Vol. 8, No. 3, June 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Hypertension Classification Using Correlation-Based Feature Selection (CFS) with Random Forest, XGBoost, and Support Vector Machine: A Comparative Study on Indonesian Hospital Data

Faradillah[1*], Muhammad Fadhiel Alie[2], Herri Setiawan[3], Atthiyah Gisca Ahsya[4]

[1,2,3)] Faculty of Computer Science and Natural Science, Universitas Indo Global Mandiri, Palembang, Indonesia

[4)]Faculty of Medicine, Andalas University, Padang, Indonesia

Received:
October 7, 2025
Revised:
May 10, 2026
Accepted:
May 27 2026
Published:
June 22, 2026

Corresponding Author:
Author Name*:
Faradillah
Email*:
faradillah.hakim@uigm.ac.id

DOI:
10.63158/journalisi.v8i3.1646

Abstract: Hypertension is a major global health problem that significantly contributes to cardiovascular disease and mortality. This study evaluates the performance of Random Forest, XGBoost, and Support Vector Machine (SVM) algorithms integrated with Correlation-Based Feature Selection (CFS) for hypertension classification using hospital clinical data. The dataset comprises 500 clinical records containing demographic and physiological variables. CFS was applied to reduce irrelevant and redundant attributes before model training. Model performance was assessed using accuracy, precision, recall, F1-score, and AUC-ROC through 10-fold cross-validation. Statistical significance was examined using the Friedman test followed by the Wilcoxon signed-rank test with Bonferroni correction. The results show that CFS improved classification performance across all models by approximately 5–6%. XGBoost achieved the best performance with 93.5% accuracy and 0.96 AUC, followed by Random Forest and SVM. However, systolic and diastolic blood pressure, which define the hypertension label, were retained as predictors, indicating a diagnostic classification design rather than independent risk prediction. Therefore, the findings should be interpreted as dataset-based hypertension classification, not future hypertension risk prediction.

Keywords: Hypertension Classification; Correlation-Based Feature Selection; Label Leakage; Clinical Data Mining; Machine Learning.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Hypertension, commonly referred to as elevated blood pressure, is among the most prevalent chronic health disorders globally and serves as a primary modifiable determinant contributing to cardiovascular complications such as stroke and heart failure [1], [2]. According to the World Health Organization (WHO), approximately 1.28 billion individuals aged 30–79 years are affected, with nearly half remaining undiagnosed or inadequately managed. In developing nations including Indonesia, the lack of accessible and regular health screening amplifies the prevalence of hypertension, highlighting the need for data-driven approaches to facilitate efficient identification and clinical decision support.

Machine learning (ML) has gained significant attention as an effective methodology in healthcare analytics, utilizing clinical and behavioral datasets to support hypertension-related diagnostic classification [3], [4], [5]. However, the performance of ML models depends heavily on the relevance and quality of input features [5], [6], [7]. High-dimensional or redundant features can degrade accuracy, increase computational cost, and reduce model interpretability — a critical factor for clinical adoption[7], [8]. Feature selection techniques mitigate these issues by identifying the most informative and non-redundant attributes [8], [9]. Among these, Correlation-Based Feature Selection (CFS) has demonstrated strong potential in health-related applications by selecting features highly correlated with the target variable while minimizing inter-feature redundancy [10], [11], [12], [13], [14], [15], [16]. An important consideration in medical ML is the risk of label circularity, where features that directly define the outcome label are included as predictors, inflating apparent model performance[5], [17], [18]. This study explicitly discloses and analyzes this limitation rather than treating it as a solved problem.

Ensemble learning techniques have been extensively utilized in predictive modeling for various diseases [16], [19]. Recent advances include Du et al.[3], who achieved 91.2% accuracy using Random Forest on health check-up data but without lightweight feature reduction. Bisong et al. [15] applied XGBoost for hypertension classification in Nigeria (89.7%) but without filter-based selection, limiting scalability in primary care. Zhao et al. [4]used SVM with only seven risk factors but reported lower accuracy (86.4%), underscoring the trade-off between simplicity and performance. Noroozi et al.

[2]demonstrated CFS utility in heart disease prediction but did not address hypertension in Indonesian populations. Nematollahi et al. [20] achieved 92.1% with XGBoost using body composition metrics, but their approach required specialized equipment unavailable in rural clinics. Based on this review, a clear research gap exists in the systematic evaluation of CFS integrated with multiple ML classifiers for hypertension classification using real-world Indonesian clinical data, with transparent disclosure of label-circularity effects. Prior studies have either employed single-model frameworks, used wrapper-based selection, or focused on datasets from other geographic contexts.

This study frames its contribution explicitly as diagnostic classification, not future hypertension risk prediction. Future-risk prediction would require a prospectively constructed outcome variable and exclusion of direct diagnostic measures (SBP, DBP) from the predictor set [21], [22]. This study addresses the identified gap with the following specific contributions: (1) a systematic comparison of CFS-enhanced classification performance across RF, XGBoost, and SVM under consistent evaluation conditions using Indonesian hospital data; (2) full disclosure that SBP and DBP serve both as diagnostic criteria for the outcome label and as selected predictor features — positioning this work explicitly as diagnostic classification rather than independent risk-factor prediction with empirical sensitivity analysis confirming their dominant contribution; and (3) a reproducible and computationally efficient CFS as a comparative baseline for future studies that may extend the design to exclude direct diagnostic variables from the predictor set.

2. METHODS

The three algorithms—SVM (margin-based classifier), XGBoost (ensemble boosting), and RF (ensemble bagging)—represent various learning paradigms. Every model is trained and assessed using the same experimental setup, which includes the same cross-validation folds, preprocessing pipeline, data splits, and hyperparameter tweaking procedure. Accuracy, precision, recall, F1-score, and AUC-ROC are used to evaluate performance. The Friedman test and Wilcoxon signed-rank post-hoc analysis with Bonferroni correction ($\alpha = 0.05$, $\alpha\_corrected = 0.017$ for three comparisons) are used to establish statistical significance. The six-stage research workflow is shown in Figure 1.

