



INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv



Detecting Need-Attention Patients Using Machine Learning

Theng-Jia Law^{a,*}, Choo-Yee Ting^a, Helmi Zakariah^b

^a Multimedia University, Cyberjaya, Selangor, Malaysia

^b HAYAT Technologies Sdn. Bhd., Centrepont North, Mid Valley City Lingkaran Syed Putra, Kuala Lumpur, Malaysia

Corresponding author: *thengjialaw@gmail.com

Abstract— In healthcare, detecting patients who need immediate attention is difficult. Identifying the critical variables is challenging in patient detection because human intervention in variable selection is required. Consequently, patients who need immediate attention often experience prolonged waiting times. Researchers have investigated various approaches to identify those who require attention. One of the techniques is leveraging Artificial Intelligence (AI). However, identifying the optimal feature set and predictive model is complex. Therefore, this study has attempted to (i) identify the critical features and (ii) develop and evaluate predictive models in detecting those who need attention. The dataset is collected from one of the healthcare companies. The dataset collected contains 67 variables and 51102 records. It consists of patient information and questionnaires answered by each participant registered in the Selangor Saring Program. Important features were identified in detecting those who need attention on treated data. Multiple classifiers were developed due to their simplicity. The models were evaluated before and after hyperparameter tuning based on accuracy, precision, recall, F1-score, Geometric Mean, and Area Under the Curve. The findings showed that the Stacking Classifier produced the highest accuracy (69.9%) when using the blood dataset. In contrast, Extreme Gradient Boosting achieved the highest accuracy (81.7%) when the urine dataset was used. This work can be extended to explore the incorporation of Points of Interest and geographical data near patients' residences and study other ensemble models to enhance the performance of detecting those who need attention.

Keywords— Need-attention patient detection; artificial intelligence; machine learning.

Manuscript received 5 Dec. 2022; revised 9 Jul. 2023; accepted 25 Sep. 2023. Date of publication 31 May 2024.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Patients under the category “Need-Attention” require medical attention due to their physical or mental condition. This category includes patients who experience several symptoms, such as pain, discomfort, or distress. The patients who require attention may require different levels of medical attention, ranging from primary care to intensive care. Hence, timely and accurate medical treatment should be provided for those needing attention to alleviate their symptoms and prevent further complications such as life-threatening ones. However, some patients require attention but do not receive timely treatment.

It causes delayed treatment and worsens symptoms, prolonged illness, and even death in severe cases [1], [2], [3], [4], [5]. For instance, [1] have concluded that delayed treatment increases the death risk by estimation from 1.2% to 3.2% weekly in early-stage cancers. Besides, delayed treatment of patients who require immediate attention is associated with an increased risk of death [2]. The researchers stated that there is a 6–8% increase in the risk of death for

every four-week delay. Moreover, [3] stated that about 30% of patients with tuberculosis are not diagnosed or reported, contributing to the 1.5 million annual deaths globally. Thus, it is essential to identify patients needing attention to prevent delayed treatment and adverse outcomes.

To prevent delayed treatment and adverse outcomes, identifying the critical factors that are used to detect patients who need attention and taking corresponding clinical measures are crucial [6], [7]. Nevertheless, [8] stated that decision-making in medical treatment is complex as there are multiple factors to consider aside from the purely medical aspects. Not all the factors have a direct impact on treatment decisions because certain factors are of primary importance for individual patients and clinicians, while others are structural and may indirectly influence the ultimate decision [9].

The Risk Assessment Questionnaire is developed based on the standard screening protocol per Malaysia's Ministry of Health (MOH) Clinical Practice Guidelines and related guidance from the medical professional body, which acts as the first filtration process to identify participants eligible for

certain health screenings. The objectives of this study are:

- To identify the essential features used in detecting those needing attention.
- To develop and evaluate predictive models to detect those who need attention.

A. Identifying the Important Variables

Many researchers did not study the laboratory variables but only the demographic characteristics of patients, which contribute to the higher hospitalization risk [10]. [11] considering general criteria such as personal, social, and other factors deemed relevant to related disease-specific outcomes can help clinicians detect patients who require attention. In the medical field, feature selection could be used to identify the most crucial factors that cause disease. Besides that, it is essential to improve learning performance, prevent overfitting, and reduce computational costs when there are many features in the dataset [12], [13], [14]. Table 1 indicates the feature selection techniques researchers used to identify the critical variables in detecting patients who need attention.

TABLE I
FEATURE SELECTION TECHNIQUES USED BY RESEARCHERS

Author	Chi-square	Information Gain	Least Absolute Shrinkage and Selection Operator	Boruta	BorutaShap	Stepwise Backward Selection	Ensemble Predictor Selection	Analysis of Variance	Mutual Information
[10]				✓					
[15]	✓								
[16]	✓								
[17]	✓								
[18]			✓						
[19]				✓					
[20]							✓		
[21]			✓						
[22]				✓					
[23]				✓					
[24]			✓						
[25]				✓					
[26]					✓				
[27]					✓				
[28]						✓			
[29]		✓							
[30]			✓						
[31]					✓				
[32]	✓							✓	✓
[33]								✓	

To identify the most significant features in diagnosing patients, the researchers applied filter-based feature selection methods such as Chi-square [15], [16], [17]. Chi-square is used to rank the independent features which depends on the class label with the filter-based feature selection method [34], [35], [36]. Additionally, Chi-square, Analysis of Variance (ANOVA), and Mutual Information (MI) are used to identify the factors that were used to determine the people with COVID-19 [32]. The researchers stated that the ANOVA

feature selection method model performed significantly better than other models. Moreover, 150 initial risk factors that were used to determine the hospitalization outcomes among geriatric patients with dementia were reduced into 35 significant risk factors using ANOVA [33].

The stepwise backward selection was applied to identify the predictors of post-COVID-19 [28]. Furthermore, Machine Learning (ML) based ensemble predictor selection (EPS) is applied to rank the variable importance and determine the minimal predictors [20]. Among 26 variables, they identified 10 essential factors contributing to diabetic retinopathy (DR). The researchers used Information Gain (IG) to determine the most significant features to improve chronic kidney disease (CKD) diagnosis [29]. On the other hand, researchers implemented the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm to select the most essential feature in detecting COVID-19 patients who require attention [18], [24], [30]. Additionally, the researchers applied the LASSO algorithm to select the features for detecting heart disease patients [21]. Embedded methods performed less computationally costly than wrapper methods [12], [14]. However, wrapper methods might be better when the dataset contains continuous variables or mixed types of variables [13].

