XGBoost-Driven Diabetes Prediction using Permutation Feature Importance for Model Interpretability

Chinwe Miracle Chituru1, a), Sin-Ban Ho1, b), Ian Chai1, c), and Chuie-Hong Tan2, d)

1Faculty of Computing and Informatics, Multimedia University, 63100 Cyberjaya, Malaysia.

2 Faculty of Management, Multimedia University, 63100 Cyberjaya, Malaysia.

b) Corresponding author: sbho@mmu.edu.my

*a) reun.love36@gmail.com  
c) ianchai@mmu.edu.my  
d) chtan@mmu.edu.my*

**Abstract.** Diabetes mellitus is a chronic metabolic disorder that has become a recurring global health crisis. Early diagnosis and accurate risk prediction require advanced analytical methods. Machine learning techniques, especially gradient-boosting methods like Extreme Gradient Boosting (XGBoost), have shown improved performance in predicting the diabetes risk from various clinical features. This paper applies classification techniques using XGBoost for prediction. It models relationships in a dataset with moderate number of features. For model interpretability and actionable insights, permutation feature importance is used to provide the relative contribution of each feature toward the model’s robust predictions. Unlike traditional methods that rely on model weights or coefficients, permutation feature importance quantifies the impact of feature shuffling on model performance, offering a more reliable and model-agnostic explanation of feature impact. The model performance metrics are evaluated before and after the feature shuffling, which is performed by random permutation. The results highlight the crucial predictors of diabetes risk, which can assist in early intervention by identifying key health factors. This paper highlights how the different features interact, providing a deeper understanding of the predictive power of various clinical features. The results reveal XGBoost and permutation importance as promising tools for health providers to promote evidence-based decisions on diabetes management and intervention strategies.

# INTRODUCTION

Diseases affect global society, and diabetes is one of the most concerning chronic health conditions, with millions of people diagnosed yearly [1]. Early detection and prompt intervention help prevent the complications that arise, such as cardiovascular diseases, diabetic neuropathy, and nephropathy. Recently, Machine Learning (ML) techniques have become well-known in the healthcare sector for their ability to analyze datasets from small to large and provide accurate predictions for diabetes disease. These techniques range from deep learning, particularly neural networks, to ensemble learning. XGBoost (Extreme Gradient Boosting), an ensemble learning method, performs outstandingly in multiple tasks owing to its excellent predictive power and applicability across domains such as medical diagnostics.

Although XGBoost has been highly effective in diabetes risk prediction [2], model interpretability is one of the notable challenges. Understanding the intricacies behind the predictions and the influential features. XGBoost, a complex machine learning model, functions as a “closed box” model that is difficult to interpret, hence making it hard for healthcare providers to understand why certain predictions were made. Lack of transparency will always hamper this trust and inhibit these models from getting wide acceptance in clinical settings where explanation is required for decision-making. Therefore, there is a growing demand for models that perform well and integrate with model interpretability [3] enhancing insight into why the model made any decision.

Permutation Feature Importance (PFI) is a widely used method for feature interpretation. It measures the increase in model prediction error when a feature is randomly permuted. It indicates which features most contribute to predicting diabetes through randomization or shuffling to measure feature importance. However, integration of the PFI technique into complex models such as Extreme Gradient Boosting is not comprehensively evaluated, and the potential to enhance interpretability in the case of diabetes prediction is still open for analysis. Moreover, the interaction of feature importance with many underlying data characteristics, such as class imbalance and feature correlation, remains under-researched.

This study aims to bridge these research gaps by proposing a model interpretability framework employing PFI with XGBoost for the prediction. This research will examine how PFI can be utilized to illustrate the features that influence the prediction. It explores the relationships between the health parameters of the dataset and the diabetes prediction, providing clearer insight into the model’s decision-making process. Furthermore, it will investigate how integrating PFI into the XGBoost model makes its real-world application more useful in healthcare settings, such that the predictions are correct and understandable to medical professionals. The findings of this study aim to bridge the gap between high-performing ML algorithms and the need for transparency in health systems.