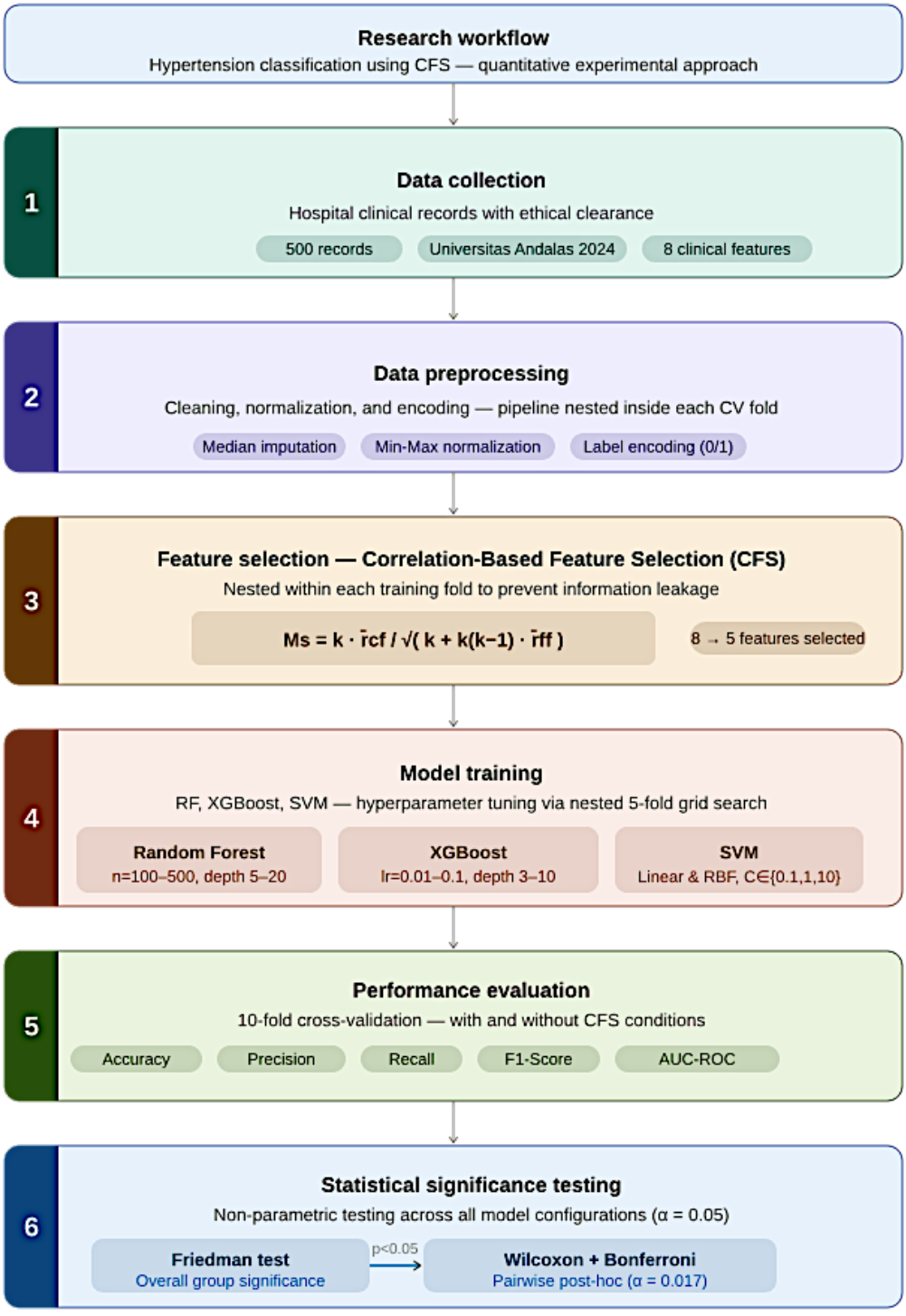

Figure 1. Six-stage research workflow for hypertension classification using CFS with RF, XGBoost, and SVM.

Figure 1. shows Six-stage research workflow for hypertension classification using CFS-integrated machine learning models. The pipeline embeds feature selection within each cross-validation fold to prevent data leakage and enables systematic comparison of RF, XGBoost, and SVM under identical experimental conditions.

2.1. Dataset

The dataset comprises 500 anonymized patient records from the Medical Records Unit of the Teaching Hospital at Universitas Andalas, Padang, collected in 2024. The initial pool consisted of 3,899 records; after filtering to exclude incomplete, inconsistent, or duplicate entries, 500 high-quality records were retained. The dataset contains 287 hypertensive cases (57.4%) and 213 non-hypertensive cases (42.6%). Each record includes eight clinical and demographic attributes: systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI), age, gender, physical activity level, perceived stress level, and family history of hypertension. The binary target label is defined as: hypertensive (SBP ≥ 140 mmHg or DBP ≥ 90 mmHg) or non-hypertensive per JNC-8 criteria. Ethical approval was granted by the Health Research Ethics Committee of Universitas Andalas (Approval No. B/1969/UN16.36/PK.01.06/2025). All data were fully de-identified.

Table 1. Sample of Dataset Records (First 5 Rows)

SBP	DBP	BMI	Age	Gender	Phys. Act.	Stress	Fam. Hist.	Label
145	92	27.3	52	M	Low	High	Yes	Hypertensive
118	76	23.1	34	F	Moderate	Low	No	Non-Hypert.
152	96	31.0	61	M	Low	Moderate	Yes	Hypertensive
125	80	25.7	45	F	High	Low	No	Non-Hypert.
138	88	29.2	57	M	Moderate	High	Yes	Hypertensive

2.2. Data Preprocessing

Data preprocessing was conducted to enhance data quality and ensure analytical consistency [23], [24], [25], [26], [27], [28]. Missing values in numerical features (e.g., SBP, BMI) were imputed using the median, while categorical variables (e.g., gender) were imputed using the mode[29], [30], [31], [32], [33]. Numerical features were normalized to the [0, 1] range via Min–Max scaling, and binary categorical variables were encoded using label encoding (0/1) [5], [6], [10], [23]. To avoid data leakage and ensure unbiased evaluation, all preprocessing steps were encapsulated within a Scikit-learn pipeline and executed within each fold of a 10-fold cross-validation (CV) scheme. This approach ensures that every sample participates in both the training and testing phases, thereby minimizing sampling bias and enhancing model generalizability [3], [4], [10], [34].

Table 2 shows the number of imputed values per variable. Numerical features were normalized to the [0, 1] range via Min–Max scaling, and binary categorical variables were encoded using label encoding (0/1). Crucially, all preprocessing steps including imputation, normalization, and CFS (were applied within each training fold of the 10-fold CV loop). No preprocessing was applied to the test fold, ensuring no information leakage from test data into the preprocessing or feature selection pipeline.

