A Risk Prediction Model for Personalized Healthcare Using Machine Learning

Md. Abu Nishan1, a), Choo-Yee Ting1, 2, b), Chih-Yang Pee1, 3, c) and Helmi Zakariah4, d)

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

2Centre for Big Data and Blockchain Technologies, CoE for Advanced Cloud, Multimedia University, Persiaran Multimedia, 63100, Cyberjaya, Selangor, Malaysia.

3Centre for Image and Vision Computing, CoE for Artificial Intelligence, Multimedia University, Persiaran Multimedia, 63100, Cyberjaya, Selangor, Malaysia.

4Hayat Technologies Sdn Bhd, Level 15, Centerpoint North, Lingkaran Syed Putra, Mid Valley City, 59200 Kuala Lumpur, Malaysia,

*b) Corresponding author: cyting@mmu.edu.my*

a) 1221404588@student.mmu.edu.my

*c) cypee@mmu.edu.my*

*d) drhelmi@myhayat.my*

**Abstract.** Personalized healthcare is revolutionizing medical practice by tailoring interventions to individual characteristics. The integration of machine learning (ML) into this domain enables the development of predictive models capable of extracting meaningful insights from complex, high-dimensional patient data. Despite its potential, ML adoption in personalized healthcare is challenged by data heterogeneity, limited interpretability, and generalizability issues. This study aims to address these challenges by developing a comprehensive ML framework. A proprietary large-scale screening dataset was used, integrating clinical and behavioral information into a unified feature space. Advanced preprocessing steps were followed by multiple feature selection strategies, including filter, wrapper, embedded, and dimensionality reduction methods. Twelve ML classifiers were trained to predict individuals' risk of specific health conditions, such as diabetes and related comorbidities, and evaluated using standard metrics, with emphasis on accuracy, precision, recall, F1-score, and ROC AUC. Among the evaluated models, ensemble methods Gradient Boosting, Voting Classifier, and Stacking Classifier achieved the highest accuracies (≥ 0.7940) and F1-scores (≥ 0.8800) in predicting health risks. SVM and Stacking Classifier demonstrated superior recall (0.9866 and 0.9862), making them effective at identifying at-risk individuals, while Naive Bayes achieved the highest precision (0.8427), indicating reliability in positive predictions. Logistic Regression and LDA provided well-balanced performance across metrics, while ensemble models such as Bagging and Random Forest also showed strong recall and F1-scores. The findings highlight the effectiveness of ensemble learning techniques for health risk prediction in personalized healthcare, with Gradient Boosting and Voting Classifier emerging as the most robust and reliable models. This study contributes a scalable and interpretable ML framework to support precision medicine through improved early detection, risk stratification, and targeted intervention.

# INTRODUCTION

Personalized healthcare marks a significant advancement in modern medicine by shifting from a one-size-fits-all approach to individualized care tailored to genetic, clinical, lifestyle, and environmental factors. The integration of machine learning (ML) into personalized healthcare enables the development of data-driven predictive models capable of uncovering complex, non-linear patterns within high-dimensional datasets. Leveraging large volumes of patient data, ML facilitates early diagnosis, risk stratification, treatment planning, and continuous monitoring, ultimately improving patient outcomes, optimizing resource use, and supporting precision medicine initiatives [1].

Despite its promise, the practical deployment of ML in personalized healthcare presents notable challenges [2]. These include the heterogeneity of medical data—ranging from structured clinical records and sensor outputs to unstructured notes and genomic data—collected from disparate sources with inconsistent quality and formats [3]. Such variability complicates data integration, model development, and generalizability across populations. Moreover, the opaque or "black-box" nature of many ML algorithms limits interpretability, which is a critical requirement in clinical settings where explainability fosters clinician trust and ensures patient safety [4]. Further barriers include restricted data accessibility, privacy regulations, and a lack of standardized data-sharing protocols. Many ML models are developed using small, homogeneous datasets, limiting their applicability in real-world, diverse populations.