Furthermore, wrapper methods could identify an optimal subset of features that results in better performance when the optimization algorithms with ML are employed. By using Boruta wrapper feature selection algorithms with CART decision tree, [19] have improved the classification accuracy of medical datasets effectively. Moreover, Boruta has been used to identify the significant features of medical datasets [10], [22], [23], [25]. The researchers recommended Boruta to select relevant variables in high-dimensional datasets [37], [38]. However, [31] stated that it is difficult to identify the variables that are close to their best shadow features in Boruta algorithm.

To improve the underlying Boruta, BorutaShap is introduced by using Shapley Additive Explanations (SHAP) rather than using classical metrics of feature importance such as gain [39], [40]. Moreover, the idea of the combination of shadow features and SHAP importance as feature score is efficient in selecting relevant features and eliminating noise [40]. By using BorutaShap, the researchers have identified significant clinical features that were associated with the class labels [26], [27], [31].

B. Machine Learning Techniques in Patient Detection

ML has assisted in effective patient monitoring and data-driven analytics systems that enable the collection of patient information and analysis in predicting health conditions [41]. In hospitals with more than thousands of patients, manual reviews of patient health records are time-consuming, resulting in difficulty in detecting patients [42]. The researchers demonstrated that patients' waiting time and the doctors' idle time are reduced by more than 50% by implementing ML in patient detection [43]. Several techniques have been used to determine patients who need attention, as shown in Table 2.

TABLE II
TOP COMMON MACHINE LEARNING TECHNIQUES USED BY RESEARCHERS

Author	Random Forest	Support Vector Machine	Logistic Regression	K-Nearest Neighbors	Extreme Gradient Boosting	Naive Bayes	Adaptive Boosting
[18]	✓	✓	✓	✓			
[24]					✓		
[36]	✓	✓	✓	✓	✓		
[44]	✓		✓		✓		
[45]	✓	✓					✓
[46]	✓		✓		✓		
[47]	✓						
[48]					✓		
[49]						✓	
[50]		✓	✓			✓	
[51]	✓						
[52]							✓
[53]	✓	✓	✓	✓			
[54]	✓		✓		✓		
[55]			✓				
[56]	✓				✓		
[57]			✓				
[58]	✓			✓			✓

In COVID-19, [47] employed Artificial Neural Network (ANN), Random Forest (RF) and CART to identify COVID-19 patients who are at high risk based on the clinical variables. [53] implemented Support Vector Machine (SVM), ANN, RF, Decision Tree (DT), Logistic Regression (LR), and K-Nearest Neighbors (KNN) to predict the mortality risk in COVID-19 patients. [18] built an ensemble model using LR, SVM, Gradient-boosted Decision Tree (GBDT) and Neural Network (NN) to determine the mortality risk in COVID-19 patients.

Additionally, [36] have effectively assisted in early detection of patients who are symptomatic for COVID-19 testing by applying Multilayer Perceptron (MLP), Gradient Boosting Machine (GBM), DT, RF, Extreme Gradient Boosting (XGBoost), KNN, SVM, and LR. The researchers stated that the proposed ML algorithms could help in COVID-19 patient detection during primary health care. [55] used LR in early detection of both COVID-19 and non-COVID-19 patients with high mortality risk to help physicians in enhancing patient management. Multinomial NB is constructed to identify suspected among citizens [49]. In COVID-19, blood test results may require long turnaround time approximately 3 to 4 hours, leading to a higher risk for the prevalence of the virus among other individuals by the patient without patient detection [52]. To solve this issue, the researchers implemented AdaBoost to predict COVID-19 patients based on their clinical symptoms.

[57] used LR to determine the mortality risk of COVID-19 patients in improving the patient categorization. On the other hand, XGBoost is constructed to determine the mortality and death risk of COVID-19 [24]. The results indicated that XGBoost performed stable performance than multivariable

LR in identifying fatal outcome of patients. Furthermore, another study has been conducted on the need for ICU care in COVID-19 patients. The ICU transfer rate of hospitalized COVID-19 patients is significantly higher than 11% for other hospitalized patients [51]. To efficiently manage the difficulties in frequent COVID-19 clinical assessments, the researchers developed RF to identify the risk of ICU transfer within the next 1 day using electronic medical records (EMR).

In the United States, the death risk during hospitalization reached 40% and the percentage of patients stay in the ICU reached approximately 22% [58]. While ICU decision making is relying on the physiological data interpretation, the researchers stated that it is difficult for medical professionals to make decisions in dynamic changing environment when there is not enough real-time patient information. To solve this problem, the researchers proposed a stacking model with Support Vector Classifier (SVC), RF, KNN, Bootstrap aggregating (Bagging), Light Gradient Boosting Machine (LGBM), AdaBoost as first-level classifier and LGBM as the meta-classifier to predict mortality in patients with heart failure. On the other hand, [44] proposed RF, XGBoost and LR to determine the patient postoperative survival within 30 days after the liver transplantation. To determine the preoperative surgical risks in ICU, [45] proposed ML models such as RF, AdaBoost, and SVM. The researchers demonstrated that the ML methods could perform better than the traditional risk stratification tools.

In ED, [54] detected patients who require attention by applying Multinomial LR, XGBoost, RF and GBDT. [56] developed XGBoost and Deep Neural Network (DNN) to predict patients in need of critical care in ED. To identify patients who have cardiac disease and those who are normal, [48] presented XGBoost to risk-stratify patients based on predicted benefits of clinically relevant evidence-based interventions efficiently. [50] implemented LR, NB and SVM to determine the patients who are higher risk in angiography. [46] proposed LR, RF, and XGBoost to determine the patients who have at high risk of mortality and cardiopulmonary arrest in ED. They developed the models to predict the composite outcome in the first 1 day after the ED triage and demonstrated that the XGBoost performs better than other models. Likewise, [54] demonstrated that XGBoost performed slightly better than LR, RF, and GBDT in identifying the triage levels of patients with suspected cardiovascular disease (CVD).

II. MATERIAL AND METHOD

This section describes the approaches used to identify the important features that could be used to detect patients who need attention, develop, and evaluate predictive models for identifying those who need attention. Figure 1 illustrates the flowcharts of methods used in this work. In this work, data preprocessing is performed to transform the raw data into a useful for predictive models. Furthermore, feature selection is performed to identify the important features that could be used to detect the patients who need attention whereas class imbalance technique is applied to balance the dataset. Several ML models are constructed to identify those who need attention and evaluated before and after hyperparameter tuning.