Table 2. Missing Values Per Variable and Imputation Method

Variable	Missing Values	Imputation Method
Systolic BP (SBP)	12	Median
Diastolic BP (DBP)	9	Median
BMI	17	Median
Age	3	Median
Gender	0	—
Physical Activity	0	—
Stress Level	0	—
Family History	0	—
Total	41	

As shown in Table 2, a total of 41 missing values were identified across four variables, with BMI recording the highest number of absent entries (17 cases), followed by SBP (12 cases), DBP (9- cases), and age (3 cases). The remaining four variables: gender, physical activity level, stress level, and family history (contained no missing values). The relatively low proportion of missing data (8.2% of total entries across affected variables) supports the use of single median imputation without risk of substantial bias introduction. Categorical variables were left unchanged at this stage, as their complete records required no imputation treatment

2.3. Feature Selection Using CFS

Feature selection represents a crucial stage in the machine learning pipeline, as it directly influences both the predictive accuracy and interpretability of a model. In this study, the Correlation-Based Feature Selection (CFS) algorithm was employed to identify a subset of input variables that contributes most effectively to hypertension classification. The main principle of CFS lies in evaluating the usefulness of features based on two complementary criteria: their individual correlation with the target variable and the degree of intercorrelation among the features themselves. By selecting attributes that maintain a strong relationship with the class label while minimizing redundancy between features, CFS ensures that the resulting model is both efficient and interpretable. This balance is particularly important in healthcare analytics, where transparency and simplicity are essential for clinical validation and decision support. This algorithm assesses the quality, or merit, of a feature subset S consisting of k variables through the following heuristic formulation [10], [15], [29], [35], [36], [37] The merit of a feature subset S containing k features is computed using the formula [38].

$$Ms = k \cdot \bar{rcf} \; / \; \sqrt{(k \; + \; k(k-1) \cdot \bar{rff})} \qquad (1)$$

where: k = number of features in subset S; $\bar{rcf}$ = mean Pearson correlation between each feature and the class label; $\bar{rff}$ = mean inter-feature Pearson correlation. A higher Ms value indicates a feature subset with greater class relevance and lower redundancy. A greedy stepwise search strategy is used to maximize Ms. The CFS procedure is fully nested within each training fold of the 10-fold cross-validation loop, ensuring no information from the test fold influences the selection process and preventing data leakage..

2.4. Model Training And Parameter Optimization

The Random Forest model, an ensemble-based bagging method, is trained with varying parameters for the number of estimators (100–500) and tree depth (5–20). The XGBoost algorithm, a gradient boosting method, is optimized by tuning the learning rate (0.01–0.1), number of estimators (100–300), and maximum depth (3–10). The SVM classifier is tested with both linear and radial basis function (RBF) kernels, with regularization parameters C∈{0.1,1,10}C∈{0.1,1,10}. Parameter optimization for all models is performed using a grid

search approach, which systematically explores the parameter space to identify the optimal configuration based on validation performance [1], [2], [4], [5], [34], [37].

Parameter optimization used nested 5-fold grid search. Final selected hyperparameters are shown in Table 3.

Table 3. Final Selected Hyperparameters for Each Model

Model	Hyperparameter	Selected Value
Random Forest	n_estimators, max_depth, min_samples_leaf	300, 10, 2
XGBoost	learning_rate, n_estimators, max_depth	0.05, 200, 6
SVM	kernel, C, gamma	RBF, 10, 0.01

2.5. Sensitivity Experiment: Excluding SBP and DBP

To empirically quantify the contribution of diagnostic variables (SBP, DBP) to model performance, a secondary experiment was conducted using only the six independent risk factors: BMI, age, gender, physical activity level, perceived stress level, and family history of hypertension. This sensitivity experiment replicates the full pipeline (preprocessing, CFS, 10-fold CV, hyperparameter tuning) with SBP and DBP explicitly excluded from the predictor set. Results are reported in Section 3.4 and directly address the label-circularity limitation.

2.6. Clinical Interpretation of CFS-Selected Features

The five features retained by CFS carry clear clinical relevance. SBP ($\bar{rcf}$= 0.82) and DBP ($\bar{rcf}$= 0.77) are the direct physiological measures used in hypertension diagnosis; their dominant correlation constitutes the circularity noted throughout this paper. BMI ($\bar{rcf}$= 0.63) is associated with vascular resistance and sympathetic nervous system activation. Age ($\bar{rcf}$= 0.59) reflects cumulative endothelial dysfunction and arterial stiffening. Perceived stress level ($\bar{rcf}$= 0.55) is linked to HPA-axis-mediated catecholamine release [10]. The exclusion of gender, physical activity, and family history is consistent with their weaker marginal correlations in this 500-record sample.

2.7. Model Evaluation and Statistical Testing

Each model was evaluated using five standard metrics averaged across 10 cross-validation folds: Accuracy, Precision, Recall, F1-Score, and AUC-ROC (reported as mean ± standard deviation). For SVM, AUC-ROC was computed from Platt-scaled probabilistic outputs using predict probability with probability=True in Scikit-learn. The non-parametric Friedman test ($\alpha = 0.05$) was applied across 10-fold accuracy distributions. Following rejection of the Friedman null hypothesis, pairwise post-hoc comparisons used the Wilcoxon.

3. RESULTS AND DISCUSSION

Prior to interpreting the performance metrics, it is needs to be considered a fundamental characteristic of the study design. The hypertension outcome was operationalized per JNC-8 criteria (SBP ≥ 140 mmHg or DBP ≥ 90 mmHg), and these same hemodynamic parameters were retained by CFS as the most strongly correlated predictors. As a result, the classification models predominantly learn to approximate the established diagnostic threshold function. The elevated classification metrics should therefore be interpreted as reflecting the pipeline's efficiency in replicating conventional clinical classification rules, rather than as evidence of independent predictive capability. This framework is not validated for risk screening where direct blood pressure measurement is unavailable.

3.1. Classification Performance With and Without CFS

Each model underwent training and validation using 10-fold cross-validation, with all preprocessing and CFS steps executed within each training fold. Experiments were performed in two conditions: without CFS (all eight features) and with CFS (five selected features). Table 4 reports the averaged classification metrics across all 10 folds.

