Original Article

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 1, No. 1, 2023

Machine Learning Approach for Diabetes Detection Using Fine-Tuned XGBoost Algorithm

Aga Maulana [1], Farassa Rani Faisal [1], Teuku Rizky Noviandy [1,*], Tatsa Rizkia [2], Ghazi Mauer Idroes [3], Trina Ekawati Tallei [4], Mohamed El-Shazly [5] and Rinaldi Idroes [6]

[1] Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; agamaulana@usk.ac.id (A.M.); farrasa.lldikti13@kemdikbud.go.id (F.R.F); trizkynoviandy@gmail.com (T.R.N)
[2] General Practitioner, School of Medicine, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; tatsa.rizkia@gmail.com (T.R)
[3] Department of Occupational Health and Safety, Faculty of Health Sciences, Universitas Abulyatama, Aceh Besar 23372, Indonesia; idroesghazi_k3@abulyatama.ac.id (G.M.I.)
[4] Department of Biology, Faculty of Mathematics and Natural Sciences, Sam Ratulangi University, Manado 95115, North Sulawesi, Indonesia; trina_tallei@unsrat.ac.id (T.E.T.)
[5] Department of Pharmacognosy, Faculty of Pharmacy, Ain-Shams University, Cairo 11566, Egypt; mohamed.elshazly@pharma.asu.edu.eg (M.E.-S.)
[6] Department of Chemistry, Faculty of Mathematics and Natural Sciences Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: trizkynoviandy@gmail.com

Article History

Received 8 July 2023
Revised 5 August 2023
Accepted 16 August 2023
Available Online 22 August 2023

Keywords:
Classification
Feature importance
Pima Indian
Hyperparameter tuning
Supervised learning

Abstract

Diabetes is a chronic condition characterized by elevated blood glucose levels which leads to organ dysfunction and an increased risk of premature death. The global prevalence of diabetes has been rising, necessitating an accurate and timely diagnosis to achieve the most effective management. Recent advancements in the field of machine learning have opened new possibilities for improving diabetes detection and management. In this study, we propose a fine-tuned XGBoost model for diabetes detection. We use the Pima Indian Diabetes dataset and employ a random search for hyperparameter tuning. The fine-tuned XGBoost model is compared with six other popular machine learning models and achieves the highest performance in accuracy, precision, sensitivity, and F1-score. This study demonstrates the potential of the fine-tuned XGBoost model as a robust and efficient tool for diabetes detection. The insights of this study advance medical diagnostics for efficient and personalized management of diabetes.

1. Introduction

Diabetes is a condition in which the level of glucose or blood sugar increases above the normal limit. Glucose accumulates in the blood because it is not properly absorbed by the body's cells, which can lead to various disturbances in organ function [1–3]. Diabetes can cause complications in many parts of the body and can ultimately increase the risk of premature death. Adults with diabetes also have two to three times a higher risk of heart attacks and strokes [4]. During pregnancy, poorly controlled diabetes can increase the risk of fetal death and other complications [5].

The number of people with diabetes has been increasing over the past decades, both in terms of cases and