# LITERATURE REVIEW

According to Chituru [4], the integration of computing in healthcare provides significant improvements in diagnostic accuracy and personalized treatment by leveraging data. This spatial integration allows for a more comprehensive understanding of a patient’s condition, which can subsequently lead to better-informed medical decisions. To build on this, the application of the XGBoost ML model in medical diagnostics can be used for diabetes prediction, and PFI can be used to rank features based on their influence on model predictions. Ensemble Methods are a class of ML algorithms that combine multiple models [5] for improved predictive performance, such as the XGBoost. Synthetic Minority Over-sampling Technique (SMOTE) is used to balance the instances across all the classes [6].

XGBoost is a scalable ML algorithm based on gradient-boosted decision trees. It iteratively builds decision trees by correcting mistakes of the previous predictor model to produce a final and accurate prediction. XGBoost was built based on gradient boosting machines with a high-performance boosting technique to minimize the loss function [7]. To evaluate the XGBoost model, the accuracy metric is the proportion of correct predictions, i.e. True Positives (TP) plus True Negatives (TN) out of all predictions. Precision measures the proportion of TP among all predicted positives. Recall (sensitivity) quantifies the proportion of actual positives correctly identified, and F1-Score is the harmonic mean of precision and recall, providing a balanced measure, particularly useful when dealing with imbalanced class distributions. According to Gündoğdu [8], in the efficient prediction of early-stage diabetes using the XGBoost classifier with the Random Forest (RF) feature selection technique on 520 samples, XGBoost and RF outperform the K-Nearest Neighbors (KNN) and Gradient Boosting algorithms, showing better and more consistent TP and TN.

The data preprocessing step involves preparing raw data for analysis and model training by cleaning, encoding, filtering, and transforming it. Data preprocessing is essential for reliable model classification [9]. Palanivinayagam proposed a method using the Support Vector Machine (SVM) classiﬁer to recover the missing values and the false classiﬁcation [6] To ensure equal class distribution, data division was implemented [7]. KNN and SVM are superior methods for filling in missing values, achieving an accuracy of 98.49% and 94.89%, respectively [10].

PFI serves as a global Explainable Artificial Intelligence (XAI) method that measures each feature’s impact on the model’s performance. Permutation is an established feature importance method in ML and software engineering communities [11], originally introduced by Breiman [12] for ranking features in a random forest classifier. PFI is often used as an interpretation method [13] to describe the relationship between input features and the model’s prediction. It works by shuffling the values of one feature across all instances while the other features in the dataset remain intact. For example, for a dataset with five columns (features), the data of one column is randomly permuted. The goal is to see how much the model’s performance drops when a feature is randomized, and this indicates its importance.

PFI is model agnostic and is carried out after the model has been trained. In comparison with alternative models, it evaluates the feature impact on the overall model’s performance, unlike the Local Interpretable Model-agnostic Explanations (LIME), which work on an instance level. SHapley Additive exPlanations (SHAP) uses game theory for feature attributions, while PFI assesses impact through performance change resulting from feature permutation.

There are connections between permutation importance estimates for a prediction model and linear model coefficients [14]. Both LIME and PFI aim to explain the model, but LIME focuses on local interpretability, PFI focuses on global, and SHAP provides both local and global explanations. Despite these strengths, benchmarking against similar works reveals that XGBoost with PFI provides comparable or superior performance, particularly in terms of stability and interpretability when compared to other techniques. Charts and plots are used for the visualization of the results from permutation, which is important for easier comprehension. The Mean Absolute Error (MAE) metric is used to quantify the prediction error, and for this study, it was computed on the permuted features.

# RESEARCH METHODOLOGY

The improved accuracy for the XGBoost model trained on a clean dataset is a necessity for it to be applied in medical systems. Data collection for the diabetes health indicator dataset was obtained from the 2021 Behavioral Risk Factor Surveillance System (BRFSS) exercise. The Centers for Disease Control and Prevention (CDC) funded the creation of the dataset to provide insights into the health status and lifestyle, and their correlation to the diagnosis of diabetes. A clean dataset of the 236,378 exercise responses to the CDC’s BRFSS2021 was updated by Kaggle’s collaborator [15], which can be referenced in past research on prediction models for type 2 diabetes. The dataset consists of 21 feature variables and a binary target variable, where class labels are 0 for non-diabetes and 1 for prediabetes or diabetes [16]. The feature variables are listed in Table 1, which describes health and lifestyle factors.