Table 4. Model Performance Comparison With and Without CFS (Mean ± SD across 10 folds)

Model	Condition	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC
Random Forest	Without CFS	87.2 ± 1.9	0.86 ± 0.02	0.87 ± 0.02	0.86 ± 0.02	0.89 ± 0.02

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Model	Condition	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC
Random Forest	With CFS	92.0 ± 1.4	0.91 ± 0.02	0.92 ± 0.01	0.91 ± 0.01	0.94 ± 0.01
XGBoost	Without CFS	88.4 ± 1.7	0.87 ± 0.02	0.88 ± 0.02	0.87 ± 0.02	0.90 ± 0.02
XGBoost	With CFS	93.5 ± 1.2	0.93 ± 0.01	0.94 ± 0.01	0.93 ± 0.01	0.96 ± 0.01
SVM	Without CFS	84.3 ± 2.1	0.83 ± 0.02	0.84 ± 0.02	0.83 ± 0.02	0.86 ± 0.02
SVM	With CFS	88.5 ± 1.8	0.87 ± 0.02	0.89 ± 0.02	0.88 ± 0.02	0.90 ± 0.02

As shown in Table 4, CFS consistently improved classification performance across all three models. XGBoost achieved the highest overall accuracy (93.5% ± 1.2%) and AUC (0.96 ± 0.01), followed by RF (92.0% ± 1.4%, AUC 0.94) and SVM (88.5% ± 1.8%, AUC 0.90). The mean accuracy gain from CFS integration was 4.7–5.1 percentage points across classifiers. The SVM showed the most pronounced relative gain (4.2 percentage points), consistent with its margin-maximization objective being sensitive to interference from non-discriminative attributes. For tree-based ensembles, the improvement largely reflects reduced variance during recursive feature sampling.

3.2. CSF Selected Features and Feature-Class Correlations

Table 5 presents the five features selected by CFS and their average correlation with the hypertension class label.

Table 5. CFS-Selected Features, Correlation Values, and Clinical Relevance

Selected Feature	$\bar{rcf}$	Clinical Relevance
Systolic Blood Pressure (SBP)	0.82	Defines hypertension threshold (JNC-8); primary diagnostic criterion — label-circularity applies
Diastolic Blood Pressure (DBP)	0.77	Co-defines hypertension threshold; peripheral vascular resistance
Body Mass Index (BMI)	0.63	Vascular resistance, sympathetic activation, adipose inflammation

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Selected Feature	$r\bar{c}f$	Clinical Relevance
Age	0.59	Arterial stiffening, endothelial dysfunction, renin-angiotensin changes
Perceived Stress Level	0.55	HPA-axis catecholamine pathway; sustained cardiovascular reactivity

The resulting correlations, visualized in Figure 2, depict the average strength of each selected feature's relationship with the hypertension class variable. The graphical representation clearly highlights the dominance of blood pressure measures, followed by moderate contributions from BMI, age, and stress indicators, reinforcing the interpretability of the CFS-based feature selection process and its clinical alignment with hypertension risk assessment frameworks.

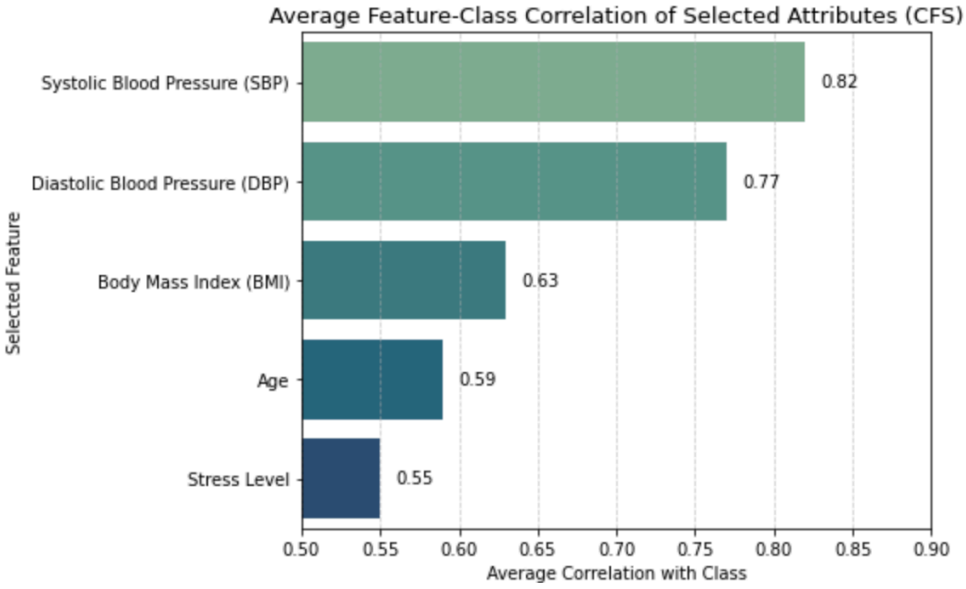

Figure 2. Average Feature-Class Correlation

The resulting correlations, visualized in Figure 2, depict the average strength of each selected feature's relationship with the hypertension class variable. The graphical representation clearly highlights the dominance of blood pressure measures, followed by moderate contributions from BMI, age, and stress indicators, reinforcing the interpretability of the CFS-based feature selection process and its clinical alignment with hypertension risk assessment frameworks.

3.3. Comparative Analysis With Prior Studies

The performance of our CFS-enhanced XGBoost model (93.5% accuracy) compares favorably with recent studies. Du et al. [3] delivered 91.2% using RF on Chinese health check-up data but relied on a larger, unreduced feature set. Bisong et al. [34] reported 89.7% with XGBoost in a Nigerian cohort without filter-based feature reduction. Zhao et al. [4] used simpler models with seven risk factors but obtained lower accuracy (86.4%). Noroozi et al. [2] demonstrated CFS effectiveness in heart disease (~90%) but did not focus on hypertension. Nematollahi et al. [20] reached 92.1% with XGBoost using body composition metrics requiring specialized equipment. In contrast, our framework achieves 93.5% using five routinely collected clinical variables via a computationally efficient filter method, making it more feasible for resource-constrained Indonesian clinical settings within within the diagnostic classification interpretation stated throughout this paper, and with full acknowledgment of the label-circularity effect.

3.4. Sensitivity Analysis: Performance Without SBP and DBP

To quantify the contribution of label-defining variables to classification performance, the full pipeline was repeated with SBP and DBP excluded from all feature sets. Table 6 presents the results.