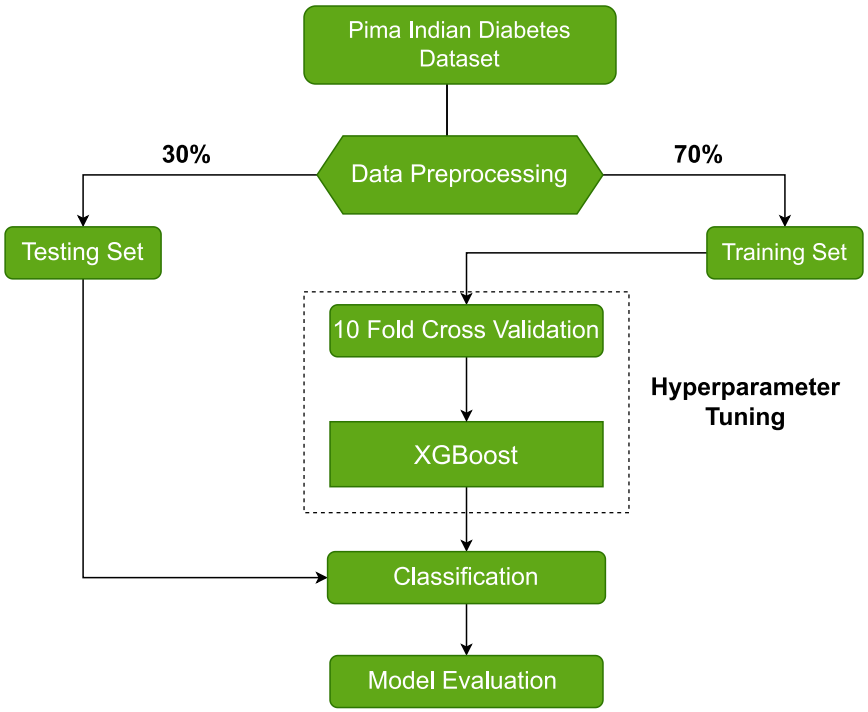


Figure 1. Flowchart of this study.

prevalence. In 2019, the number of people with diabetes worldwide reached 463 million, and it is predicted to continue rising to 700 million by the year 2045. The majority of diabetes patients live in low- and middle-income countries, and 1.6 million deaths are directly caused by diabetes each year [6–8].

Diabetes is a chronic condition that, currently, cannot be cured. However, early detection of diabetes provides a crucial opportunity to delay or prevent its progression into acute stages. Lifestyle changes and timely medical intervention can effectively manage blood glucose levels, reducing the risk of complications and minimizing the financial burden of treatment [9]. Therefore, the rapid and accurate identification, diagnosis, and analysis of diabetes are topics of research that are highly beneficial and crucial to be pursued.

Recent advancements in technology have opened new possibilities for the accurate and efficient diagnosis of diabetes [10]. Machine learning algorithms gained prominence as powerful tools in this domain due to their ability to analyze datasets and extract meaningful patterns from complex medical information. Numerous studies were conducted to develop and optimize machine learning models for diabetes detection [11, 12]. Chang et al. [13] utilized three machine-learning models for diabetes prediction and found that the random forest emerged as the most effective with an F1-score of 85.17%. Kumari et al [14] proposed an ensemble approach using a soft voting classifier and successfully achieved high performance in diabetes classification, attaining an F1-score of 80.60%. However, there is still room for further improvement in utilizing machine learning algorithms for diabetes detection. There is a possibility to further improve the performance of machine learning algorithms for diabetes detection.

In this study, we propose a fine-tuned XGBoost model for diabetes detection. Our objective was to achieve enhanced predictive performance by systematically exploring various hyperparameters during the training process. To accomplish this, we carefully selected a range of hyperparameters to tune and evaluate the combinations of these hyperparameters, so the model can get the optimal settings that maximize the model's predictive capabilities.

2. Materials and Methods

Figure 1 illustrates the flowchart of our study which consists of three main stages including data preprocessing, our proposed approach using XGBoost, and model evaluation.

2.1. Dataset and Preprocessing

In this study, we used the Pima Indian Diabetes dataset, which has been widely used in research related to diabetes diagnosis. The dataset was originally collected by the National Institute of Diabetes and Digestive and Kidney Diseases and is publicly available for research purposes [15]. It comprises 768 individuals, all of whom

Table 1. The variables in the dataset.

Variables	Description	Range
Pregnancies	Indicates the number of times pregnancy	0 to 17
Glucose	Refers to the plasma glucose concentration at 2 hours during an oral glucose tolerance test	0 to 199
Blood Pressure	Represents the diastolic blood pressure measured in mm Hg	0 to 122
SkinThickness	Measures the triceps skin fold thickness in millimeters	0 to 99
Insulin	Denotes the 2-hour serum insulin level in μU/mL	0 to 846
BMI	Body Mass Index, calculated as weight in kilograms divided by height in meters squared	0 to 67.1
Diabetes Pedigree	Represents the diabetes pedigree function	0.078 to 2.42
Age	Age of the individual in years	21 to 81
Outcome	A binary value indicating whether the individual is non-diabetic (0) or diabetic (1)	0 and 1

are Pima-Indian females aged 21 or older and residing close to Phoenix, Arizona, USA. Among the individuals, 500 are non-diabetic cases, while 268 represent diabetic cases.

The variables included in this dataset and the description are presented in Table 1. However, the presence of zero values in the dataset for variables such as “Glucose”, “Blood Pressure”, “Skin Thickness”, “Insulin”, and “BMI” is not feasible based on medical and scientific understanding. These zero values are due to the data entry errors or missing measurements. To address this issue, we used data imputation using the median values of the respective columns [16].