**TABLE 1.** Health and lifestyle factors

|  |  |  |
| --- | --- | --- |
| **Column name (Abbreviation)** | **Data Detail** | **Description** |
| HighBP | 0 = No, 1 = Yes | High Blood pressure |
| HighChol | 0 = No, 1 = Yes | Cholesterol awareness |
| CholCheck  BMI  Smoker  Stroke  HeartDiseaseorAttack  PhysActivity  Fruits  Veggies  HvyAlcoholConsump | 0 = No, 1 = Yes  Continuous variable  0 = No, 1 = Yes  0 = No, 1 = Yes  0 = No, 1 = Yes  0 = No, 1 = Yes  0 = No, 1 = Yes  0 = No, 1 = Yes  0 = No, 1 = Yes | Cholesterol check (last 5 years)  Body Mass Index (Scale)  Smoked at least 100 cigarettes  Ever had a stroke  Heart disease or heart attack  Physical activity (last 30 days)  Consume fruits (≥ 1 per day)  Take vegetables (≥ 1 per day)  Heavy alcohol drinker |

Feature selection regarding the demographics and access to healthcare category is listed in Table 2, with key insights on general healthcare. With specific aspects of the details and description. The features from this table help to analyze the patterns and correlation between the health indicators listed and the prevalence of diabetes.

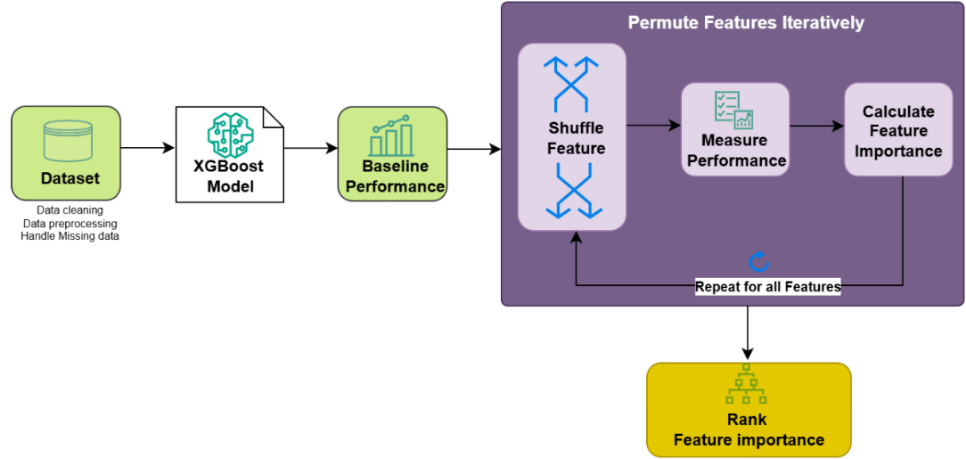
**TABLE 2.** Demographics and access to healthcare

|  |  |  |
| --- | --- | --- |
| **Column name (Abbreviation)** | **Data Detail** | **Description** |
| AnyHealthcare | 0 = No, 1 = Yes | Any kind of health coverage |
| NoDocbcCost | 0 = No, 1 = Yes | Doctor cost barrier (12 months) |
| GenHlth  MentHlth  Physhlth  DiffWalk  Sex  Age  Education  Income | Ordinal categorical  Continuous numerical (0 to 30)  Continuous numerical (0 to 30)  0 = No, 1 = Yes  0 = Female, 1 = male  Ordinal categorical  Ordinal categorical  Ordinal categorical | General health (scale 1 – 5)  Mental health (days)  Physical health (days)  Difficulty walking  Gender  Age group (scale 1 – 13)  Education level (1 – 6)  Income (scale 1 – 8) |