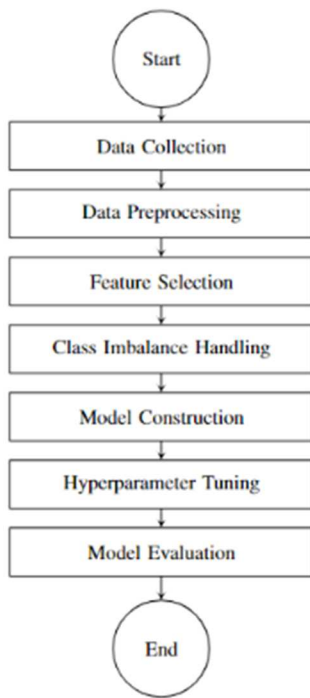


Fig. 1 Flowcharts of Methods

A. Data Source

Selangor Saring Program is a health program, initiated by the Government of Selangor (Kerajaan Negeri Selangor), focused on providing subsidized health screenings for chronic diseases such as diabetes, hypertension, and cardiovascular diseases, primarily for the residents of Selangor. The program's registration and administration process is digitalized and handled solely using the Selangkah application, aiming to increase the program's efficiency. The program was established to provide systematic health screening methods and channels for Selangor residents. Each participant registered in the program would need to answer the Risk Assessment Questionnaire to determine the eligibility for types of screening available.

This work collects data from one of the healthcare companies. The data contains the questionnaire that is answered by each respondent who has registered for the Selangor Saring Program. The data contains 51102 records and 67 features, including the target class. The names of two target classes are *need attention* and *normal*.

TABLE III
FEATURES IN THE DATA COLLECTED

Data	Features
Patient Information	Selangkah ID, age, gender, race, height, weight, the type of screening test taken, and the result status of screening test taken
Questionnaire	Other disease diagnosed, history of Sexually Transmitted Illness (STI), eye history, eye disease, glaucoma, cancer and disease diagnosed by family, nearsighted or farsighted, recipient of Skim Insuran Peduli Sihat / Itizam Sihat Selangor, mammogram interest, married, medical insurance, medication, alcohol consumer, steroids usage, symptoms, enough sleep, allergy, eye trauma history, medical card benefit, weekly steps, annual medical checkup, smoker, type of smoking, disease

diagnosed, autoimmune kidney diagnosed, cancer diagnosed, been to optometry clinic, covid vaccine receiver, menopause, Human Papillomavirus (HPV) vaccine, heard of Selangor Saring Program, commute daily, work time, times per week, spending out-of-pocket per GP visit, spending out-of-pocket per pharmacy visit, recreational area usage, children up-to-date immunization, total children less than 12 years old, between 12 and 17 years old, more than 18 years old, married status, migraine, number of households and children, Raynaud phenomenon, snoring, average income, steroids usage reason, last eye and health checkup, child age, and past medical history.

TABLE IV
BLOOD AND URINE DISTRIBUTION

	Normal	Need Attention	Total
Blood	4063	8943	13006
Urine	11451	2198	13649
Total	15514	11141	26655

Table 3 indicates the features in the data collected where 3 duplicated features are not included, such as the Selangkah ID, eye disease diagnosed by family, and eye trauma history. Table 4 indicates the distribution of blood and urine in the data. The data consists of 13006 instances of blood and 13649 cases of urine. Furthermore, 8943 participants require attention in their blood results, whereas 4063 participants are normal. On the other hand, 2198 participants require attention in their urine results, while 11451 participants are normal.

B. Data Preprocessing

Before applying ML models, the dataset is preprocessed to check for duplicate rows and missing values. The real-world data is seldom clean and complete, especially in the healthcare field and thus, data preprocessing is an essential step to provide processed data to improve the prediction accuracy [59]. The data collected consists of long feature names, which refer to a set of questions that gather information from the respondents. Due to the long feature names in the screening questionnaires, each feature is renamed to shorten the feature names. Furthermore, the data collected contains different tests that the participants have taken. In this work, blood and urine tests are selected as most of the participants take both tests than other tests. In the screening questionnaires, participants can choose multiple responses to a question. Thus, to prevent longer feature names after dummifying, the characters of categorical features such as /, ', and / are removed and placed in lowercase. The English version of the answers is selected.

In addition, participants could input the date and time of their last eye and health checkup. However, some responses that can cause errors will be considered outliers, such as 0022-01-24. To prevent the error, the date and time are replaced with NaT. The Identification Number of participants and duplicated rows or responses are removed. The duplicates should be removed because duplicates can add weight to samples, which can cause bias in ML models.

The data collected contains missing values due to the lack of responses from the respondents. The researchers stated that questionnaires are particularly vulnerable, with missing data being out of the researcher's hands, as respondents may

choose to leave items unanswered [60]. This can lead to performance degradation and biased outcomes as most statistical and ML algorithms are not robust enough to handle missing values [61]. Therefore, missing value imputation is performed as described in this subsection.

First, the dataset is split into two sets: blood and urine tests. Let D_{blood} and D_{urine} be the dataset that contains participants who take blood and urine tests, respectively. For each test, the row will be removed if the row contains no response from the participants in the screening questionnaires. Table 5 shows the distribution of missing values in D_{blood} and D_{urine} where G_1 , G_2 , and G_3 are the number of children who are less than 12 years old, between 12 and 17 years old, and more than 18 years old, respectively.

TABLE V
DISTRIBUTION OF MISSING VALUES IN D_{blood} AND D_{urine}

	Height	Weight	G_1	G_2	G_3
D_{blood}	7166	7106	10165	11067	9409
D_{urine}	7465	7391	10647	11580	9774

Based on Table 5, at least 50% of D_{blood} and D_{urine} have missing values for height, weight, G_1 , G_2 , and G_3 . Besides that, more than 70% of D_{blood} and D_{urine} have missing values for G_1 , G_2 , and G_3 . This indicates that most respondents might choose to leave the items unanswered for G_1 , G_2 , and G_3 . In this project, the missing values in age, height, and weight are filled with its median. On the other hand, the missing value in the number of children groups is filled with 0 to indicate that there is no child in the corresponding household. The missing value in the number of households is filled with 1 to indicate the respondent. The features are removed if the features contain empty values for each respondent.