The dataset is randomly split into two subsets including the training set, comprising 70% of the data, and the testing set, comprising the remaining 30%. The training set was utilized to train the fine-tuned XGBoost model, allowing it to learn from the data and optimize its parameters for diabetes detection. On the other hand, the testing set was completely independent and served as a "holdout" dataset to assess the model's performance on the unseen data [17].

2.2. Proposed Approach

XGBoost is a super popular machine learning framework introduced in 2014 by Chen et al. [18]. We selected it for our study because it has a reputation for performing exceptionally well in a wide range of tasks [19–21]. XGBoost uses a method called ensemble learning, which combines several weak prediction models (decision trees) to make a strong and accurate overall model. This approach results in impressive predictions and finds applications in various fields.

In a previous study, Li et al. successfully applied XGBoost for diabetes prediction, but their approach lacked the reporting of details about the hyperparameter tuning process [22]. Even though XGBoost performs well in many tasks, training the model can be challenging due to its many hyperparameters. Therefore, it is necessary to carry out the hyperparameter tuning so that the performance of the trained model is good. To achieve this, we employed a random search approach with a cross-validation of 10 folds. Random search efficiently explores the hyperparameter space, allowing us to find the best combination of settings for our specific task. The hyperparameters were tuned and the explanation is summarized in Table 2.

2.3. Model Evaluation

The model that has been trained was evaluated using five metrics including accuracy, precision, sensitivity, specificity, and F1-score. Accuracy measures the overall correctness of the model's predictions, while precision assesses the accuracy of positive predictions [23]. Sensitivity gauges the ability to identify positive instances, and specificity evaluates the ability to correctly identify negative instances. The F1-score provides a balanced representation of precision and sensitivity to represent the model's overall performance in binary classification tasks [24]. The metrics are presented in equations 1, 2, 3, 4, and 5, respectively:

$$Accuracy = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})} \quad (1)$$

$$Precision = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FP})} \quad (2)$$

$$Sensitivity = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FN})} \quad (3)$$

$$Specificity = \frac{(\mathrm{TN})}{(\mathrm{TN} + \mathrm{FP})} \quad (4)$$

$$F1\text{–}Score = 2 \times \frac{(\mathrm{Precision} \times \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recal})} \quad (5)$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative. In this study, it is important to note that TP refers to individuals who were correctly identified as non-diabetic, and TN refers to individuals who were correctly identified as diabetic.

Additionally, we conducted a comparative analysis of our fine-tuned XGBoost model with six other popular machine learning algorithms, namely k-Nearest

Table 2. Hyperparameter and the explanation.

Hyperparameter	Range	Description
Max depth	2 to 30	Maximum depth of each decision tree, influencing model complexity and its ability to capture intricate patterns.
Learning rate	0.01 to 0.5	Step size at each iteration during model training, impacting convergence speed and potential for overshooting solutions.
Minimum child weight	1 to 30	Minimum sum of instance weight required in a child node, affecting the partitioning of data points in the decision tree.
Subsample	0.5 to 1	Fraction of samples used for training each tree, influencing model variance and robustness.
Colsample by tree	0.5 to 1	Fraction of features used for training each tree, impacting the diversity of trees in the ensemble.
N estimators	50 to 500	Number of boosting rounds or trees in the ensemble, affecting model complexity and generalization.
Gamma	0 to 5	Minimum loss reduction required for further partitioning on a leaf node during tree building, impacting regularization.

Table 3. Model performance on the testing set.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
KNN	77.92	80.25	86.30	63.53	67.92
Decision Tree	77.49	82.19	82.19	69.41	69.41
Random Forest	78.79	79.75	89.04	61.18	84.14
Naïve Bayesian	76.19	80.13	82.88	64.71	66.67
Support Vector Machine	77.49	77.33	91.10	54.12	63.89
Multilayer Perceptron	76.19	75.14	**93.15**	47.06	59.26
XGBoost	**82.68**	**84.42**	89.04	**71.76**	**86.67**

