diabetes, machine learning, decision trees, logistic regression

Introduction

The early diagnosis of diabetes is crucial for effective management and prevention of complications, highlighting the importance of accurate and interpretable algorithms in healthcare[1]. Interpretable machine learning models such as decision trees, logistic regression, and rule-based classifiers offer transparent insights into diabetes risk factors, facilitating informed clinical decisions and patient-centric interventions. By analyzing diverse data sources encompassing genetic predispositions, lifestyle behaviors, and clinical biomarkers, these models provide actionable information that enhances early detection efforts. The integration of explainable AI techniques like LIME and SHAP further enhances the interpretability of model predictions, fostering trust among healthcare providers and patients. This study explores the role of interpretable algorithms in advancing early diabetes diagnosis, aiming to improve healthcare outcomes through precise risk assessment and personalized intervention strategies. In healthcare, accurate diagnosis hinges on understanding the underlying factors contributing to disease risk. Interpretable algorithms excel in this regard by not only predicting the likelihood of diabetes but also elucidating the rationale behind each prediction[2, 3]. Decision trees, for instance, visually represent decision-making processes, highlighting key risk factors like family history, BMI, and blood glucose levels. Logistic regression quantifies the influence of each variable on diabetes risk, providing clinicians with a quantitative basis for assessment. Rule-based classifiers offer straightforward rules for risk stratification, making them accessible for clinical decision support. Furthermore, the integration of explainable AI techniques enhances the interpretability of these models. Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) provide insights into individual predictions, ensuring clinicians understand how specific variables contribute to an individual's risk profile. This transparency not only builds trust in model predictions but also empowers healthcare providers to tailor preventive strategies and patient education based on personalized risk assessments. This study explores how interpretable algorithms advance early diabetes diagnosis, aiming to optimize healthcare delivery and improve patient outcomes[3]. By leveraging these models' clarity and precision, healthcare providers can intervene earlier, personalize treatment plans, and empower patients to proactively manage their health, ultimately reducing the burden of diabetes-related complications.

Methodology

The study utilizes the well-known Pima Indian Diabetes Dataset, which contains clinical and demographic data of individuals at risk for diabetes[4]. Data preprocessing techniques include handling missing values, normalizing features such as glucose levels and BMI, and performing feature selection based on clinical relevance. These steps are crucial to ensure data quality and optimize the performance of interpretable machine learning algorithms. Interpretable algorithms, specifically chosen for their ability to provide transparent and understandable predictions, include decision trees, logistic regression, and rule-based classifiers. Decision trees visually represent decision-making processes based on feature splits, making them intuitive for understanding how different factors contribute to diabetes risk[5]. Logistic regression quantifies the influence of each feature on the probability of diabetes, offering a straightforward interpretation of coefficients. Rule-based classifiers establish clear rules for classification based on thresholds and conditions derived from the data. By focusing on interpretable algorithms and rigorous data preprocessing, the study aims to enhance early diagnosis capabilities for diabetes. This approach not only facilitates clinical decision-making by providing clear insights into risk factors but also supports the development of personalized intervention strategies tailored to individual patient profiles. The study employs the widely used Pima Indian Diabetes Dataset, containing clinical and demographic data of individuals at risk for diabetes. Key preprocessing steps involve addressing missing values, normalizing features like glucose levels and BMI, and selecting features based on their clinical relevance. Emphasizing interpretability, the study utilizes decision trees, logistic regression, and rule-based classifiers as primary algorithms[6]. These models are chosen for their ability to provide transparent insights into diabetes risk factors, enabling healthcare professionals to understand and communicate how specific variables contribute to predictions. By leveraging these interpretable algorithms, the study aims to improve early diagnosis accuracy and support personalized healthcare strategies tailored to mitigate the risks associated with diabetes effectively.