Shape of features (*X*) matrix: (236378, 21) and shape of target (*y*) vector: (236378,) with counts of 202,810 for class 0 and 33,568 for class 1. Python libraries and methods were run on Google Colaboratory for model training and analysis. Data preprocessing was done to handle class imbalances using SMOTE [17] and outliers using a robust scaler. Duplicates were not removed due to the inherent structure of the dataset and the preprocessing steps applied. The original dataset was large and included categorical grouping and scaling of continuous features, which resulted in identical feature values across multiple observations. Given that participants with the same group are likely to share the same values, they do not represent actual redundancies, as it is statistically inevitable that these groupings will result in duplicate rows. Removing these duplicates would delete valid observations, and as such, introduce artificial bias and reduce data diversity. Keeping the duplicates intact was required to preserve the integrity of the categorical transformations and prevent loss of valid variance, for the accuracy of the model to reflect the underlying distributions in the original data. Data types were consistent having *float64* and *int64,* plus the missing values in the dataset were handled using the *dropna()* method from the pandas’ library, which eliminates any rows that have at least one missing value, specifically, rows containing any missing or *NaN* (Not a Number) values to ensure completeness with no missing data points. With a focus on selectively increasing the minority class while maintaining the original majority, the SMOTE technique created new data points for the minority class without oversaturating the dataset to avoid overfitting.

Predictive accuracy was prioritized by developing a tuned, interpretable XGBoost model, utilizing the 80:20 train-test split, as it provides an optimal outcome [18]. Due to page limitations, the detailed comparison with alternative tree-based models (e.g., random forest, decision trees) and the SHAP interpretability method is not included. However, PFI was chosen for its global interpretability and model-agnostic nature. Cross-validation was performed on the full dataset using Stratified K-Fold with 5 folds, demonstrating the robustness of the model's performance.

The PFI technique was subsequently applied to explain the model's predictions. The baseline performance of the trained model on the test set was evaluated using the performance metrics, including F1-score, precision, accuracy, and recall. PFI will be done sequentially. After the initial feature permutation, measure the performance and calculate the importance. The feature importance is the difference between the baseline performance and the model performance after shuffling. Then repeat steps from feature shuffling as shown in Figure 1 for all the features in the dataset. Rank the features by their importance based on the performance drop and visualize the results via charts and plots.



**FIGURE 1.** XGBoost model and permutation feature ranking

|  |  |
| --- | --- |
|  |  |

Figure 1 depicts the methodology for XGBoost and feature ranking after iteratively permuting all the features. The concept is to measure the feature’s importance by calculating the increase in the model’s prediction error after permuting the feature. After the ML model is trained on the dataset, baseline performance is established by evaluating the model’s performance on the test set using the evaluation metrics. The feature values are shuffled randomly, and this interferes with the correlation between the feature and target variable. Model performance is quantified using shuffled feature data. The decrease in performance determines the significance of the feature. This is executed for all selected features, and the outcome is used to rank the features based on their impact on the performance of the model. Importance analysis of the PFI suggests feature permuting and splitting the dataset.

# RESULTS AND DISCUSSION

During model training, the datasets showed some issues of duplicates and outliers, and the features with major outlier issues were BMI, mental health, and physical health, while general health, cholesterol check, stroke, and heart disease had very minor outliers. Handling this requires the implementation of various techniques. To minimize the effect of outliers, we use the robust scaler method [19], and normalized the heart disease dataset by eliminating duplicates and using Principal Component Analysis (PCA) for dimensionality reduction [20]. The original dataset was reduced to an imbalanced set of 236,378 after handling missing values. Imbalanced datasets can skew model performance and affect the reliability of results [10]. For data preprocessing, missing values were dropped, and 67,136 samples were randomly selected during dataset reduction to achieve a balanced split for training and testing the model. This resulted in the binary diabetes classes having 33,568 count each. In this study, PCA and robust scaler methods were applied to the 67,136 instances of the 5050 BRFSS split dataset, but did not improve the model accuracy (74%).

However, using the imbalanced dataset of 236,378 instances and implementing SMOTE to handle the class imbalance of the original class distribution, the model was trained and evaluated using the resampled class distribution. The combination of these techniques with SMOTE, feature engineering, hyperparameter tuning, and a well-structured XGBoost pipeline improved the accuracy of the model to 90%. For the classification report, the recall of 97% and 84% for classes 0 and 1, respectively. The F1-score achieved was 91% for class 0 and a precision of 96% for class 1.

Permutation importance was evaluated on the test set. The number of instances was 81,124 samples with 21 features, and a total of 21 permutations were performed. The Shape used in permutation importance: [81124, 21]. To explore how feature importance varies across individual predictions, a 3D plot was generated first using the 81124 instances and then using a sample of 100 for a clearer version.