C. Feature Selection

The feature selection uses BorutaShap to select the features that explain the dataset with SHAP values. Firstly, the dataset is split into 70% training and 30% testing data using the result status as the target variable. Stratification is used during the splitting to ensure that the proportion of values in the sample produced is the same as the proportion of values provided in the target variable. The training data and testing data is scaled using Standard Scaler to avoid bias towards the features. RF with balanced class weight is used as an estimator of the BorutaShap.

D. Class Imbalance Treatment

D_{blood} consists of 4063 normal and 8943 need attention. On the other hand, D_{urine} consists of 11451 normal and 2198 need attention. In D_{blood} and D_{urine} , the class distributions are not balanced. The predictive models will be biased towards the majority class as the data sets have higher proportions of majority class. As a result, the minority classes are most likely to be misclassified. Hence, class imbalance treatment is performed to alleviate this problem. First, the data set is split into 70% training and 30% testing data using the result status as the target variable. Stratification is used during the splitting. SMOTE-Tomek is employed to resolve the class imbalance among *normal* and *need-attention* classes. The minority class is oversampled, whereas the majority class is undersampled, so the participants are equally distributed. The

training and testing data are scaled using Standard Scaler to avoid bias towards the features.

E. Model Construction

In this work, predictive models are constructed to identify the patients who require attention. The binary class variable in the models reflects whether a patient requires attention. The classification models used are LR, Gaussian NB, RF, AdaBoost, XGBoost, and Stacking classifier. Balanced class weight is applied on LR, and RF to adjust the weights inversely proportional to class frequencies in the input data. In this work, the stacking classifier, as shown in Figure 2, comprises Nu-SVC, KNN, Gaussian NB, and XGBoost as the base classifiers, and LDA as the meta classifier.

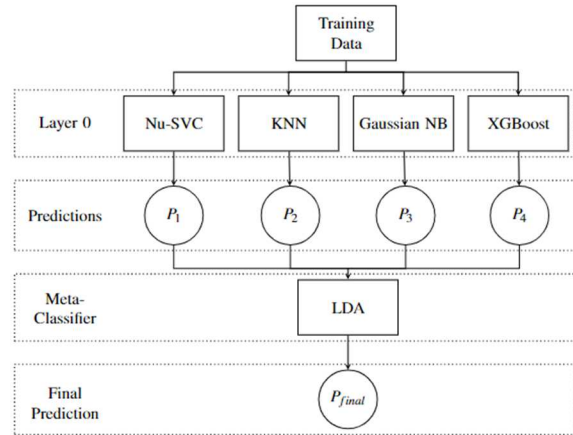


Fig. 2 Stacking Classifier

F. Hyperparameter Tuning

Table 6 indicates the specified hyperparameter values to be searched in each predictive model. Hyperparameter tuning is choosing a set of optimal hyperparameters for a predictive model. In this project, grid-based hyperparameter tuning, GridSearchCV, is carried out to perform an exhaustive search over specified hyperparameter values for a predictive model. Besides that, a cross-validated grid search is conducted over a parameter grid to optimize the performance of a predictive model.

TABLE VI
SPECIFIED HYPERPARAMETER VALUES SEARCH

Model	Hyperparameter Tuning	Search Value	
LR	penalty	11, 12	
	C	0.0001, 0.01, 1, 100, 10000	
Gaussian NB	solver	liblinear	
	var_smoothing	1, 0.0056, 3.162×10^{-5} , 1.7782×10^{-7} , 1×10^{-9}	
RF	max_features	6, 11, 16, 21, 26, 31	
	max_depth	3, 4, 5, 6, 7	
AdaBoost	n_estimators	50, 100	
	learning_rate	0.001, 0.1	
XGBoost	reg_alpha	1, 2	
	reg_lambda	1, 2	
	min_child_weight	1, 2, 3, 4, 5, 6, 7	
Stacking Classifier	Nu-SVC	nu	0.1, 0.2, 0.3, 0.4, 0.5, 0.6
		KNN	n_neighbors

Model	Hyperparameter Tuning	Search Value
Gaussian NB	var_smoothing	1, 0.0056, 3.162×10^{-5} , 1.7782×10^{-7} , 1×10^{-9}
XGBoost	reg_alpha	1, 2
	reg_lambda	1, 2
	min_child_weight	1, 2, 3, 4, 5, 6, 7
LDA	solver	svd, lsqr, eigen

G. Model Evaluation

This work evaluates predictive models based on accuracy, precision, recall, F1-score, and Geometric Mean (G-Mean). Besides that, Area Under the Curve (AUC) is computed to indicate the trade-off between correctly predicted positive classes and incorrectly predicted negative classes. True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) are four measurements used in the evaluation metric. TP records the total number of positive classes that are identified correctly. On the other hand, TN refers to the total number of negative classes that are identified correctly. If the negative class is predicted as positive, it is known as FP. However, FN refers to the total number of positive classes predicted as negative. For further comparison and evaluation, Confidence Interval (CI) is used to estimate the performance of a ML model on unseen data or a likelihood over a range. Furthermore, ANOVA and Tukey’s Honestly Significant Different (HSD) test are performed.

A CI refers to an interval statistic used to quantify the uncertainty on an estimate. This work applies CI to present the likelihood of ML model performance over a range. For example, CI could indicate a 95% likelihood that the range x to y covers the true model accuracy. The larger the confidence interval, the larger the margin of error. In this work, each prediction made by an ML model is a binary decision that indicates normal or needs *attention*. This is known as Bernoulli trial. The binomial distribution can be approximated with a Gaussian distribution for larger sample sizes. The CI of each ML model is calculated using the number of correct predictions and instances.

Stratified 5-folds cross validation is performed to produce different sets in the statistical hypothesis test. For each fold, SMOTE-Tomek and Standard Scaler are used to balance and scale the sample distribution. The ANOVA test is performed to determine the significant difference among ML models in terms of accuracy. Null hypothesis, H_0 and alternative hypothesis, H_a of ANOVA tests are formed as below.

H_0 : There is no significant difference among ML models in terms of accuracy.

H_a : There is a significant difference among ML models in terms of accuracy.

If the p-value obtained is higher than 0.05, the H_0 is failed to reject. Thus, multiple pairwise comparison analysis using Tukey’s HSD test is performed to identify the pairs of significant differences. This is because the ANOVA test only concludes that ML models have a significant difference in accuracy rather than determining which pairs have a significant difference.

