Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 2, April 2026, pp. 1442~1451
ISSN: 2302-9285, DOI: 10.11591/eei.v15i2.11247

A robust model for early detection of chronic kidney disease leveraging machine learning and data balancing techniques

Helmi Imaduddin[1], Siti Agrippina Alodia Yusuf[2], Muhammad Syahriandi Adhantoro[3]
[1]Department of Informatics Engineering, Faculty of Communication and Informatics, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia
[2]Department of Information Systems and Technology, Faculty of Engineering, Universitas Muhammadiyah Mataram, Mataram, Indonesia
[3]Department of Information System, Faculty of Communication and Informatics, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

Article Info

Article history:

Received Aug 20, 2025
Revised Jan 14, 2026
Accepted Mar 10, 2026

Keywords:

Chronic kidney disease
Class imbalance
Cost-sensitive learning
Machine learning
Synthetic minority over-sampling technique

ABSTRACT

Chronic kidney disease (CKD) requires reliable early screening, yet clinical datasets are often highly class imbalanced, which can bias machine learning models and reduce detection quality. This study presents a unified evaluation of two imbalance mitigation strategies, synthetic minority over-sampling technique (SMOTE), and cost-sensitive learning, across six classifiers: decision tree (DT), K-nearest neighbors (KNN), logistic regression (LR), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XGBoost). Experiments were conducted on a public CKD dataset with 1,659 records and 54 features using a consistent pipeline including preprocessing, feature selection, imbalance handling, and stratified k-fold cross-validation. Models were assessed with accuracy, precision, recall, and F1-score. Results show that the imbalance strategy materially changes model behavior: cost-sensitive learning generally improves precision, while SMOTE more often increases recall and F1-score. The best overall performance was achieved by XGBoost with cost-sensitive learning, reaching 93% accuracy and 92% precision, outperforming prior reports on the same dataset. RF remained stable across both strategies, whereas KNN was sensitive to SMOTE induced distribution shifts. These findings provide practical guidance for selecting imbalance handling methods to improve healthcare machine learning for CKD detection.

Corresponding Author:

Helmi Imaduddin
Department of Informatics Engineering, Faculty of Communication and Informatics
Universitas Muhammadiyah Surakarta
Surakarta, Indonesia
Email: hi776@ums.ac.id

1. INTRODUCTION

Chronic kidney disease (CKD) represents an increasingly pressing global health concern, exhibiting considerable prevalence across numerous nations. According to data from the World Health Organization (WHO), roughly 10% of the planet's population is impacted by CKD, translating to over 850 million individuals [1]. CKD is marked by a gradual and irreversible deterioration in renal function, potentially resulting in kidney failure necessitating life-saving interventions such as dialysis or organ transplantation. Identifying CKD at an early stage is crucial to effective medical administration, as timely intervention can decelerate disease progression and enhance the quality of life for patients [2].

Machine learning technology has advanced rapidly in recent years. Machine learning serves as a potent tool for analyzing intricate health data, enabling the refinement of diagnostic models for specific diseases that augment detection abilities. The primary challenge in training machine learning algorithms lies in acquiring balanced datasets, especially when collected data is unbalanced [3]. In the context of CKD, the dataset often contains a markedly higher number of cases from one class compared to the other, which can lead to biased models that underperform in predictive accuracy and fail to generalize effectively.

Despite the growing use of machine learning for CKD prediction, a persistent methodological challenge is severe class imbalance in clinical datasets. When class distributions are highly skewed, standard learners optimized for overall accuracy tend to favor the majority class, leading to biased decision boundaries and suboptimal clinical utility. This issue is particularly critical in screening scenarios where misclassification can translate into delayed intervention or unnecessary follow-up testing.

To address these challenges, recent studies have explored ensemble and hybrid learning for CKD prediction, often coupled with imbalance mitigation techniques. One widely used method is the synthetic minority over-sampling technique (SMOTE), which generates synthetic examples of minority class instances to improve class distribution and enhance model performance [4]. Another promising approach is cost-sensitive learning, which adjusts the classification process by assigning higher misclassification costs to minority class errors, thereby directing the algorithm's focus toward these critical cases. Both methods aim to strengthen the predictive capacity of models in imbalanced data scenarios, ultimately improving early CKD detection.

Numerous studies have been conducted on SMOTE, Rahayu *et al.* [5] addressed class disproportion in an employee performance dataset using C4.5 and reported that SMOTE increased accuracy from 17% to 86% across multiple train test splits. Consistently, Wadhawan *et al.* [6] showed that SMOTE improved several classifiers, including support vector machine (SVM), K-nearest neighbors (KNN), decision tree (DT), Naïve Bayes, and random forest (RF), with gains ranging from 2.88% to 15.08%. Evidence also indicates that high quality synthetic samples generated via SMOTE can further enhance ensemble learning, for example improving RF accuracy from 97.75% to 100% [7].

Recent CKD prediction studies increasingly rely on machine learning and ensemble learning to support early screening and risk stratification. Multiple works report strong performance on public CKD datasets by combining robust classifiers, feature selection, and explainability. For example, an explainable machine learning interface for CKD diagnosis evaluated several algorithms and highlighted strong performance for modern ensemble learners while providing model interpretability using explanation techniques such as SHAP or related approaches [8]. Beyond single model development, systematic reviews and meta-analyses emphasize that reported performance is often difficult to compare across studies due to heterogeneous datasets, validation protocols, and inconsistent reporting of imbalance handling and clinically meaningful metrics [9], [10].