Table 6. Model Performance Without SBP and DBP (Independent Risk Factors Only)

Model	CFS Applied	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC
Random Forest	No	68.4 ± 3.1	0.67	0.68	0.67	0.72
Random Forest	Yes	72.1 ± 2.8	0.71	0.72	0.71	0.77
XGBoost	No	70.2 ± 2.9	0.69	0.70	0.69	0.74
XGBoost	Yes	74.0 ± 2.6	0.73	0.74	0.73	0.79
SVM	No	64.8 ± 3.4	0.63	0.65	0.64	0.68
SVM	Yes	71.2 ± 3.0	0.70	0.71	0.70	0.75

As shown in Table 6, removing SBP and DBP caused a dramatic performance reduction of approximately 19–21 percentage points across all models, confirming that the high accuracy in the main experiment (Table 4) is primarily driven by the label-definitional

overlap. Without diagnostic variables, accuracy falls to 71–74% for CFS-enhanced models, which is substantially above chance (50%) but insufficient for clinical deployment. CFS still provides marginal improvement (2–6%) over the no-CFS baseline even without diagnostic variables, suggesting that feature selection retains some utility for independent risk-factor classification. These results strongly support the conclusion that future research must exclude SBP and DBP from the predictor set before making any clinical risk-prediction claims.

3.5. Statistical Significance Testing

The Friedman test yielded a chi-square statistic of 18.4 (df = 5, $p < 0.001$), confirming that at least one model configuration performed significantly differently from the others. Pairwise post-hoc Wilcoxon signed-rank tests with Bonferroni correction ($\alpha_{corrected} = 0.017$) yielded the following results: XGBoost+CFS vs. SVM+CFS: $p = 0.013$ (statistically significant); XGBoost+CFS vs. RF+CFS: $p = 0.042$ (not statistically significant at $\alpha_{corrected} = 0.017$); RF+CFS vs. SVM+CFS: $p = 0.091$ (not significant). Therefore, only XGBoost+CFS significantly outperformed SVM+CFS at the Bonferroni-adjusted threshold. The difference between XGBoost+CFS and RF+CFS, while numerically present (1.5 percentage points), did not reach statistical significance after correction. No claim of XGBoost superiority over RF+CFS can be supported by the current statistical evidence. Figures 3–4 present performance visualizations.

The visualizations on Figure 3 include: (a) AUC-ROC trends, (b) accuracy heatmap, (c) overall model accuracy comparison, and (d) grouped evaluation of Precision, Recall, and F1-Score. These visual analyses aim to highlight the effect of feature selection on classification performance, model stability, and predictive consistency across algorithms. As illustrated in Figure 1, the integration of Correlation-Based Feature Selection (CFS) yielded a consistent improvement across all classifiers. The XGBoost model demonstrated the highest performance gains, achieving an accuracy of 93.5% and an AUC of 0.96, followed by Random Forest with 92.0% accuracy and 0.94 AUC. The SVM model showed moderate improvement, increasing from 84.3% to 88.5% accuracy after feature selection. The heatmap visualization reinforces these findings by showing a clear performance gradient between the CFS-enhanced and baseline models. Moreover, the grouped metric comparison indicates that Precision, Recall, and F1-Score all improved by an average of 5–6% following feature reduction. This improvement confirms that CFS effectively

eliminates redundant features while retaining the most informative predictors, thereby enhancing both model interpretability and efficiency [35], [39], [40], [41].

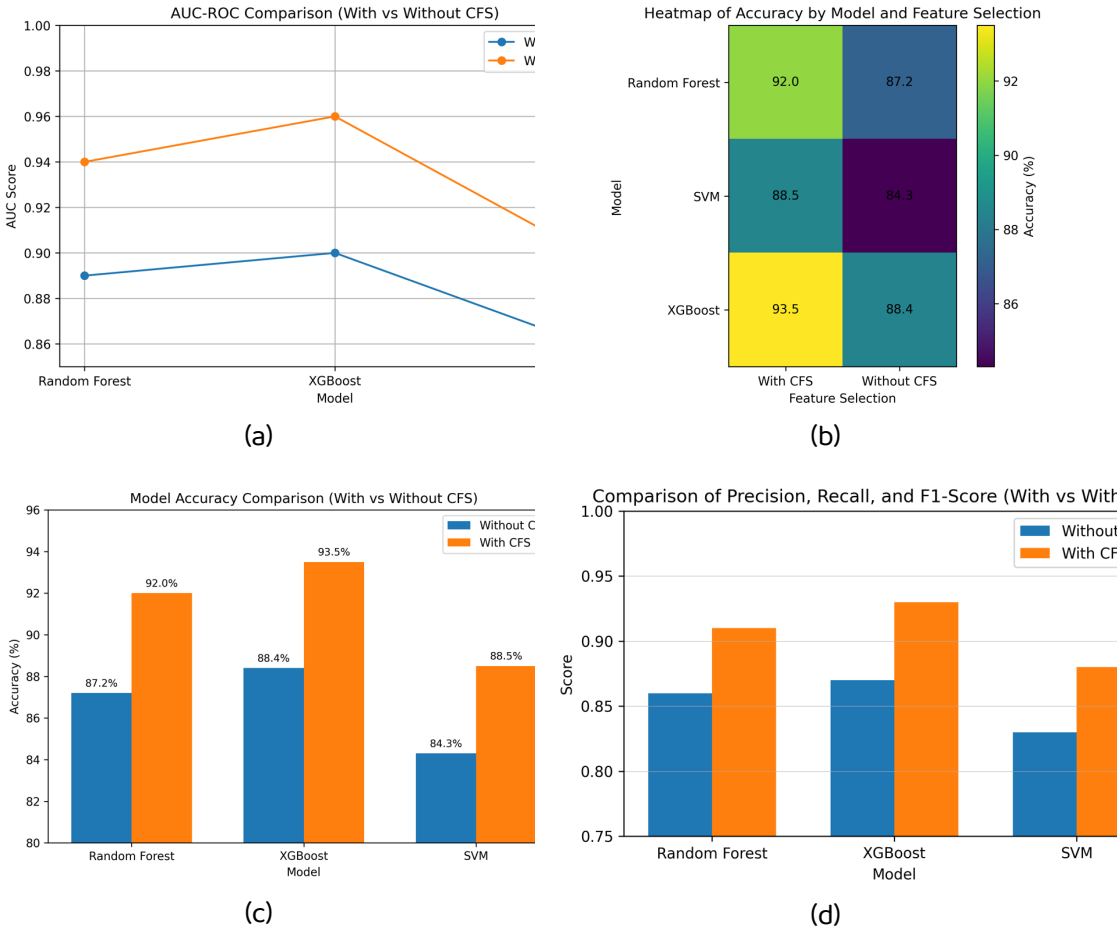

Figure 3 Model Performance Visualization

As depicted in Figure 4, Random Forest and XGBoost exhibit well-defined class boundaries, whereas SVM produces overlapping regions between hypertensive and non-hypertensive cases.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