In this section, the important features that contributed to detect patients who need attention are discussed when the D_{blood} and D_{urine} were used. Furthermore, the performance of each ML model is presented based on accuracy, precision, recall, F1-score, G-Mean and AUC before and after hyperparameter tuning using the important features. Each model is then further evaluated using CI, ANOVA, and Tukey’s HSD test based on accuracy after hyperparameter tuning.

A. Features Contributing Need-Attention in Patient Detection

The original dataset contains 66 features, excluding the target class. After dummifying, the dataset contains 143 features, which could increase to high computational complexity. In this work, BorutaShap is applied to identify the important features which contribute to detecting patients who need attention. As a result, 23 and 22 significant features are identified in D_{blood} and D_{urine} respectively. In D_{blood} and D_{urine} , BorutaShap has identified age, symptoms such as obesity, snoring, weight, and children’s up-to-date immunization as significant features in detecting patients who need attention. In D_{blood} , BorutaShap identified gender, mammogram interest, menopause, HPV vaccine, the total number of children that are more than 18 years old, number of children, and the child’s age as the crucial factors. On the other hand, BorutaShap identified race, average income, number of households, work time, medical insurance, alcohol consumption, steroid usage, and number of children that are between 12 and 17 years old as the essential features in detecting patients who need attention in D_{urine} . Furthermore, the type of daily commutation such as motorcycle, steroids usage, been to optometry clinic, eye trauma history, married status, and the type of disease diagnosed, such as hypertension and diabetes, have been identified as other significant features to detect the patients who need attention in D_{urine} .

B. Comparison of Model Performance

Table 7 indicates the model performance when D_{blood} and D_{urine} was used before and after hyperparameter tuning. When D_{blood} was used before hyperparameter tuning, AdaBoost outperformed all other models, by obtaining the highest accuracy (69.7%), highest recall (92.0%), highest F1-score (80.7%) and highest AUC (64.0%). LR outperformed other models in balancing the classification performance in both majority and minority classes by obtaining the highest G-Mean (60.3%). After hyperparameter tuning, there is improvement in AUC of all models except Gaussian NB and XGBoost. Besides that, there has been an improvement in the accuracy of Gaussian NB, AdaBoost, and the stacking classifier. After hyperparameter tuning, the Stacking Classifier obtained the highest accuracy (69.9%). Although LR achieved the highest G-Mean (60.3%) before hyperparameter tuning, RF achieved the highest G-Mean (61.0%) by performing hyperparameter tuning.

TABLE VII
COMPARISON OF MODEL PERFORMANCE

Dataset	Hyperparameter Tuning	Model	Accuracy	Precision	Recall	F1-Score	G-Mean	AUC
D_{blood}	No	LR	0.617	0.765	0.639	0.697	0.603	0.631
		Gaussian NB	0.496	0.757	0.392	0.517	0.533	0.597
		RF	0.661	0.727	0.813	0.767	0.516	0.609
		AdaBoost	0.697	0.718	0.920	0.807	0.435	0.640
		XGBoost	0.683	0.725	0.870	0.791	0.487	0.627
	Stacking Classifier	0.697	0.722	0.908	0.804	0.459	0.632	
	Yes	LR	0.616	0.765	0.638	0.696	0.602	0.632
		Gaussian NB	0.586	0.753	0.592	0.663	0.583	0.597
		RF	0.655	0.767	0.717	0.741	0.610	0.657
		AdaBoost	0.698	0.712	0.942	0.811	0.390	0.642
XGBoost		0.685	0.728	0.864	0.790	0.501	0.627	
Stacking Classifier	0.699	0.727	0.902	0.805	0.477	0.633		
D_{urine}	No	LR	0.660	0.233	0.484	0.314	0.580	0.615
		Gaussian NB	0.678	0.254	0.514	0.340	0.604	0.631
		RF	0.803	0.286	0.150	0.197	0.373	0.611
		AdaBoost	0.792	0.285	0.193	0.230	0.418	0.618
		XGBoost	0.821	0.360	0.149	0.211	0.376	0.618
	Stacking Classifier	0.816	0.342	0.153	0.212	0.380	0.623	
	Yes	LR	0.661	0.233	0.484	0.315	0.580	0.615
		Gaussian NB	0.694	0.260	0.489	0.339	0.599	0.636
		RF	0.703	0.256	0.442	0.324	0.577	0.631
		AdaBoost	0.800	0.296	0.178	0.222	0.404	0.636
XGBoost		0.817	0.343	0.150	0.209	0.377	0.609	
Stacking Classifier	0.811	0.314	0.150	0.203	0.375	0.618		

When D_{urine} was used before hyperparameter tuning, XGBoost achieved the highest accuracy (82.1%) and precision (36.0%). After hyperparameter tuning, the AUC of Gaussian NB, RF, and AdaBoost improved. Besides, after hyperparameter tuning, XGBoost has the highest accuracy (81.7%) and precision (34.3%). Nevertheless, Gaussian NB achieved the highest recall (48.9%) in correctly identifying patients who need attention.

TABLE VIII
SPECIFIED HYPERPARAMETER VALUES SEARCH

Model	Hyperparameter	Best Hyperparameter Value	
		D_{blood}	D_{urine}
LR	penalty	11	11
	C	100	1
Gaussian NB	solver	liblinear	liblinear
	var_smoothing	1	1
RF	max_features	11	11
	max_depth	7	7
AdaBoost	n_estimators	100	100
	learning_rate	0.1	0.1
XGBoost	reg_alpha	1	1
	reg_lambda	2	1
	min_child_weight	1	2
Stacking Classifier			
Nu-SVC	nu	0.6	0.5
KNN	n_neighbors	5	2
Gaussian NB	var_smoothing	1	1
XGBoost	reg_alpha	1	1
	reg_lambda	2	2
	min_child_weight	1	1
LDA	solver	svd	svd

Table 8 indicates the best hyperparameter value of each predictive model returned by GridSearchCV when D_{blood} and D_{urine} was used. By using these best hyperparameter values, each predictive model is evaluated when D_{blood} and D_{urine} were used. For further analysis of the model performance, 95% CI is computed on the number of correct predictions by each model after hyperparameter tuning when D_{blood} and D_{urine} were used.