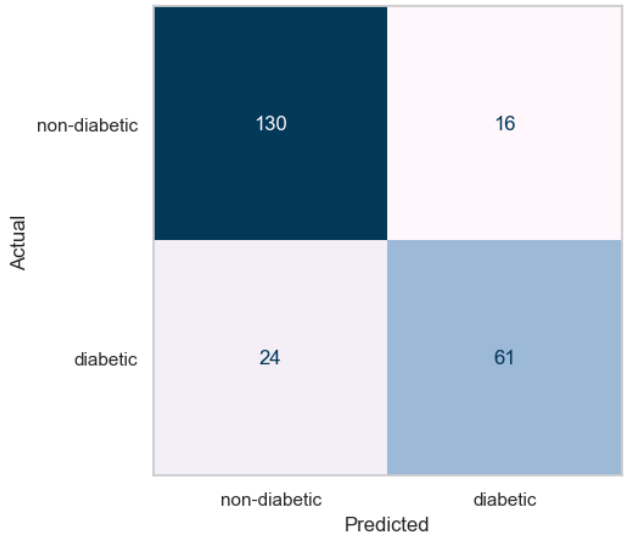


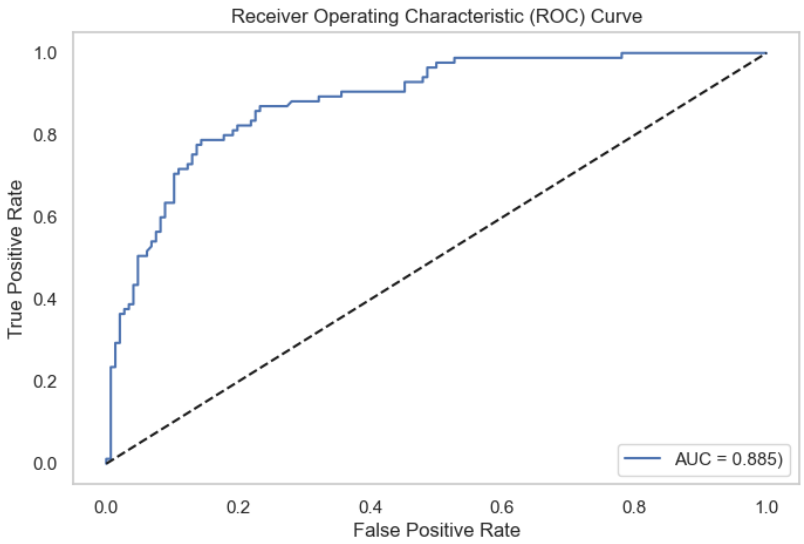
Figure 2. Confusion matrix of XGBoost model on the testing set.

Neighbors, Decision Tree, Random Forest, Naïve Bayesian, Support Vector Machine, and Multilayer Perceptron. The comparative study aimed to determine whether our fine-tuned XGBoost model outperforms or is on par with these widely used algorithms.