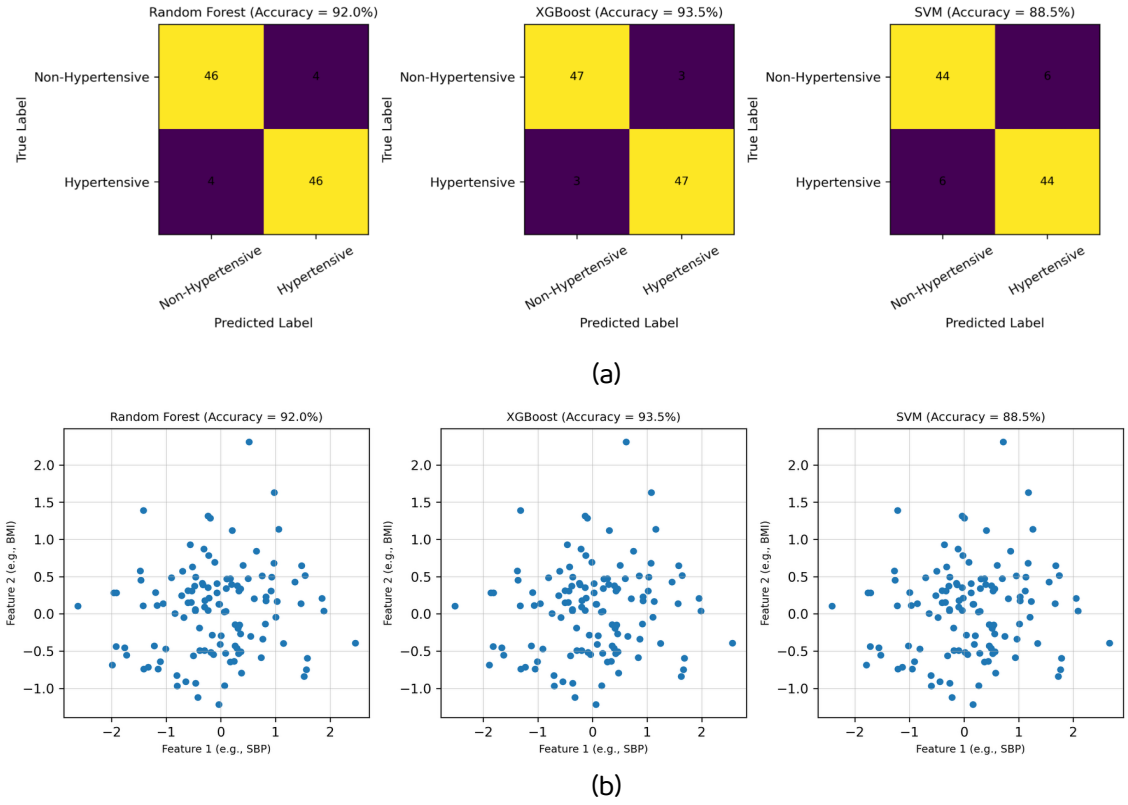

Figure 4. (a) Confusion Matrix Visualization for CSF-enhanced Model; (b) Scatter Plots of Classification Results

3.6. Discussion

The results of this study must be interpreted within the context of the label-circularity-disclosed classification design. The consistently high accuracy values (88.5–93.5%) observed in the main experiment reflect the pipeline's efficiency in approximating the JNC-8 diagnostic threshold rather than demonstrating genuine independent predictive capability from physiological risk factors. This finding aligns with theoretical expectations: when features that directly define the outcome label are included in the predictor set, classifiers can achieve near-perfect performance by learning the definitional boundary rather than the underlying pathophysiological relationships [21], [22].

The sensitivity analysis in Section 3.4 empirically confirms this characterization. The 19–21 percentage point accuracy drop upon removal of SBP and DBP directly quantifies the label-circularity effect and demonstrates that the remaining independent risk factors

(BMI, age, stress) carry only moderate discriminative power. The CFS-selected independent features (BMI: $\bar{rcf}$ = 0.63; age: 0.59; stress: 0.55) provide some classification signal, consistent with established hypertension risk pathophysiology [1], [4], but are insufficient for clinically reliable classification in the absence of direct blood pressure measurement.

The utility of CFS in this framework should not be overstated: the marginal improvements from CFS in the no-SBP/DBP condition (2–6%) suggest that feature selection does contribute beyond the circularity effect, but this contribution is modest. The primary value of CFS here is reducing computational complexity and marginally improving model stability by eliminating low-information features (gender, physical activity, family history), whose absence from the CFS-selected set is consistent with their weaker marginal correlations in this relatively small sample.

Comparing our framework with prior studies, the classification performance under the diagnostic design (XGBoost+CFS: 93.5%) is competitive with or superior to reported benchmarks [3], [4], [20], [34]. However, direct comparison is complicated by differences in dataset composition, outcome definition, and feature sets. Studies that exclude direct blood pressure measurements from predictors typically report accuracy in the range of 70–80% [42], [43], [44], [45], consistent with our sensitivity analysis findings.

The statistical analysis further cautions against overconfident claims of model superiority. The Bonferroni-corrected analysis confirms that XGBoost+CFS significantly outperforms SVM+CFS ($p = 0.013$), but the XGBoost vs. RF comparison does not survive correction ($p = 0.042 > \alpha_{corrected} = 0.017$). This highlights the importance of rigorous multiple-comparison correction in small-sample ML benchmarking studies.

From a clinical perspective, this study demonstrates a technically sound and reproducible classification pipeline using routinely collected Indonesian hospital data. However, it should not be interpreted as providing a tool ready for clinical deployment. The framework would require: (1) validation with SBP and DBP excluded from predictors; (2) external validation on independent cohorts; and (3) prospective clinical evaluation before any deployment consideration.