Figure 2 offers a granular 3D view of the permutation importance values and data samples across features (X-axis), instances (Y-axis), and the Z-axis is the computed importance value for a specific feature on an instance. The color ranges from low importance (purple) to high importance (yellow). Peaks indicate localized relevance from the combination of feature and datapoint, while flat surfaces indicate little influence. Figure 2 shows a wider variation of permutation importance scores, ranging approximately from 0.30 to 0.70 with a threshold of 0.47. Unlike the previous plot with a narrow range of 0.497 to 0.503, this version indicates greater heterogeneity in how features influence predictions across individual instances. Such local variation is critical in healthcare, where understanding individualized predictions matters.

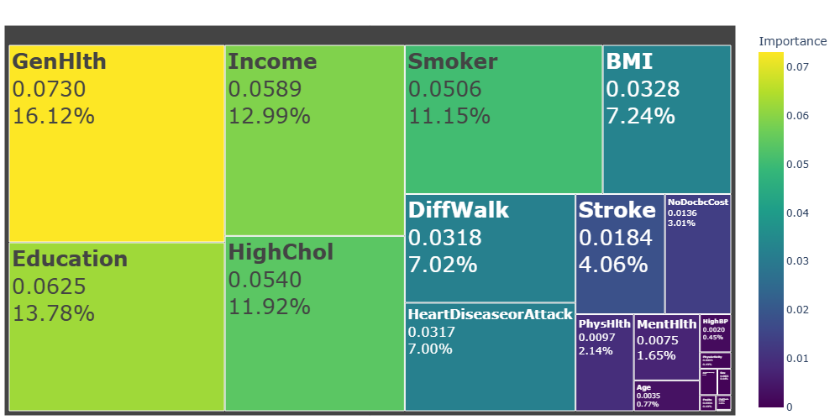
Figure 3 depicts the PFI of the plot. Y-axis lists the feature, X-axis shows the importance, and higher values indicate greater impact on the prediction. Longer bars mean the model’s prediction drops significantly if the feature is permuted. General health is of extreme importance and above the 0.035 midpoint threshold, reflecting its well-known correlation with diabetes risk and strongly influencing the model’s prediction. BMI and difficulty walking are moderately important, in contrast to veggies, any healthcare, fruits, and cholesterol check; when shuffled, they have minimal impact. Education, income, and high cholesterol features are significantly important.

|  |  |
| --- | --- |
| **FIGURE 2.** 3D Permutation importance with values | **FIGURE 3.** Bar plot for PFI method |

The scatter plot was used to show a relationship between features and the target. General health and diabetes binary compares the actual versus predicted values before and after permutation. It contained two plots. The predictions before permutation were closely aligned with the actuals, and a lower MAE of 0.0993 indicated relatively good performance. The permuted general health values had an MAE of 0.1733. The drop in prediction accuracy is shown in the higher MAE (from 0.0993 to 0.1733) after permutation. This indicates the general health’s predictive power for diabetes classification. Furthermore, the ~10% (0.0993) error rate is consistent with a 90% model accuracy.

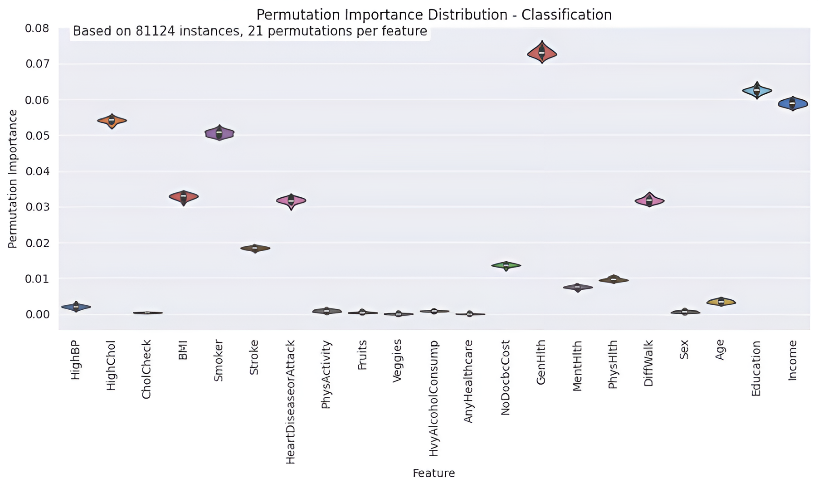
Another scatter plot that evaluated the effect of BMI on predicting diabetes, like the general health analysis above. Using the unshuffled BMI values, the subplot with no permutation had an MAE of 0.0993, and the permuted BMI plot displayed an MAE of 0.1316. The MAE increase suggests a clear drop in model performance, and prediction errors are more noticeable. The increase in MAE from 0.0993 to 0.1316 implies that BMI has a meaningful relationship with the diabetes status. This change is more pronounced with values (∆MAE ≈ 0.0323 BMI vs. 0.074 in GenHlth), indicating that general health is a stronger predictor in this dataset, as confirmed with the Fig. 3 bar plot.