TABLE IX
CI OF ML MODELS BASED ON NUMBER OF CORRECT PREDICTIONS

Model	D_{blood}		D_{urine}	
	Lower	Upper	Lower	Upper
LR	0.601	0.632	0.646	0.675
Gaussian NB	0.571	0.602	0.680	0.708
RF	0.640	0.670	0.689	0.717
AdaBoost	0.684	0.713	0.787	0.812
XGBoost	0.670	0.699	0.805	0.829
Stacking Classifier	0.685	0.714	0.798	0.823

There is a 95% likelihood that the range 68.5% to 71.4% covers the true accuracy of the Stacking Classifier when D_{blood} was used while there is a 95% likelihood that the range 80.5% to 82.9% covers the true accuracy of XGBoost when D_{urine} was used. Besides, every model has an approximately similar margin of error and precision when D_{blood} and D_{urine} were used. Statistical hypothesis tests such as ANOVA and Tukey's HSD test are performed on the accuracies recorded of each fold after hyperparameter tuning and stratified 5-fold cross-validation. When D_{blood} was used, the p-value computed was smaller than the significant level (0.05) based on the p-value (approximately 0) and F value (91.731). Hence, the null hypothesis is rejected, and there is a difference between the classification accuracies. By performing Tukey's HSD test, the results showed that the Stacking Classifier performed best in accuracy. There is a significant difference

($p < 0.05$) from other models except AdaBoost and XGBoost. Although there is no significant difference between Stacking Classifier and AdaBoost and XGBoost, the average accuracy of Stacking Classifier is higher than AdaBoost and XGBoost.

When D_{urine} was used, the p-value computed is smaller than the significant level (0.05) based on the p-value (approximately 0) and F value (8.104). Hence, the null hypothesis is rejected, and there is a difference between the classification accuracies. Tukey's HSD test indicated that the XGBoost performed best based on accuracy, and there is a significant difference ($p < 0.05$) from other models except AdaBoost and Stacking Classifier. Although there is no significant difference between XGBoost and AdaBoost and Stacking Classifier, the average accuracy of XGBoost is higher than that of AdaBoost and Stacking Classifier.

IV. CONCLUSION

In this study, BorutaShap was used to select the essential features for detecting patients who need attention. As a result, BorutaShap has identified 23 and 22 significant features in D_{blood} and D_{urine} respectively. BorutaShap identified the common significant features: age, symptoms such as obesity, snoring, weight, and children's up-to-date immunization. In D_{blood} , BorutaShap identified gender, mammogram interest, menopause, HPV vaccine, the total number of children that are more than 18 years old, the number of children, and the child's age as the crucial factors. On the other hand, BorutaShap identified race, average income, number of households, work time, medical insurance, alcohol consumption, steroid usage, and number of children that are between 12 and 17 years old as the essential features in detecting patients who need attention in D_{urine} . Furthermore, the type of daily commutation such as motorcycle, steroids usage, been to optometry clinic, eye trauma history, married status, and the type of disease diagnosed, such as hypertension and diabetes, have been identified as other significant features to detect the patients who need attention in D_{urine} . LR, Gaussian NB, RF, AdaBoost, XGBoost, and Stacking Classifier were implemented to detect those who need attention. The results showed that the Stacking Classifier achieved the highest accuracy (69.9%) when the blood dataset was used, whereas XGBoost achieved the highest accuracy (81.7%) when the urine dataset was used. To further advance the performance of detecting need-attention patients, future investigations could focus on integrating Point of Interest (POI) and geographical data relevant to patient's living environments. Additionally, the exploration of diverse ensemble models could hold the potential to enhance the prediction performance significantly.