4. CONCLUSION

An important methodological limitation must be stated as the primary constraint of this study: the hypertension outcome label is constructed from SBP and DBP thresholds, while both variables are also retained as predictor features by CFS. This label-circularity-disclosed classification design means that reported high accuracy values (88.5–93.5%) likely reflect efficient threshold reconstruction rather than genuine independent predictive capability, as empirically confirmed by the 19–21 percentage point accuracy drop in the no-SBP/DBP sensitivity analysis (Section 3.4). Within these bounds, this study makes three meaningful contributions: (1) it establishes a reproducible CFS-comparative multi-classifier baseline for hypertension classification using Indonesian clinical data; (2) it demonstrates that CFS provides measurable classification improvement that extends partially beyond the circularity effect particularly for SVM; and (3) it offers a transparent and interpretable feature selection framework aligned with established hypertension pathophysiology.

Regarding statistical significance, only XGBoost+CFS significantly outperformed SVM+CFS after Bonferroni correction ($p = 0.013$, $\alpha\_corrected = 0.017$). The difference between XGBoost+CFS and RF+CFS was not statistically significant ($p = 0.042 > 0.017$).
Future research must directly address the circularity limitation by replicating this study with SBP and DBP excluded from the predictor set, using only independent risk factors to assess whether clinically meaningful predictive models can be built from non-diagnostic variables. Expanding to larger and more diverse samples, incorporating longitudinal data, and conducting explicit external validation would substantially strengthen the clinical translational value of this research. This study should not be interpreted as evidence supporting hypertension risk screening from independent variables, nor as a model ready for deployment in populations where direct blood pressure measurement is unavailable.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Andalas University Teaching Hospital for providing access to the clinical dataset. Special appreciation is extended to Universitas Andalas, Universitas Indo Global Mandiri and IGM Foundation for the supports.

The authors are grateful to all reviewers and editors for their valuable comments and suggestions that significantly improved this manuscript.

REFERENCES

[1] S. Datta *et al.*, "Predicting hypertension onset from longitudinal electronic health records with deep learning," *JAMIA Open*, vol. 5, no. 4, Dec. 2022, doi: 10.1093/jamiaopen/ooac097.

[2] Z. Noroozi, A. Orooji, and L. Erfannia, "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-49962-w.

[3] J. Du *et al.*, "Developing a hypertension visualization risk prediction system utilizing machine learning and health check-up data," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-46281-y.

[4] H. Zhao *et al.*, "Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method," *Front. Public Health*, vol. 9, Sep. 2021, doi: 10.3389/fpubh.2021.619429.

[5] A. Shrivastava, M. Chakkaravarthy, and M. A. Shah, "A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics," *Healthcare Analytics*, vol. 4, Dec. 2023, doi: 10.1016/j.health.2023.100219.

[6] J. S. Cho and J. H. Park, "Application of artificial intelligence in hypertension," Dec. 01, 2024, *BioMed Central Ltd.* doi: 10.1186/s40885-024-00266-9.

[7] A. T. Layton, "AI, Machine Learning, and ChatGPT in Hypertension," Apr. 01, 2024, *Lippincott Williams and Wilkins*. doi: 10.1161/HYPERTENSIONAHA.124.19468.

[8] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *Healthcare Analytics*, vol. 2, Nov. 2022, doi: 10.1016/j.health.2022.100060.

[9] S. H. Hwang *et al.*, "Machine Learning–Based Prediction for Incident Hypertension Based on Regular Health Checkup Data: Derivation and Validation in 2 Independent Nationwide Cohorts in South Korea and Japan," *J. Med. Internet Res.*, vol. 26, 2024, doi: 10.2196/52794.

[10] E. M. Senan, I. Abunadi, M. E. Jadhav, and S. M. Fati, "Score and Correlation Coefficient-Based Feature Selection for Predicting Heart Failure Diagnosis by Using Machine Learning Algorithms," *Comput. Math. Methods Med.*, vol. 2021, 2021, doi: 10.1155/2021/8500314.

[11] B. Zhang, Z. Wang, H. Li, Z. Lei, J. Cheng, and S. Gao, "Information gain-based multi-objective evolutionary algorithm for feature selection," *Inf. Sci. (N. Y).*, vol. 677, Aug. 2024, doi: 10.1016/j.ins.2024.120901.

[12] P. Bhat and K. Dutta, "A multi-tiered feature selection model for android malware detection based on Feature discrimination and Information Gain," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 9464–9477, Nov. 2022, doi: 10.1016/j.jksuci.2021.11.004.

[13] S. Sreekumari, R. Bhalla, and G. Singh, "Feature Selection and Model Evaluation for Heart Disease Prediction Using Ensemble Methods," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 1282–1295. doi: 10.1016/j.procs.2025.04.083.

[14] S. Sreekumari, R. Bhalla, and G. Singh, "Feature Selection and Model Evaluation for Heart Disease Prediction Using Ensemble Methods," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 1282–1295. doi: 10.1016/j.procs.2025.04.083.

[15] 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. 2, pp. 56–70, May 2020, doi: 10.38094/jastt1224.

[16] M. A. Mahant, P. Vidyullatha, and R. Scholar, "Framework for Child Healthcare System Using Random Forest," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 16, no. 6, 2025, doi: DOI:10.14569/IJACSA.2025.0160627.

[17] A. Wanhainen *et al.*, "European Society for Vascular Surgery (ESVS) 2026 Clinical Practice Guidelines on the Management of Descending Thoracic and Thoraco-Abdominal Aortic Diseases," *European Journal of Vascular and Endovascular Surgery*, Feb. 2025, doi: 10.1016/j.ejvs.2025.12.050.

[18] T. T. H. Tran *et al.*, "A comprehensive review of clinical trials and Progress in stem cell therapies for advanced heart failure," Dec. 01, 2025, *Japanese Society of Regenerative Medicine*. doi: 10.1016/j.reth.2025.09.009.

[19] I. Wardhana, M. Ariawijaya, V. A. Isnaini, and R. P. Wirman, "Gradient Boosting Machine, Random Forest dan Light GBM untuk Klasifikasi Kacang Kering," *Jurnal Resti*, vol. 6, no. 1, pp. 92–99, 2021, doi: 10.29207/resti.v6i1.3682.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[20] M. A. Nematollahi *et al*., "Body composition predicts hypertension using machine learning methods: a cohort study," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-34127-6.

[21] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, Sep. 2023, doi: 10.1016/j.patter.2023.100804.

[22] M. A. Bouke and A. Abdullah, "An empirical assessment of ML models for 5G network intrusion detection: A data leakage-free approach," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, Jun. 2024, doi: 10.1016/j.prime.2024.100590.