Figure 4 visualizes the PFI tree-map for the classification model. It highlights the most influential feature on the XGBoost model prediction. Each rectangle represents a feature, and the color scale on the right ranges from 0 to 0.07, with lighter colors indicating higher feature impact. The numerical values within each rectangle represent the percentage contribution of the feature to the total permutation importance. General Health (GenHlth) is the most important feature (~0.07, yellow). Education and income (~0.05-0.06, green) follow, while features like age and high blood pressure have lower importance (~0.001-0.003, dark blue). BMI having 0.0328 indicates its significance; similar works highlight BMI among the most important features, suggesting that body mass significantly contributes to the model’s ability to differentiate between diabetes outcomes [21].



**FIGURE 4.** Tree-map classification

Figure 5 reflects a violin plot to show the distribution of different model features. The y-axis represents the PI values, indicating how much impact each feature has. The Kernel Density Estimation (KDE) displays the density of the data at different values. The miniature box plot in the violin provides additional details about the data distribution, showing the key summary statistics such as the median and interquartile range (IQR).



**FIGURE 5.** Distribution of PFI ccores

To analyze the importance of permutation in scores across the features, in contrast to the bar plot, Fig.5 violin plot for High BP has a PI of ~0.001 to 0.005 with higher concentration around 0.003 to 0.004, and general health displays a PI of 0.070 to 0.075, implying a strong magnitude in distribution. Low KDE can be seen in fruits, veggies, and physical activity, indicating little variation in their importance. The BMI suggests a larger density at 0.032 to 0.037 and mirrors the density estimation on either side of its axis.

# CONCLUSION

This study focuses on utilizing the XGBoost model for diabetes prediction and Permutation feature importance for interpretability. This study is important to bridge the gap between black box models and explain their predictions. By implementing data preprocessing techniques and SMOTE for handling class imbalance, model accuracy improves. Integration of Explainable Artificial Intelligence methods and XGBoost in the health system gives insight into the features that are influential to the model’s prediction. This collaboration can enhance trust and enable healthcare providers to make well-informed decisions and give personalized treatment.

# References

1. K. Hacker, “The Burden of Chronic Disease,” Mayo Clinic Proceedings: Innovations, Quality & Outcomes **8**(1), 112–119 (2024). https://doi.org/10.1016/j.mayocpiqo.2023.08.005

2. H. Alshari̇, and A. Odabas, “Machine Learning Model to Diagnose Diabetes Type 2 Based on Health Behavior,” Gazi University Journal of Science **35**(3), 834–852 (2022). https://doi.org/10.35378/gujs.931760

3. E. ŞAHiN, N.N. Arslan, and D. Özdemir, “Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning,” Neural Comput & Applic **37**(2), 859–965 (2025). https://doi.org/10.1007/s00521-024-10437-2

4. C.M. Chituru, S.-B. Ho, and I. Chai, “Integrating Spatial Computing with Clinical Pathology for Enhanced Diagnosis and Treatment Informatics in Healthcare,” JOIV: Int. J. Inform. Visualization **8**(3–2), 1762 (2024). https://doi.org/10.62527/joiv.8.3-2.2951

5. K. Natarajan, D. Baskaran, and S. Kamalanathan, “An adaptive ensemble feature selection technique for model-agnostic diabetes prediction,” Sci Rep **15**(1), 6907 (2025). https://doi.org/10.1038/s41598-025-91282-8