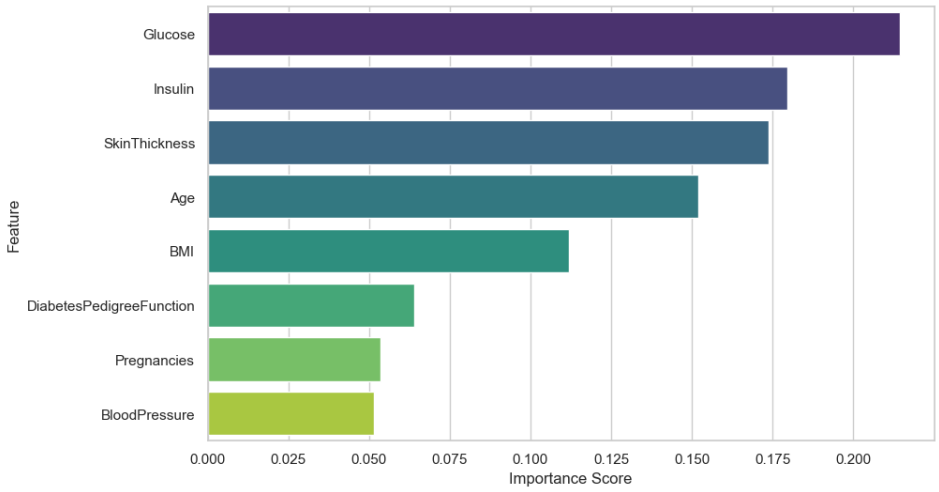
3. Results and Discussion

In this study, we trained a fine-tuned XGBoost model for diabetes detection. The training process involved hyperparameter tuning with randomized search, which was conducted for 100 iterations to explore different combinations of hyperparameters and optimize the performance. The model was also trained using 10-fold cross-validation. This technique involves dividing the dataset into 10 subsets, training the model nine times on different combinations of nine subsets, and evaluating its performance on the remaining subsets. This process was repeated 10 times, and the final performance metrics were obtained by averaging the results from the 10 iterations. It was done to ensure the model's robustness and generalizability by evaluating its ability to perform consistently on different subsets of the data. The selected hyperparameter values after conducting the tuning process were: max depth = 22, learning rate = 0.5, minimum child weight = 1.15, subsample = 1, colsample bytree = 0.5, N estimators = 100, and Gamma = 5. These hyperparameters were used to train the model. We used the testing set to evaluate the model's performance on the unseen data.

Table 3 provides a summary of the model performance on the testing set. Our fine-tuned XGBoost model achieved the highest accuracy of 82.68% among all tested models. It also exhibited the highest precision (84.42%) and specificity (71.76%) compared with other models, indicating its ability to correctly classify non-diabetic individuals. However, its sensitivity (89.04%) was slightly lower compared with the MLP model (93.15). It also showed that there is a significant difference between the values of sensitivity and specificity for all models. This discrepancy is likely due to the class imbalance in the dataset, which caused the models to be biased towards



Figures 3. ROC Curve of XGBoost model.



Figures 4. Feature importance of the fine-tuned XGBoost model.

the non-diabetic class, the majority class in this study. The confusion matrix, presented in Figure 2, illustrates the performance of the fine-tuned XGBoost model in the detection of diabetes. The matrix shows the classification results and gives a summary of the model's ability to differentiate between diabetic and non-diabetic cases. We can observe that the number of individuals correctly classified as non-diabetic was 130. Additionally, 16 individuals were incorrectly classified as non-diabetic while they were diabetic. On the other hand, the number of individuals correctly classified as diabetic is 61, while there were cases in which individuals were incorrectly classified as diabetic instead of non-diabetic and these individuals amounted to 24.

From these results, it is also important to note that for the context of our study, high specificity is essential because it ensures that diabetic individuals are less likely to be incorrectly classified as non-diabetic. This is crucial as misclassifying a diabetic person as non-diabetic could lead to the delay of necessary medical intervention and treatment. On the other hand, high sensitivity ensures that non-diabetic individuals are correctly identified, reducing the chances of false negatives. While our fine-tuned XGBoost model achieved lower sensitivity compared with the MLP model, it still achieved the highest F1-score, which reinforces the fact that our fine-tuned XGBoost model is indeed well-balanced for the task of diabetes detection. It performs well in minimizing non-diabetic individuals misclassified as diabetic and diabetic individuals misclassified as non-diabetic.

The Receiver Operating Characteristic (ROC) curve is presented in Figure 3. This figure visualizes the true positive rate (sensitivity) against the false positive rate at various threshold values to show the model's ability to

distinguish between diabetic and non-diabetic cases. From the ROC curve, a metric named Area Under the Curve (AUC) was calculated. This metric measures the overall discriminatory power of the model across all possible classification thresholds ranging from 0 to 1, where a score of 1 indicates perfect discrimination (the model perfectly separates the diabetic and non-diabetic), while a score of 0.5 suggests no discrimination or random guessing. In this study, the fine-tuned XGBoost model achieved an AUC score of 0.885. This high AUC score demonstrates the model's strong discriminative ability in correctly classifying diabetic and non-diabetic cases. The curve's positioning towards the upper-left corner signifies that the model effectively balances sensitivity and specificity, maximizing the true positive rate while minimizing the false positive rate.

To explain our model, we employed XGBoost feature importance analysis. The Gain metric was employed in this method to assess the significance of each feature. It quantifies the impact of individual features on enhancing the model's loss function during the decision tree construction process. The more times a feature is used in different trees and the more significant the gain achieved, the higher its feature importance score becomes. Higher Gain scores indicate more crucial features in influencing the model's predictions. The results are visualized in Figure 4. In our study, “Glucose” is the most important feature with a score of 0.214, followed by “Insulin” with 0.179 and “SkinThickness” with 0.173655. While the other features are less important compared with the top three features, they still provide information that helps models to make accurate predictions in diabetes detection.