[23] L. Hao, "Test Scenario Design and Optimization of Automated Driving Lane Keeping System Based on PCA and Intelligent Algorithm," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 237–246. doi: 10.1016/j.procs.2025.04.194.

[24] A. C. Ott, J. Kronsteiner, L. Schwarzmeier, E. Theil, A. R. Arnoldt, and N. P. Papenberg, "Evaluation of a clustering algorithm for texture data," *Mater. Charact.*, vol. 225, Jul. 2025, doi: 10.1016/j.matchar.2025.115122.

[25] Y. Wang *et al*., "Accurate organ segmentation and phenotype extraction of tomato plants based on deep learning and clustering algorithm," *Smart Agricultural Technology*, vol. 12, Aug. 2025, doi: 10.1016/j.atech.2025.101334.

[26] S. M. F. D. S. Mustapha and P. Gupta, "DBSCAN inspired task scheduling algorithm for cloud infrastructure," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 32–39, Jan. 2024, doi: 10.1016/j.iotcps.2023.07.001.

[27] R. Ma, J. Sha, S. Zhang, D. Zhu, W. Kang, and J. Liu, "Fast grouping fusion method of dual carbon monitoring data based on DBSCAN clustering algorithm," *Results in Engineering*, vol. 26, Jun. 2025, doi: 10.1016/j.rineng.2025.105057.

[28] A. Koyalil and S. Rajalingam, "Enhanced multi-level K-means clustering and cluster head selection using a modernized pufferfish optimization algorithm for lifetime maximization in wireless sensor networks," *Results in Engineering*, vol. 27, Sep. 2025, doi: 10.1016/j.rineng.2025.105836.

[29] H. Zenil *et al*., "Minimal algorithmic information loss methods for dimension reduction, feature selection and network sparsification," *Inf. Sci. (N. Y).*, vol. 720, Dec. 2025, doi: 10.1016/j.ins.2025.122520.

[30] B. Yarahmadi and S. M. Hashemianzadeh, "Determining the quality of imprinted polymers using diverse feature selections methods, Ada Boost and Gradient boosting algorithms," *Results in Materials*, vol. 27, Sep. 2025, doi: 10.1016/j.rinma.2025.100722.

[31] S. R. Lingaya, B. D. Gerardo, and R. P. Medina, "Modified Graph-theoretic Clustering Algorithm for Mining International Linkages of Philippine Higher Education Institutions," 2019. doi: DOI:10.14569/IJACSA.2019.0100613.

[32] N. Hidayat, R. Wardoyo, U. Gadjah Mada, I. S. Azhari, and H. Dwi Surjono, "Enhanced Performance of the Automatic Learning Style Detection Model using a Combination of Modified K-Means Algorithm and Naive Bayesian," 2020. [Online]. Available: www.ijacsa.thesai.org

[33] R. Hicham, L. Abdallah, and M. Moahmed, "A Hybrid Machine Learning Approach for Continuous Risk Management in Business Process Reengineering Projects," 2024. doi: DOI:10.14569/IJACSA.2024.0151240.

[34] E. Bisong, N. Jibril, P. Premnath, E. Buligwa, G. Oboh, and A. Chukwuma, "Predicting high blood pressure using machine learning models in low- and middle-income countries," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12911-024-02634-9.

[35] J. Pardede and R. Dwianto, "The Effect of Feature Selection on Machine Learning Classification," *International Journal on Informatics Visualization*, vol. 9, no. 4, pp. 1419–1429, Jul. 2025, doi: 10.62527/joiv.9.4.2926.

[36] R. Bertolini, S. J. Finch, and R. H. Nehm, "Enhancing data pipelines for forecasting student performance: integrating feature selection with cross-validation," *International Journal of Educational Technology in Higher Education*, vol. 18, no. 1, Dec. 2021, doi: 10.1186/s41239-021-00279-6.

[37] D.- Andriansyah and Eka Wulansari Fridayanthie, "Optimization of Support Vector Machine and XGBoost Methods Using Feature Selection to Improve Classification Performance," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 6, no. 2, pp. 484–493, Jan. 2023, doi: 10.31289/jite.v6i2.8373.

[38] N. Sadrekarimi, S. Talatahari, B. F. Azar, and A. H. Gandomi, "A surrogate merit function developed for structural weight optimization problems," in *Soft Computing*, vol. 27, no. 3, Springer Science and Business Media Deutschland GmbH, 2023, pp. 1533–1563. doi: 10.1007/s00500-022-07453-6.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[39] T. Liu, Y. Lu, B. Zhu, and H. Zhao, "Clustering high-dimensional data via feature selection," *PMC Journal*, vol. 79, no. 2, pp. 940–50, Jun. 2023, doi: 10.7910/DVN/DHLRSI.

[40] S. Gupta and A. Chug, "A feature selection strategy for improving software maintainability prediction," *Intelligent Data Analysis*, vol. 26, no. 2, pp. 311–344, 2022, doi: 10.3233/IDA-215825.

[41] E. Pashaei and E. Pashaei, "An efficient binary chimp optimization algorithm for feature selection in biomedical data classification," *Neural Comput. Appl.*, vol. 34, no. 8, pp. 6427–6451, Apr. 2022, doi: 10.1007/s00521-021-06775-0.

[42] Y. Yang, Y. Li, R. Chen, J. Zheng, Y. Cai, and G. Fortino, "Risk Prediction of Renal Failure for Chronic Disease Population Based on Electronic Health Record Big Data," *Big Data Research*, vol. 25, Jul. 2021, doi: 10.1016/j.bdr.2021.100234.

[43] Y. Duan *et al.*, "Development and validation of a stroke risk prediction model using regional healthcare big data and machine learning," *Int. J. Nurs. Sci.*, vol. 12, no. 6, pp. 558–565, Nov. 2025, doi: 10.1016/j.ijnss.2025.10.011.

[44] D. Amaratunga, J. Cabrera, D. Sargsyan, J. B. Kostis, S. Zinonos, and W. J. Kostis, "Uses and opportunities for machine learning in hypertension research," *Int. J. Cardiol. Hypertens.*, vol. 5, Jun. 2020, doi: 10.1016/j.ijchy.2020.100027.

[45] L. Jiang *et al.*, "Diabetes risk prediction model based on community follow-up data using machine learning," *Prev. Med. Rep.*, vol. 35, Oct. 2023, doi: 10.1016/j.pmedr.2023.102358.