6. Palanivinayagam, and R. Damaševičius, “Effective Handling of Missing Values in Datasets for Classification Using Machine Learning Methods,” Information **14**(2), 92 (2023). https://doi.org/10.3390/info14020092

7. J. Wu, Y. Fang, X. Tan, S. Kang, X. Yue, Y. Rao, H. Huang, M. Liu, S. Qiu, and P.-T. Yap, “Detecting type 2 diabetes mellitus cognitive impairment using whole-brain functional connectivity,” Sci Rep **13**(1), 3940 (2023). https://doi.org/10.1038/s41598-023-28163-5

8. S. Gündoğdu, “Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique,” Multimed Tools Appl **82**(22), 34163–34181 (2023). https://doi.org/10.1007/s11042-023-15165-8

9. Ž.Ð. Vujovic, “Classification Model Evaluation Metrics,” International Journal of Advanced Computer Science and Applications (IJACSA) **12**(6), (2021). https://doi.org/10.14569/IJACSA.2021.0120670

10. M. Isnan, G.N. Elwirehardja, and B. Pardamean, “A review: Data pre-processing techniques used for diabetes prediction,” Procedia Computer Science **245**, 667–676 (2024). https://doi.org/10.1016/j.procs.2024.10.293

11. G.K. Rajbahadur, S. Wang, G.A. Oliva, Y. Kamei, and A.E. Hassan, “The Impact of Feature Importance Methods on the Interpretation of Defect Classifiers,” IIEEE Trans. Software Eng. **48**(7), 2245–2261 (2022). https://doi.org/10.1109/TSE.2021.3056941

12. Breiman, L.: Random Forest. Mach Learn. 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

13. W.-X. Ong, S.-B. Ho, and C.-H. Tan, “Enhancing migraine management system through weather forecasting for a better daily life,” Journal of Informatics and Web Engineering **2**(2), 201–217 (2023). https://doi.org/10.33093/jiwe.2023.2.2.15

14. Fisher, C. Rudin, and F. Dominici, “All Models are Wrong, but Many are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously,” (2018). https://doi.org/10.48550/arXiv.1801.01489

15. J. Nazreen, “Diabetes Health Indicators Dataset.” https://www.kaggle.com/datasets/julnazz/diabetes-health-indicators-dataset

16. Z. Xie, O. Nikolayeva, J. Luo, and D. Li, “Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques,” Prev. Chronic Dis. **16**, 190109 (2019). https://doi.org/10.5888/pcd16.190109

17. C. Azad, B. Bhushan, R. Sharma, A. Shankar, K.K. Singh, and A. Khamparia, “Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus,” Multimedia Systems **28**(4), 1289–1307 (2022). https://doi.org/10.1007/s00530-021-00817-2

18. R. Haque, S.-B. Ho, I. Chai, and A. Abdullah, “Parameter and hyperparameter optimisation of deep neural network model for personalised predictions of asthma,” Journal of Advances in Information Technology Vol **13**(5), (2022). https://doi.org/10.12720/jait.13.5.512-517

19. J.-L. Goh, S.-B. Ho, and C.-H. Tan, “Weather-based arthritis tracking: A mobile mechanism for preventive strategies,” Journal of Informatics and Web Engineering (JIWE) **3**(1), 210–225 (2024). https://doi.org/10.33093/jiwe.2024.3.1.14

20. C.-W. Teoh, S.-B. Ho, K.S. Dollmat, and C.-H. Tan, “Machine Learning Prediction Model for Early Student Academic Performance Evaluation in Video-Based Learning,” International Journal of Membrane Science and Technology **10**(2), 1529–1544 (2023). https://doi.org/10.15379/ijmst.v10i2.1822

21. U. Mitra, P. Sarkar, J. Mondal, and J. Kundu, “Enhancing Interpretability in Diabetics Prediction: A Comparative Study of SHAP, LIME and Permutation Feature Importance,” in *2025 AI-Driven Smart Healthcare for Society 5.0*, (IEEE, Kolkata, India, 2025), pp. 1–6. https://doi.org/10.1109/IEEECONF64992.2025.10962890