Furthermore, we compared our results with those obtained in a previous study conducted by other researchers to assess the performance of our fine-tuned XGBoost model in diabetes detection and to understand how it measures up against the findings reported in the previous study. We compared the F1-score since the dataset has imbalanced classes, and the F1-score was a more suitable metric for evaluation in such a case. The comparison showed that our proposed fine-tuned XGBoost achieved a higher F1-score compared with previous studies conducted by Chang et al. (86.24%) [13] and Kumari et al. (80.60%) [14].

4. Conclusions

In this study, we proposed and fine-tuned an XGBoost model for diabetes detection using the Pima Indian Diabetes dataset. The results demonstrated that our model outperformed six other popular machine learning algorithms in terms of accuracy, precision, sensitivity, and F1-score. The model's excellent performance indicates its potential as a reliable tool for early diabetes detection and contributes valuable insights toward improved healthcare outcomes for individuals living with diabetes.

Despite its promising results, this study has some limitations that should be considered. First, this study uses the data imputation method used to address missing values in the dataset, which may introduce biases or inaccuracies. Second, the class imbalance in the dataset might have influenced the model's performance metrics to be biased toward the majority class. Lastly, it should be noted that diabetes is a complex and multifaceted condition influenced by various genetic, lifestyle, and environmental factors. Future studies should address these limitations, by introducing alternative imputation methods, employing class balancing methods, and conducting a comprehensive approach involving a combination of machine learning models, genetic analysis, and clinical expertise.

Author Contributions: Conceptualization, A.M., F.R.F and T.R.N.; methodology, A.M. and F.R.F.; software, A.M. and T.R. validation, T.R.N., T.E.T., M.E.-S. and R.I.; formal analysis, A.M. and G.M.I.; investigation, F.R.F. and T.R.; resources, F.R.F. and G.M.I.; data curation, T.R.N. and R.I.; writing—original draft preparation, A.M., F.R.F., T.R. and G.M.I.; writing—review and editing, T.R.N., T.E.T., M.E.-S. and R.I.; visualization, A.M.; supervision, T.R.N. and R.I.; project administration, T.R.N.; All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive any external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study available at Pima Indians Diabetes Database on Kaggle. Link: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.

Acknowledgments: The authors thank their institution and university.

Conflicts of Interest: All the authors declare that there are no conflicts of interest.

References

1. Vaishali, R., Sasikala, R., Ramasubbareddy, S., Remya, S., and Nalluri, S. (2017). Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset, *Proceedings of the IEEE International Conference on Computing, Networking and Informatics, ICCNI 2017*, Vols 2017-Janua, 1–5. doi:10.1109/ICCNI.2017.8123815.
2. Zimmet, P. Z., Magliano, D. J., Herman, W. H., and Shaw, J. E. (2014). Diabetes: a 21st century challenge, *The Lancet Diabetes & Endocrinology*, Vol. 2, No. 1, 56–64. doi:10.1016/S2213-8587(13)70112-8.
3. Quazi, A., Patwekar, M., Patwekar, F., Alghamdi, S., Rajab, B. S., Babalghith, A. O., and Islam, F. (2022). *In Vitro* Alpha-Amylase

Enzyme Assay of Hydroalcoholic Polyherbal Extract: Proof of Concept for the Development of Polyherbal Teabag Formulation for the Treatment of Diabetes, *Evidence-Based Complementary and Alternative Medicine*, Vol. 2022, 1577957. doi:10.1155/2022/1577957.

4. Rao, Y. K., Lee, M.-J., Chen, K., Lee, Y.-C., Wu, W.-S., and Tzeng, Y.-M. (2011). Insulin-mimetic action of rhoifolin and cosmosiin isolated from Citrus grandis (L.) Osbeck leaves: enhanced adiponectin secretion and insulin receptor phosphorylation in 3T3-L1 cells, *Evidence-Based Complementary and Alternative Medicine*, Vol. 2011.

5. Ye, W., Luo, C., Huang, J., Li, C., Liu, Z., and Liu, F. (2022). Gestational diabetes mellitus and adverse pregnancy outcomes: systematic review and meta-analysis, *BMJ*, e067946. doi:10.1136/bmj-2021-067946.