REFERENCES

- [1] A. A. Khorana *et al.*, "Time to initial cancer treatment in the United States and association with survival over time: an observational study," *PLoS One*, vol. 14, no. 3, p. e0213209, 2019.
- [2] T. P. Hanna *et al.*, "Mortality due to cancer treatment delay: systematic review and meta-analysis," *BMJ*, p. m4087, Nov. 2020, doi:10.1136/bmj.m4087.
- [3] K. A. Fleming *et al.*, "The Lancet Commission on diagnostics: transforming access to diagnostics," *The Lancet*, vol. 398, no. 10315, pp. 1997–2050, 2021.
- [4] M. A. F. Pimentel *et al.*, "Detecting Deteriorating Patients in the Hospital: Development and Validation of a Novel Scoring System," *Am J Respir Crit Care Med*, vol. 204, no. 1, pp. 44–52, Jul. 2021.
- [5] N.-C. Shih, P.-T. Kung, W.-Y. Kuo, and W.-C. Tsai, "Association of treatment delay and stage with mortality in breast cancer: a nationwide cohort study in Taiwan," *Scientific Reports*, vol. 12, no. 1, Nov. 2022, doi: 10.1038/s41598-022-23683-y.
- [6] N. Zaki, H. Alashwal, and S. Ibrahim, "Association of hypertension, diabetes, stroke, cancer, kidney disease, and high-cholesterol with COVID-19 disease severity and fatality: A systematic review," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 5, pp. 1133–1142, Sep. 2020, doi:10.1016/j.dsx.2020.07.005.
- [7] Y. Gao *et al.*, "Risk factors for severe and critically ill COVID-19 patients: a review," *Allergy*, vol. 76, no. 2, pp. 428–455, 2021.
- [8] A. Shrestha, C. Martin, M. Burton, S. Walters, K. Collins, and L. Wyld, "Quality of life versus length of life considerations in cancer patients: a systematic literature review," *Psychooncology*, vol. 28, no. 7, pp. 1367–1380, 2019.
- [9] T. Chandrasekar, S. A. Bootjian, U. Capitanio, B. Gershman, M. C. Mir, and A. Kutikov, "Collaborative Review: Factors Influencing Treatment Decisions for Patients with a Localized Solid Renal Mass," *European Urology*, vol. 80, no. 5, pp. 575–588, Nov. 2021, doi:10.1016/j.eururo.2021.01.021.
- [10] J. Mayneris-Perxachs *et al.*, "Blood Hemoglobin Substantially Modulates the Impact of Gender, Morbid Obesity, and Hyperglycemia on COVID-19 Death Risk: A Multicenter Study in Italy and Spain," *Front Endocrinol (Lausanne)*, vol. 12, p. 741248, 2021.
- [11] J. Déry *et al.*, "A systematic review of patient prioritization tools in non-emergency healthcare services," *Systematic Reviews*, vol. 9, no. 1, Oct. 2020, doi: 10.1186/s13643-020-01482-8.
- [12] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in Biology and Medicine*, vol. 112, p. 103375, Sep. 2019, doi:10.1016/j.compbiomed.2019.103375.
- [13] E. Hancer, B. Xue, and M. Zhang, "A survey on feature selection approaches for clustering," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4519–4545, Jan. 2020, doi: 10.1007/s10462-019-09800-w.
- [14] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 1, pp. 56–70, May 2020, doi:10.38094/jastt1224.
- [15] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring feature selection and classification methods for predicting heart disease," *Digital Health*, vol. 6, p. 205520762091477, Jan. 2020, doi: 10.1177/2055207620914777.
- [16] S. Ray, K. Alshouily, A. Roy, A. AlGhamdi, and D. P. Agrawal, "Chi-Squared Based Feature Selection for Stroke Prediction using AzureML," *2020 Intermountain Engineering, Technology and Computing (IETC)*, Oct. 2020, doi: 10.1109/ietc47856.2020.9249117.
- [17] M. F. Ijaz, M. Attique, and Y. Son, "Data-Driven Cervical Cancer Prediction Model with Outlier Detection and Over-Sampling Methods," *Sensors*, vol. 20, no. 10, p. 2809, May 2020, doi:10.3390/s20102809.
- [18] G. Yue *et al.*, "Machine learning based early warning system enables accurate mortality risk prediction for COVID-19," *Nat Commun*, vol. 11, no. 1, p. 5033, 2020.
- [19] R. Tang and X. Zhang, "CART Decision Tree Combined with Boruta Feature Selection for Medical Data Classification," *2020 5th IEEE International Conference on Big Data Analytics (ICBDA)*, May 2020, doi: 10.1109/icbda49040.2020.9101199.
- [20] R. Wang *et al.*, "Derivation and Validation of Essential Predictors and Risk Index for Early Detection of Diabetic Retinopathy Using Electronic Health Records," *Journal of Clinical Medicine*, vol. 10, no. 7, p. 1473, Apr. 2021, doi: 10.3390/jcm10071473.
- [21] V. S. Narayanan, K. N. Raj, K. Kumar, and M. Kumar, "Survival Prediction of Heart Failure Patients Using Lasso Algorithm and Gaussian Naive Bayes Classifier," 2021.
- [22] Khan *et al.*, "Examining the prevalence of hypertension by urban–rural stratification: A Cross-sectional study of nepal demographic and health survey," *Asian Journal of Social Health and Behavior*, vol. 4, no. 1, p. 15, 2021.
- [23] L. Wang, Y. Zhou, F. Wang, L. Ding, P. E. D. Love, and S. Li, "The Influence of the Built Environment on People's Mental Health: An Empirical Classification of Causal Factors," *Sustainable Cities and Society*, vol. 74, p. 103185, Nov. 2021, doi:10.1016/j.scs.2021.103185.
- [24] X. Guan *et al.*, "Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19