In response to these challenges, this study proposes a robust methodology that applies multiple ML algorithms to a large-scale, real-world screening dataset. This dataset combines clinical, demographic, and lifestyle information from 14,713 individuals and contains 582 features after integration and preprocessing. The objective is to identify individuals at risk of specific health conditions by developing accurate and interpretable predictive models. To enhance model learning, we applied a comprehensive feature selection pipeline incorporating filter, wrapper, embedded, and dimensionality reduction techniques to identify the most informative predictors. We then trained and evaluated twelve classification models to assess their performance. The emphasis is not only on predictive accuracy but also on transparency and fairness, contributing toward more reliable, generalizable, and equitable personalized healthcare solutions.

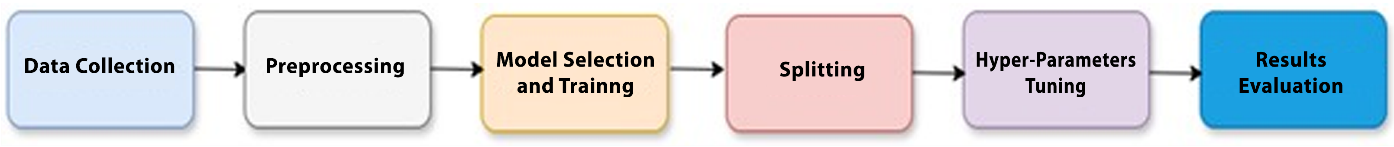
# LITERATURE REVIEW

Recent literature highlights the increasing application of machine learning (ML) in predictive healthcare, particularly for chronic disease diagnosis such as diabetes. Various studies have utilized both public and private datasets with differing methodologies and success rates. In contrast, RF, Support Vector Machine (SVM), and Decision Tree (DT) were applied on the St. Paul’s Hospital dataset [5], achieving accuracies of 96.8%, 99.6%, and 98.2%, though dataset size and representativeness were noted as limitations. However, concerns regarding model interpretability and generalizability remained. Similarly, Maximum Likelihood Estimation (MLE) was applied on genomic data [6], achieving 77.47% accuracy but limited by data encryption constraints that restricted access and optimization. Emerging approaches like ML4CAD were explored [7], yielding a modest 69.48% accuracy on the Alberta Provincial dataset, with limited generalizability. Ensemble learning proved more effective conducting the study [8], using LR, SVM, NB, and RF on public datasets, reaching 95.4% accuracy but noted regulatory limitations. The broader role of machine learning in advancing precision medicine and personalized healthcare has also been emphasized in [9]. Some studies focused on specific clinical goals. For instance, logistic regression used on the Khulna Diabetes Center dataset achieved 88% accuracy but faced limitations in early-stage prediction [10]. XGBoost achieved 99.90% accuracy on the Cuenca Hospital dataset [11], though it focused solely on non-invasive detection. Another study explored interpretable ML using RF and NB on the Pima Indian dataset [12], with moderate accuracies of 79.6% and 79.1%, and highlighted a lack of trust in black-box models. SVM was applied to GEO microarray data for gene biomarker selection [13], achieving 100% accuracy. Ensemble learning was further applied as a stacking ensemble to Saudi hospital data [14], achieving 94.5% accuracy but struggling to distinguish between T1DM, T2DM, and pre-diabetes. Despite the promising result, its narrow focus on T2D diagnosis raised concerns about overfitting and limited real-world applicability. Overall, these studies reflect increasing sophistication in ML applications but also reveal recurring challenges in dataset diversity, model transparency, and clinical translation.

In summary, while many ML models demonstrate strong predictive capabilities—especially those leveraging ensemble methods or curated datasets—critical limitations persist. These include poor generalizability, limited interpretability, and dataset-related constraints such as heterogeneity and small sample sizes. Additionally, few studies address regulatory and clinical integration concerns. Moving forward, there is a need for robust, explainable models trained on diverse datasets, accompanied by transparent validation and clear reporting to enhance their practical relevance in clinical settings.

# METHODOLOGY

Figure 1 outlines the key steps to conduct this study.