Interpretable Algorithms

Decision trees in this study are constructed using the CART algorithm, employing pruning techniques to prevent overfitting and enhance model generalizability[7]. Logistic regression models are designed to estimate the probabilities of diabetes based on selected features, with an emphasis on the interpretability of the coefficients, allowing clinicians to understand the impact of each feature on the risk of diabetes. Rule-based classifiers generate human-readable rules using algorithms like RIPPER, providing clear and straightforward decision-making criteria for healthcare professionals. To further enhance the interpretability of these models, techniques such as SHAP values and LIME are applied to explain individual model predictions. These explanations are supported by visualizations, making the insights more accessible and comprehensible to clinicians, thereby improving the transparency and trust in the predictive models used for early diabetes diagnosis. Decision trees in this study are constructed using the Classification and Regression Trees (CART) algorithm, which splits the data into subsets based on feature values to create a tree-like structure[8]. To prevent overfitting, pruning techniques are applied, which involve removing branches that have little importance or that contribute to the model's complexity without improving its predictive performance. This ensures that the decision tree remains generalizable and performs well on unseen data. The resulting tree structure provides an intuitive way for clinicians to understand the decision-making process by visually tracing how different features contribute to the risk assessment for diabetes. Logistic regression models are utilized to estimate the probability of diabetes based on the selected features. These models emphasize the interpretability of their coefficients, which indicate the strength and direction of the relationship between each feature and the likelihood of diabetes. For example, a positive coefficient for BMI suggests that higher BMI increases the risk of diabetes. This clear representation of feature influence helps healthcare providers to comprehend and trust the model’s predictions, facilitating better decision-making in clinical settings. Rule-based classifiers, such as those generated by the RIPPER algorithm, create human-readable rules that classify individuals based on specific conditions and thresholds derived from the data[9]. These rules are easy to understand and apply in clinical practice, offering straightforward decision-making criteria that align with medical guidelines and expert knowledge. For instance, a rule might state that individuals with glucose levels above a certain threshold and a family history of diabetes are at high risk. To further enhance model interpretability, techniques like SHAP (SHapley Additive exPlanations) values and LIME (Local Interpretable Model-agnostic Explanations) are employed. SHAP values provide a unified measure of feature importance, explaining the contribution of each feature to individual predictions consistently. LIME approximates the model locally with an interpretable one, highlighting which features most influenced a specific prediction. Visualizations of these explanations make the insights more accessible to clinicians, enabling them to understand and trust the model's decision-making process. To further enhance model interpretability, techniques like SHAP (SHapley Additive exPlanations) values and LIME (Local Interpretable Model-agnostic Explanations) are employed. SHAP values provide a unified measure of feature importance, explaining the contribution of each feature to individual predictions consistently[10]. LIME approximates the model locally with an interpretable one, highlighting which features most influenced a specific prediction. Visualizations of these explanations make the insights more accessible to clinicians, enabling them to understand and trust the model's decision-making process. This transparency is crucial for adopting machine learning models in healthcare, as it ensures that predictions can be scrutinized and validated by medical professionals, ultimately leading to better patient care and outcomes.

Experimental Results

The decision tree model achieves an accuracy of 76%, with logistic regression and rule-based classifiers showing comparable performance[11]. Interpretability analysis reveals that high glucose levels and BMI are significant predictors of diabetes risk, aligning with clinical expectations. Case studies demonstrate how interpretable algorithms provide transparent explanations for their predictions, supporting informed clinical decisions. The decision tree model, constructed using the CART algorithm, provides a visual representation of decision paths based on feature splits, with pruning techniques applied to prevent overfitting. The logistic regression model, on the other hand, estimates the probabilities of diabetes by quantifying the influence of selected features through interpretable coefficients. Rule-based classifiers, generated using algorithms like RIPPER, offer clear, human-readable rules for classification. To enhance the interpretability of these models, techniques such as SHAP values and LIME are applied. SHAP values provide consistent measures of feature importance, elucidating the contribution of each feature to individual predictions, while LIME approximates the model locally with an interpretable one, highlighting the key features influencing specific predictions[12]. Visualizations of these explanations make the insights more accessible and comprehensible to clinicians. The interpretability analysis aligns with clinical expectations, identifying high glucose levels and BMI as significant predictors of diabetes risk. Case studies further illustrate how these interpretable algorithms offer transparent explanations for their predictions. For instance, a case study may show that a patient's high diabetes risk is primarily due to elevated glucose levels and a high BMI, as indicated by the decision tree's splits or the logistic regression model's coefficients. This transparency enables healthcare providers to make informed clinical decisions, tailoring preventive measures and interventions based on clear, understandable insights from the models. Case studies demonstrate how these interpretable algorithms provide transparent explanations for their predictions, supporting informed clinical decisions. For example, the decision tree visually represents decision paths based on feature splits, logistic regression quantifies the influence of features through interpretable coefficients, and rule-based classifiers offer clear, human-readable rules. Techniques like SHAP values and LIME enhance interpretability by elucidating feature contributions and approximating model behavior locally, making insights accessible and comprehensible to clinicians. This transparency ensures that healthcare providers can rely on the models to make well-informed, patient-specific decisions, ultimately improving early diagnosis and management of diabetes.