- patients: results from a retrospective cohort study,” *Annals of Medicine*, vol. 53, no. 1, pp. 257–266, Jan. 2021, doi:10.1080/07853890.2020.1868564.
- [25] V. Maeda-Gutiérrez et al., “Risk-Profile and Feature Selection Comparison in Diabetic Retinopathy,” *Journal of Personalized Medicine*, vol. 11, no. 12, p. 1327, Dec. 2021, doi:10.3390/jpm11121327.
- [26] S. A. Antor, J. Ahmed, Z. Nayen, F. Tabassum, and R. Mahbub, “A feature selection approach to determine obesity using machine learning method,” *Brac University*, 2021.
- [27] J. Emakhu et al., “Acute coronary syndrome prediction in emergency care: A machine learning approach,” *Computer Methods and Programs in Biomedicine*, vol. 225, p. 107080, Oct. 2022, doi:10.1016/j.cmpb.2022.107080.
- [28] C. Förster, M. G. Colombo, A.-J. Wetzel, P. Martus, and S. Joos, “Persisting Symptoms After COVID-19,” *Dtsch Arztebl Int*, vol. 119, no. 10, pp. 167–174, 2022.
- [29] S. A. Ebiaredoh-Mienye, T. G. Swart, E. Esenogho, and I. D. Mienye, “A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease,” *Bioengineering*, vol. 9, no. 8, p. 350, Jul. 2022, doi: 10.3390/bioengineering9080350.
- [30] M. R. Afrash, H. Kazemi-Arpanahi, M. Shanbehzadeh, R. Nopour, and E. Mirbagheri, “Predicting hospital readmission risk in patients with COVID-19: A machine learning approach,” *Informatics in Medicine Unlocked*, vol. 30, p. 100908, 2022, doi: 10.1016/j.imu.2022.100908.
- [31] J. Kim, J. Lee, and M. Park, “Identification of Smartwatch-Collected Lifelog Variables Affecting Body Mass Index in Middle-Aged People Using Regression Machine Learning Algorithms and SHapley Additive Explanations,” *Applied Sciences*, vol. 12, no. 8, p. 3819, Apr. 2022, doi: 10.3390/app12083819.
- [32] H. Chauhan, K. Modi, and S. Shrivastava, “Development of a classifier with analysis of feature selection methods for COVID-19 diagnosis,” *World Journal of Engineering*, vol. 19, no. 1, pp. 49–57, 2022.
- [33] Wang et al., “Risk factors and machine learning model for predicting hospitalization outcomes in geriatric patients with dementia,” *Alzheimer’s & Dementia: Translational Research & Clinical Interventions*, vol. 8, no. 1, p. e12351, 2022.
- [34] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrés, “Classification models for heart disease prediction using feature selection and PCA,” *Informatics in Medicine Unlocked*, vol. 19, p. 100330, 2020, doi: 10.1016/j.imu.2020.100330.
- [35] S. Sawangarereak and P. Thanathamathee, “Random Forest with Sampling Techniques for Handling Imbalanced Prediction of University Student Depression,” *Information*, vol. 11, no. 11, p. 519, Nov. 2020, doi: 10.3390/info11110519.
- [36] Í. Viana dos Santos Santana et al., “Classification Models for COVID-19 Test Prioritization in Brazil: Machine Learning Approach,” *Journal of Medical Internet Research*, vol. 23, no. 4, p. e27293, Apr. 2021, doi:10.2196/27293.
- [37] F. Degenhardt, S. Seifert, and S. Szymczak, “Evaluation of variable selection methods for random forests and omics data sets,” *Brief Bioinform*, vol. 20, no. 2, pp. 492–503, 2019.
- [38] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, “A comparison of random forest variable selection methods for classification prediction modeling,” *Expert Systems with Applications*, vol. 134, pp. 93–101, Nov. 2019, doi: 10.1016/j.eswa.2019.05.028.
- [39] J. Verhaeghe, J. Van Der Donckt, F. Ongenaes, and S. Van Hoecke, “Powershap: A Power-full Shapley Feature Selection Method.” *arXiv*, 2022. doi: 10.48550/ARXIV.2206.08394.
- [40] A. Gramegna and P. Giudici, “Shapley Feature Selection,” *FinTech*, vol. 1, no. 1, pp. 72–80, Feb. 2022, doi: 10.3390/fintech1010006.
- [41] D. Chang, D. Chang, and M. Pourhomayoun, “Risk Prediction of Critical Vital Signs for ICU Patients Using Recurrent Neural Network,” 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Dec. 2019, doi:10.1109/csci49370.2019.00191.
- [42] P. Radha and B. MeenaPreethi, “Machine learning approaches for disease prediction from radiology and pathology reports,” *J Green Eng*, vol. 9, no. 2, pp. 149–166, 2019.
- [43] S. Srinivas and H. Salah, “Consultation length and no-show prediction for improving appointment scheduling efficiency at a cardiology clinic: A data analytics approach,” *International Journal of Medical Informatics*, vol. 145, p. 104290, Jan. 2021, doi:10.1016/j.ijmedinf.2020.104290.
- [44] C.-L. Liu, R.-S. Soong, W.-C. Lee, G.-W. Jiang, and Y.-C. Lin, “Predicting Short-term Survival after Liver Transplantation using Machine Learning,” *Scientific Reports*, vol. 10, no. 1, Mar. 2020, doi:10.1038/s41598-020-62387-z.
- [45] C. J. Chiew, N. Liu, T. H. Wong, Y. E. Sim, and H. R. Abdullah, “Utilizing machine learning methods for preoperative prediction of postsurgical mortality and intensive care unit admission,” *Ann Surg*, vol. 272, no. 6, p. 1133, 2020.
- [46] M. Fernandes et al., “Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing,” *Plos One*, vol. 15, no. 4, p. e0230876, Apr. 2020, doi: 10.1371/journal.pone.0230876.
- [47] D. Assaf et al., “Utilization of machine-learning models to accurately predict the risk for critical COVID-19,” *Internal and Emergency Medicine*, vol. 15, no. 8, pp. 1435–1443, Aug. 2020, doi:10.1007/s11739-020-02475-0.
- [48] L. Jing et al., “A Machine Learning Approach to Management of Heart Failure Populations,” *JACC: Heart Failure*, vol. 8, no. 7, pp. 578–587, Jul. 2020, doi: 10.1016/j.jchf.2020.01.012.
- [49] J. Awwalu, N. A. Umar, M. S. Ibrahim, and O. F. Nonyelum, “A Multinomial Naïve Bayes Decision Support System for COVID-19 Detection,” *Fudma Journal Of Sciences*, vol. 4, no. 2, pp. 704–711, Oct. 2020, doi: 10.33003/fjs-2020-0402-331.
- [50] P. Golpour et al., “Comparison of Support Vector Machine, Naïve Bayes and Logistic Regression for Assessing the Necessity for Coronary Angiography,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, p. 6449, Sep. 2020, doi:10.3390/ijerph17186449.
- [51] F.-Y. Cheng et al., “Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients,” *Journal of Clinical Medicine*, vol. 9, no. 6, p. 1668, Jun. 2020, doi: 10.3390/jcm9061668.
- [52] M. Soui, N. Mansouri, R. Alhamad, M. Kessentini, and K. Ghedira, “NSGA-II as feature selection technique and AdaBoost classifier for COVID-19 prediction using patient’s symptoms,” *Nonlinear Dynamics*, vol. 106, no. 2, pp. 1453–1475, May 2021, doi:10.1007/s11071-021-06504-1.
- [53] M. Pourhomayoun and M. Shakibi, “Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making,” *Smart Health*, vol. 20, p. 100178, Apr. 2021, doi:10.1016/j.smhl.2020.100178.
- [54] H. Jiang et al., “Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease,” *International Journal of Medical Informatics*, vol. 145, p. 104326, Jan. 2021, doi: 10.1016/j.ijmedinf.2020.104326.
- [55] T. Rahman et al., “Mortality Prediction Utilizing Blood Biomarkers to Predict the Severity of COVID-19 Using Machine Learning Technique,” *Diagnostics*, vol. 11, no. 9, p. 1582, Aug. 2021, doi:10.3390/diagnostics11091582.
- [56] H. Yun, J. Choi, and J. H. Park, “Prediction of Critical Care Outcome for Adult Patients Presenting to Emergency Department Using Initial Triage Information: An XGBoost Algorithm Analysis,” *JMIR Medical Informatics*, vol. 9, no. 9, p. e30770, Sep. 2021, doi: 10.2196/30770.
- [57] M. E. H. Chowdhury et al., “An Early Warning Tool for Predicting Mortality Risk of COVID-19 Patients Using Machine Learning,” *Cognitive Computation*, Apr. 2021, doi: 10.1007/s12559-020-09812-7.
- [58] C.-C. Chiu, C.-M. Wu, T.-N. Chien, L.-J. Kao, C. Li, and H.-L. Jiang, “Applying an Improved Stacking Ensemble Model to Predict the Mortality of ICU Patients with Heart Failure,” *Journal of Clinical Medicine*, vol. 11, no. 21, p. 6460, Oct. 2022, doi:10.3390/jcm11216460.
- [59] P. Misra and A. S. Yadav, “Impact of preprocessing methods on healthcare predictions,” in *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*, 2019.
- [60] M. R. Stavseth, T. Clausen, and J. Roislien, “How handling missing data may impact conclusions: A comparison of six different imputation methods for categorical questionnaire data,” *SAGE Open Med*, vol. 7, p. 2050312118822912, 2019.
- [61] A. Jadhav, D. Pramod, and K. Ramanathan, “Comparison of Performance of Data Imputation Methods for Numeric Dataset,” *Applied Artificial Intelligence*, vol. 33, no. 10, pp. 913–933, Jul. 2019, doi: 10.1080/08839514.2019.1637138.