**FIGURE 1.** Research framework

## Dataset

For this study, a private dataset obtained from a proprietary source was used to develop predictive models for identifying individuals at risk of specific health conditions. The primary dataset comprises 19,536 individual records with 67 features, including demographic information, screening questionnaire responses, and clinical parameters. A secondary dataset consists of 14,729 unique applicants with 472 derived features generated through domain-specific feature extraction techniques, capturing detailed aspects of participants’ medical history, lifestyle behaviors, and self-reported health status. After a systematic data integration process, a final merged dataset was created, containing 14,713 matched records with a total of 582 features. These variables encompass key indicators such as age, gender, ethnicity, comorbidities (e.g., hypertension, hyperlipidemia, family history of diabetes), lifestyle factors (e.g., smoking status, dietary patterns, physical activity), and various clinical and anthropometric measurements, providing a comprehensive representation of the target population enrolled in the screening program. The objective of this methodology is to apply and evaluate machine learning classification algorithms to accurately predict health risk outcomes. The dependent variable is a binary indicator representing the presence or absence of a specific health condition, while the independent variables consist of the 582 features derived from demographic, clinical, lifestyle, and medical history data. Multiple classification models were trained and evaluated to identify the most effective approach for early risk detection and intervention.

## Data Preprocessing

Data preprocessing was conducted through a series of systematic steps to ensure the quality and consistency of the dataset prior to analysis. Initially, all feature labels and textual entries, originally recorded in Malay, were translated into English to standardize language representation and facilitate subsequent processing. Missing values were addressed using appropriate imputation strategies tailored to nature and distribution of the data. The dataset was then subjected to data splitting to enable robust model training and evaluation. Categorical variables were encoded using one-hot encoding to convert them into a machine-readable numerical format. Further data cleaning steps involved the identification and removal of irrelevant or redundant features, including those with high proportions of missing or null values (i.e., NA or None). In addition, features representing complex combinations or non-informative categorical variables were eliminated to reduce dimensionality and mitigate noise. These preprocessing steps were essential to enhance data integrity and ensure the suitability of the dataset for downstream machine learning and statistical analysis.

The methodological framework employed in this study comprises a series of structured steps designed to ensure rigorous data handling and model development. The process begins with data collection, where a proprietary dataset, as detailed in the Dataset section, was acquired for analysis. This is followed by a comprehensive data preprocessing phase, which includes cleaning and transforming the raw data to ensure quality and consistency. Key tasks in this phase include handling missing values, encoding categorical variables through one-hot encoding, scaling numerical features, and generating derived features where necessary to enhance predictive potential. Model selection was performed by evaluating several machine learning algorithms suited to the nature of the task and the structural characteristics of the dataset. Subsequently, the data were split into training (75%) and testing (25%) subsets to facilitate effective model training while enabling unbiased evaluation on unseen data. During the training phase, models were fitted to the training data to learn underlying patterns and relationships. Hyperparameter tuning was carried out using systematic search strategies to optimize model configurations and improve performance. Finally, model evaluation and prediction were conducted on the hold-out test set using standard performance metrics such as accuracy, precision, recall, and F1-score to assess the model’s ability to generalize to unseen data. This structured methodology ensures a robust and reproducible approach to data-driven analysis.

## Feature Selection Methods

To enhance model performance and interpretability while mitigating the challenges posed by high dimensionality, a comprehensive feature selection strategy was employed in this study following initial preprocessing steps, which included imputation, encoding, and feature scaling. Multiple classes of feature selection techniques—filter, wrapper, embedded, and dimensionality reduction methods—were systematically applied to identify the most informative predictors from the high-dimensional dataset. Filter methods were first utilized due to their computational efficiency and ability to independently assess each feature's relevance [15]. Among these, Mutual Information was used to capture the statistical dependence between individual features and the target variable, effectively identifying features that convey the most information about the outcome [16]. The Chi-squared test was applied to evaluate the association between categorical features and the target variable, particularly suited for identifying predictive categorical traits [17]. Additionally, ANOVA (Analysis of Variance) was employed to assess whether numerical features had statistically significant differences in mean values across categorical target classes [18]. To further improve model robustness and interpretability, correlation-based feature selection was conducted to eliminate highly correlated variables, thus reducing redundancy and multicollinearity. Wrapper methods were subsequently employed to refine feature subsets based on their impact on predictive performance. Recursive Feature Elimination (RFE) iteratively removed the least important features as determined by an estimator, allowing the model to retain only the most influential variables [19]. The Boruta algorithm, a robust wrapper method built around a Random Forest classifier, was particularly valuable for identifying all relevant features, including those that may be redundant, though it required significant computational resources [20]. Sequential Feature Selection, both forward (SFS) and backward (SBS), was also tested, incrementally adding or removing features to optimize performance, albeit with the risk of local optima due to its greedy nature [21]. Embedded methods, which integrate feature selection within the model training process itself, were also explored. LASSO (L1 regularization) was especially effective for high-dimensional data, as it encouraged sparsity by driving less informative feature coefficients to zero. Ridge regression (L2 regularization), though it retained all features, helped mitigate the influence of less relevant ones by shrinking their coefficients [22]. Additionally, tree-based models such as Random Forest and Gradient Boosting were employed to derive feature importance scores, from which features exceeding a defined threshold were selected—this approach balanced interpretability with model-driven relevance assessment. Lastly, dimensionality reduction techniques were considered to project the data into a lower-dimensional space while preserving the underlying structure. Principal Component Analysis (PCA) was used to capture the most significant variance in the data through orthogonal transformations, aiding in both dimensionality and noise reduction. Linear Discriminant Analysis (LDA), a supervised method, was applied to maximize class separability, proving particularly beneficial in classification contexts. Together, these methods formed a robust multi-faceted framework for feature selection, ensuring that the final model was both parsimonious and performant. After analyzing the results, Boruta is chosen as the feature selection method to conduct this study. It is chosen because it is robust and helps to identify all relevant features, including those with complex interactions. Can be more computationally expensive than other methods.