6. Association, A. D. (n.d.). Diabetes Overview The path to understanding diabetes starts here.

7. Hanson, M. A., Gluckman, P. D., Ma, R. C. W., Matzen, P., and Biesma, R. G. (2012). Early life opportunities for prevention of diabetes in low and middle income countries, *BMC Public Health*, Vol. 12, 1–9.

8. Dunachie, S., and Chamnan, P. (2019). The double burden of diabetes and global infection in low and middle-income countries, *Transactions of The Royal Society of Tropical Medicine and Hygiene*, Vol. 113, No. 2, 56–64.

9. Awah, P. K., Unwin, N., and Phillimore, P. (2008). Cure or control: complying with biomedical regime of diabetes in Cameroon, *BMC Health Services Research*, Vol. 8, No. 1, 43. doi:10.1186/1472-6963-8-43.

10. Ahsan, M. M., Luna, S. A., and Siddique, Z. (2022). Machine-Learning-Based Disease Diagnosis: A Comprehensive Review, *Healthcare*, Vol. 10, No. 3, 541. doi:10.3390/healthcare10030541.

11. Edeh, M. O., Khalaf, O. I., Tavera, C. A., Tayeb, S., Ghouali, S., Abdulsahib, G. M., Richard-Nnabu, N. E., and Louni, A. (2022). A Classification Algorithm-Based Hybrid Diabetes Prediction Model, *Frontiers in Public Health*, Vol. 10. doi:10.3389/fpubh.2022.829519.

12. Kumar, P. S., K, A. K., Mohapatra, S., Naik, B., Nayak, J., and Mishra, M. (2021). CatBoost Ensemble Approach for Diabetes Risk Prediction at Early Stages, *2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology(ODICON)*, IEEE, 1–6. doi:10.1109/ODICON50556.2021.9428943.

13. Chang, V., Bailey, J., Xu, Q. A., and Sun, Z. (2022). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms, *Neural Computing and Applications*. doi:10.1007/s00521-022-07049-z.

14. Kumari, S., Kumar, D., and Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier, *International Journal of Cognitive Computing in Engineering*, Vol. 2, 40–46. doi:10.1016/j.ijcce.2021.01.001.

15. Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., and Johannes, R. S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus, *Proceedings of the Annual Symposium on Computer Application in Medical Care*, American Medical Informatics Association, 261.

16. Jadhav, A., Pramod, D., and Ramanathan, K. (2019). Comparison of performance of data imputation methods for numeric dataset, *Applied Artificial Intelligence*, Vol. 33, No. 10, 913–933.

17. Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., and Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery, *Malacca Pharmaceutics*, Vol. 1, No. 2, 48–54. doi:10.60084/mp.v1i2.60.

18. Chen, T., and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794. doi:10.1145/2939672.2939785.

19. Alves, A. H. R., and Cerri, R. (2022). A Two-step Model for Drug-Target Interaction Prediction with Predictive Bi-Clustering Trees and XGBoost, *2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 1–8.

20. Maulana, A., Noviandy, T. R., Sasmita, N. R., Paristiowati, M., Suhendra, R., Yandri, E., Satrio, J., and Idroes, R. (2023). Optimizing University Admissions: A Machine Learning Perspective, *Journal of Educational Management and Learning*, Vol. 1, No. 1, 1–7. doi:10.60084/jeml.v1i1.46.

21. Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P., and Amjad, U. (2022). Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation, *Applied Sciences*, Vol. 12, No. 4, 2126.

22. Li, M., Fu, X., and Li, D. (2020). Diabetes Prediction Based on XGBoost Algorithm, *IOP Conference Series: Materials Science and Engineering*, Vol. 768, No. 7, 072093. doi:10.1088/1757-899X/768/7/072093.

23. Idroes, G. M., Maulana, A., Suhendra, R., Lala, A., Karma, T., Kusumo, F., Hewindati, Y. T., and Noviandy, T. R. (2023). TeutongNet: A Fine-Tuned Deep Learning Model for Improved Forest Fire Detection, *Leuser Journal of Environmental Studies*, Vol. 1, No. 1, 1–8. doi:10.60084/ljes.v1i1.42.

24. Noviandy, T. R., Maulana, A., Emran, T. B., Idroes, G. M., and Idroes, R. (2023). QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms, *Heca Journal of Applied Sciences*, Vol. 1, No. 1, 1–7. doi:10.60084/hjas.v1i1.12.