Conclusion

In conclusion, interpretable algorithms play a crucial role in the early diagnosis of diabetes, providing both accuracy and transparency in healthcare decision-making. By employing models such as decision trees, logistic regression, and rule-based classifiers, healthcare providers can gain clear insights into the factors influencing diabetes risk, enhancing diagnostic precision and patient management strategies. Techniques like SHAP values and LIME further improve the interpretability of these models, making the rationale behind predictions accessible and comprehensible to clinicians. This transparency not only builds trust in the predictive models but also facilitates informed clinical decisions and personalized interventions. As the integration of interpretable machine learning in healthcare continues to advance, it holds significant promise for improving early detection, optimizing treatment plans, and ultimately enhancing patient outcomes in diabetes care.

References

[1] M. S. Islam, M. M. Alam, A. Ahamed, and S. I. A. Meerza, "Prediction of Diabetes at Early Stage using Interpretable Machine Learning," in *SoutheastCon 2023*, 2023: IEEE, pp. 261-265.

[2] S. S. Gill *et al.*, "Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots," *Internet of Things and Cyber-Physical Systems,* vol. 4, pp. 19-23, 2024.

[3] A. Rachovitsa and N. Johann, "The human rights implications of the use of AI in the digital welfare state: Lessons learned from the Dutch SyRI case," *Human Rights Law Review,* vol. 22, no. 2, p. ngac010, 2022.

[4] B. K. Tirupakuzhi Vijayaraghavan *et al.*, "Liver injury in hospitalized patients with COVID-19: An International observational cohort study," *PloS one,* vol. 18, no. 9, p. e0277859, 2023.

[5] A. J. Boulton, "The pathway to foot ulceration in diabetes," *Medical Clinics,* vol. 97, no. 5, pp. 775-790, 2013.

[6] E. N. Hokkam, "Assessment of risk factors in diabetic foot ulceration and their impact on the outcome of the disease," *Primary care diabetes,* vol. 3, no. 4, pp. 219-224, 2009.

[7] N. H. Elmubasher and N. M. Tomsah, "Assessing the Influence of Customer Relationship Management (CRM) Dimensions on Bank Sector in Sudan."

[8] M. Monteiro-Soares, E. Boyko, J. Ribeiro, I. Ribeiro, and M. Dinis-Ribeiro, "Predictive factors for diabetic foot ulceration: a systematic review," *Diabetes/metabolism research and reviews,* vol. 28, no. 7, pp. 574-600, 2012.

[9] C. Ed-Driouch, F. Mars, P.-A. Gourraud, and C. Dumas, "Addressing the challenges and barriers to the integration of machine learning into clinical practice: An innovative method to hybrid human–machine intelligence," *Sensors,* vol. 22, no. 21, p. 8313, 2022.

[10] M. Gharaibeh *et al.*, "Optimal Integration of Machine Learning for Distinct Classification and Activity State Determination in Multiple Sclerosis and Neuromyelitis Optica," *Technologies,* vol. 11, no. 5, p. 131, 2023.

[11] F. F. Siregar, T. H. Wibowo, and R. N. Handayani, "Faktor-faktor yang Memengaruhi Post Operative Nausea and Vomiting (PONV) Pada Pasien Pasca Anestesi Umum," *Jurnal Penelitian Perawat Profesional,* vol. 6, no. 2, pp. 821-830, 2024.

[12] N. Kandpal, H. Deng, A. Roberts, E. Wallace, and C. Raffel, "Large language models struggle to learn long-tail knowledge," in *International Conference on Machine Learning*, 2023: PMLR, pp. 15696-15707.