# Results AND Discussion

Several classification models were trained and evaluated on the preprocessed data using Python's scikit-learn library based on the features selected by Boruta. A total of 12 classifiers are used. These are Logistic Regression, SVM, Decision Tree, Random Forest, Gradient Boosting, Naive Bayes, K-Nearest Neighbors, LDA, Neural Network (MLP), Voting Classifier, Bagging Classifier, Stacking Classifier. The results are demonstrated in Table 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TABLE 1. Summary of accuracy, precision, recall, F1-score, and ROC-AUC | | | | | | |
| Classifier | Accuracy | Precision | Recall | F1-score | ROC AUC | Curve Ref |
| Logistic Regression | 0.7899 | 0.7976 | 0.9742 | 0.8771 | 0.7204 | Figure 5 |
| SVM | 0.7929 | 0.7941 | 0.9866 | 0.8800 | 0.6195 | Figure 6 |
| Decision Tree | 0.7382 | 0.7921 | 0.8947 | 0.8403 | 0.6410 | Figure 7 |
| Random Forest | 0.7747 | 0.7969 | 0.9491 | 0.8664 | 0.6733 | Figure 8 |
| Gradient Boosting | **0.7959** | 0.7982 | 0.9834 | 0.8812 | 0.7247 | Figure 9 |
| Naive Bayes | 0.7328 | 0.8427 | 0.8025 | 0.8221 | 0.7117 | Figure 10 |
| K-Nearest Neighbors | 0.7429 | 0.8132 | 0.8644 | 0.8380 | 0.6453 | Figure 11 |
| LDA | 0.7880 | 0.8003 | 0.9654 | 0.8751 | 0.7177 | Figure 12 |
| Neural Network (MLP) | 0.7782 | 0.7961 | 0.9569 | 0.8691 | 0.6784 | Figure 13 |
| Voting Classifier | **0.7956** | 0.7971 | 0.9852 | 0.8812 | 0.7003 | Figure 14 |
| Bagging Classifier | 0.7641 | 0.7959 | 0.9325 | 0.8588 | 0.6672 | Figure 15 |
| Stacking Classifier | **0.7940** | 0.7952 | 0.9862 | 0.8805 | 0.6748 | Figure 16 |

Figures 2, 3, and 4 represent the confusion matrices for the top-performing classifiers based on classification accuracy: Gradient Boosting, Voting Classifier, and Stacking Classifier, respectively. These visualizations provide insights into the classification performance by detailing the number of true positives, true negatives, false positives, and false negatives for each model. Figure 2 shows the confusion matrix for the Gradient Boosting model, which achieved a strong balance between sensitivity and specificity. Figure 3 presents the results of the Voting Classifier, indicating consistent performance in correctly identifying positive instances. Figure 4 displays the Stacking Classifier, which demonstrated high recall, particularly in detecting the positive class, but produced slightly more false positives compared to the others. These matrices highlight each model's strengths and trade-offs, offering deeper insight beyond overall accuracy, especially for datasets with class imbalance.

|  |  |  |
| --- | --- | --- |
| **FIGURE 2.** Confusion matrix of GBM | **FIGURE 3.** Confusion matrix of Voting classifier | **FIGURE 4.** Confusion matrix of Stacking classifier |

The ROC-AUC figures of all 12 classifiers are shown in Figures 5-16.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **FIGURE 5.** Logistic regression | **FIGURE 6.** SVM | **FIGURE 7.** Decision tree |
|  |  |  |
| **FIGURE 8.** Random forest | **FIGURE 9.** Gradient boosting | **FIGURE 10.** Naive bayes |
|  |  |  |
| **FIGURE 11.** K-nearest neighbors | **FIGURE 12.** LDA | **FIGURE 13.** Neural network (MLP) |
|  |  |  |
|  |  |  |
| **FIGURE 14.** Voting classifier | **FIGURE 15.** Bagging classifier | **FIGURE 16.** Stacking classifier |

Twelve classifiers were evaluated on the personalized healthcare dataset using features selected by the Boruta algorithm, with the results summarized in Table 1. Among these, ensemble methods such as Gradient Boosting and the Voting Classifier consistently outperformed simpler models like Decision Tree and Naive Bayes, demonstrating superior accuracy, F1-score, and AUC as illustrated in Figure 9 and Figure 14. Notably, Support Vector Machine (SVM) and Stacking Classifier achieved exceptionally high recall values (0.9866 and 0.9862, respectively), highlighting their excellent sensitivity for detecting positive cases, a critical aspect in healthcare risk prediction. However, SVM’s relatively low ROC AUC score (0.6195) indicates a trade-off with specificity, which may stem from overfitting or sensitivity to feature scaling. Although Naive Bayes showed lower overall accuracy (0.7328), it attained the highest precision (0.8427), reflecting its capability to minimize false positives—a valuable trait in clinical settings where the cost of incorrect positive predictions is significant. Feature importance analysis conducted via Boruta identified key predictors that influenced classifier performance, particularly benefiting models that assume linear relationships. Computational cost analysis revealed that ensemble methods demand more training time, which is justified by their enhanced predictive performance. Hyperparameter tuning was systematically executed using a randomized search combined with cross-validation to ensure optimized model configurations. The presence of class imbalance in the dataset contributed to disparities among evaluation metrics, and although synthetic balancing techniques were not applied in this study, future research will investigate their effects. Additionally, statistical significance testing will be incorporated to validate the robustness of observed performance differences. Confusion matrix metrics for the leading models, also presented in Figure 2, Figure 3 and Figure 4, provide further insight into the balance between sensitivity and specificity, thereby supporting informed model selection aligned with clinical priorities.

# CONCLUSION

This study demonstrates the practical impact of machine learning in supporting personalized healthcare, particularly in screening and risk prediction. Among the twelve models evaluated, ensemble methods—especially Gradient Boosting, Voting Classifier, and Stacking Classifier—emerged as the most reliable and consistent performers across multiple metrics. Gradient Boosting achieved the highest accuracy of 0.7959, along with a strong F1-score of 0.8812 and a robust ROC AUC of 0.7247, making it the top-performing model overall.

These models demonstrated strong capabilities in identifying individuals at risk of health conditions, which is critical in real-world settings where early intervention can save lives, reduce long-term complications, and lower healthcare costs. Their ability to detect subtle patterns in large, complex datasets enables timely, targeted, and data-informed decision-making, ultimately improving patient outcomes and optimizing resource allocation.

Looking ahead, future work will focus on increasing the transparency, fairness, and robustness of these machine learning models. By integrating Explainable AI (XAI) techniques such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations), we aim to provide clear, human-understandable explanations of model predictions—an essential step in gaining trust from clinicians and patients. Additionally, applying SMOTE (Synthetic Minority Over-sampling Technique) will help address class imbalance in real-world healthcare data, ensuring that predictions remain reliable even for underrepresented groups. Together, these enhancements will bring machine learning models closer to real-world clinical adoption, supporting more equitable, transparent, and effective personalized healthcare solutions.

**REFERENCES**

1. M. Athar, “Potentials of artificial intelligence in familial hypercholesterolemia: Advances in screening, diagnosis, and risk stratification for early intervention and treatment,” International Journal of Cardiology **412**, 132315 (2024).
2. Y. H. Li, Y. L. Li, M. Y. Wei, and G. Y. Li, “Innovation and challenges of artificial intelligence technology in personalized healthcare,” Scientific Reports **14**, 18994 (2024).
3. J. Peng, E. C. Jury, P. Dönnes, and C. Ciurtin, “Machine learning techniques for personalised medicine approaches in immune-mediated chronic inflammatory diseases: Applications and challenges,” Frontiers in Pharmacology **12**, 720694 (2021).
4. B. Y. Kasula, “Harnessing machine learning for personalized patient care,” Transactions on Latest Trends in Artificial Intelligence **4**, 4 (2023).
5. M. L. Prasad, A. Kiran, and P. C. Shaker Reddy, “Chronic kidney disease risk prediction using machine learning techniques,” Journal of Information Technology Management **16**, 118–134 (2024).
6. W. Briguglio, P. Moghaddam, W. A. Yousef, I. Traoré, and M. Mamun, “Machine learning in precision medicine to preserve privacy via encryption,” Pattern Recognition Letters **151**, 148–154 (2021).
7. P. Ghasemi, M. Greenberg, D. A. Southern, B. Li, J. A. White, and J. Lee, “Personalized decision making for coronary artery disease treatment using offline reinforcement learning,” npj Digital Medicine **8**, 99 (2025).
8. S. K. S. Modak and V. K. Jha, “Diabetes prediction model using machine learning techniques,” Multimedia Tools and Applications **83**, 38523–38549 (2024).
9. B. Srinivasaiah, “The power of personalized healthcare: Harnessing the potential of machine learning in precision medicine,” International Journal of Science and Research **2024**, 426–429 (2024).
10. M. M. Hassan, Z. J. Peya, S. Mollick, M. A. M. Billah, M. M. H. Shakil, and A. U. Dulla, “Diabetes prediction in healthcare at early stage using machine learning approach,” in 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 01–05 (IEEE, 2021).
11. A. Prabha, J. Yadav, A. Rani, and V. Singh, “Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier,” Computers in Biology and Medicine **136**, 104664 (2021).
12. V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, “Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms,” Neural Computing and Applications **35**, 16157–16173 (2023).
13. J. Li, J. Ding, D. U. Zhi, K. Gu, and H. Wang, “Identification of type 2 diabetes based on a ten-gene biomarker prediction model constructed using a support vector machine algorithm,” BioMed Research International **2022**, 1230761 (2022).
14. M. Gollapalli, A. Alansari, H. Alkhorasani, M. Alsubaii, R. Sakloua, R. Alzahrani, … and W. Albaker, “A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM,” Computers in Biology and Medicine **147**, 105757 (2022).
15. M. Khodarahmi and V. Maihami, “A review on Kalman filter models,” Archives of Computational Methods in Engineering **30**, 727–747 (2023).
16. H. Zhou, X. Wang, and R. Zhu, “Feature selection based on mutual information with correlation coefficient,” Applied Intelligence **52**, 5457–5474 (2022).
17. A. C. Miola and H. A. Miot, “Comparing categorical variables in clinical and experimental studies,” Jornal Vascular Brasileiro **21**, e20210225 (2022).
18. S. Lakshmi and C. P. Maheswaran, “Effective deep learning-based grade prediction system using gated recurrent unit (GRU) with feature optimization using analysis of variance (ANOVA),” Automatika **65**, 425–440 (2024).
19. M. Awad and S. Fraihat, “Recursive feature elimination with cross-validation with decision tree: Feature selection method for machine learning-based intrusion detection systems,” Journal of Sensor and Actuator Networks **12**, 67 (2023).
20. N. Farhana, A. Firdaus, M. F. Darmawan, and M. F. Ab Razak, “Evaluation of Boruta algorithm in DDoS detection,” Egyptian Informatics Journal **24**, 27–42 (2023).
21. G. N. Ahmad, S. Ullah, A. Algethami, H. Fatima, and S. M. H. Akhter, “Comparative study of optimum medical diagnosis of human heart disease using machine learning technique with and without sequential feature selection,” IEEE Access **10**, 23808–23828 (2022).
22. M. Arashi, M. Roozbeh, N. A. Hamzah, and M. Gasparini, “Ridge regression and its applications in genetic studies,” PLOS ONE **16**, e0245376 (2021).