A common strategy is to pair ensemble learners, such as RF and gradient boosting, with oversampling, most frequently SMOTE, to mitigate skewed class distributions. Alturki *et al.* [11] introduced TrioNet, an ensemble framework integrating extreme gradient boosting (XGBoost) and RF with SMOTE, and reported very high accuracy on a UCI-style CKD benchmark. Dey *et al.* [12] combined hybrid feature selection with SMOTE and observed strong performance for tree based ensembles across multiple algorithms. Similarly, Ramu *et al.* [13] applied SMOTE and reported improved results using a hybrid CNN SVM approach. More complex pipelines further incorporate optimization, feature selection, and heterogeneous voting or stacking, as in DOVE-FELM and related fusion frameworks, again reporting high accuracy and underscoring the value of ensemble diversity [14], [15].

Although SMOTE remains the predominant choice in CKD pipelines, the medical diagnostics literature increasingly explores alternatives that address limitations of standard synthetic sampling, including mixed data types, feature structure, and noisy decision boundaries. For example, clustering and autoencoding have been used to refine oversampling and improve synthetic sample quality for mixed numerical and categorical data, which is relevant to CKD style tabular settings even when not CKD specific [16]. In parallel, cost-aware learning and model design choices, such as class weighting, stacking, and robust feature engineering, are frequently embedded in end-to-end frameworks, yet many CKD studies do not offer a controlled comparison against oversampling under matched experimental conditions.

This study addresses the lack of like-for-like evidence on how modern machine learning algorithms perform for CKD prediction under consistent experimental settings and class imbalance conditions. To close this gap, we conduct a controlled benchmark using a single public tabular dataset and a unified evaluation pipeline, comparing multiple classifiers under three imbalance-handling strategies: baseline training, SMOTE-based oversampling, and cost-sensitive learning. Model performance is assessed with standard classification metrics, enabling a clear comparison of both algorithm choice and imbalance mitigation. The results provide practical guidance on selecting robust models and balancing techniques for early CKD detection in imbalanced clinical data.

We evaluate the effectiveness of SMOTE and cost-sensitive learning for CKD classification across six widely used machine learning algorithms: DT, KNN, logistic regression (LR), RF, SVM, and XGBoost. These models were selected to represent complementary learning paradigms that are frequently applied in medical data analysis. Specifically, DT and RF offer interpretability and robustness, KNN provides an instance-based baseline, LR serves as a simple and well-established reference model, SVM is effective in high-dimensional settings, and XGBoost represents a state-of-the-art boosting approach. By pairing each classifier with both imbalance-handling strategies under a consistent evaluation pipeline, we aim to determine how the choice of imbalance mitigation interacts with model characteristics and to identify the most effective combinations for early CKD detection. Beyond identifying the best performing model, the study provides practical guidance on the expected precision, recall, and F1-score trade-offs introduced by oversampling versus cost-sensitive learning across different classifier families. These insights can support the development of robust and clinically useful machine learning systems for CKD screening.

Despite the growing body of research on CKD prediction using machine learning, a clearly articulated technical gap remains. Most prior studies focus primarily on maximizing accuracy or proposing complex ensemble architectures, yet they often evaluate a single imbalance-handling strategy, predominantly SMOTE, without conducting controlled comparisons against alternative approaches under identical experimental conditions. In addition, several studies report strong performance without systematically isolating the effect of imbalance mitigation from model architecture, making it difficult to determine whether performance gains stem from the classifier design, data resampling, or validation protocol. Furthermore, heterogeneous preprocessing pipelines, inconsistent validation schemes, and limited reporting of precision-recall trade-offs restrict the reproducibility and comparability of findings. Consequently, there is limited empirical evidence on how different classifier families respond to distinct imbalance-handling strategies within a unified and controlled evaluation framework. This methodological limitation motivates the present study, which provides a like-for-like comparative benchmark of baseline training, SMOTE, and cost-sensitive learning across multiple machine learning algorithms using a consistent preprocessing and validation pipeline.

This study is guided by two primary research questions. First, how effective are SMOTE and cost-sensitive learning in improving the performance of classification models on imbalanced chronic kidney disease (CKD) data. Second, which combination of machine learning algorithm and imbalance-handling technique yields the highest classification performance for early CKD detection?.

Through these research questions, the study aims to provide a comprehensive understanding of the comparative performance of multiple algorithms under different imbalance mitigation strategies, offering practical guidance for healthcare practitioners and researchers in implementing machine learning for early CKD detection. The findings are expected to contribute to the development of advanced, technology-based solutions that enhance predictive accuracy, support timely clinical intervention, and ultimately improve public health outcomes.

2. METHOD

This study aims to develop and benchmark robust machine learning models for early CKD prediction using a public tabular dataset under class imbalance conditions. We compare multiple classifiers using a unified pipeline across three training setups, baseline training, SMOTE-based oversampling, and cost-sensitive learning, and report performance using standard classification metrics to ensure fair and reproducible comparison. Figure 1 shows the flow of the research conducted.

2.1. Data collection

The data source used in this study is a public dataset available on the Kaggle platform, which contains medical information about patients with and without CKD [17]. This dataset includes 54 features. The use of public datasets facilitates replication of the study by other researchers and provides transparency in the methodology used. The total dataset used was 1,659 data consisting of 1,524 positive CKD cases and 135 negative CKD cases. Table 1 shows the number of datasets used.

We also perform preliminary assessments of the quality of the data, focusing on any missing values or outliers that could bias the results of the examination. In this context, choosing a dataset encompassing an adequate number of entries and diverse variables is pivotal for bolstering the model's generalizability. Following the compilation of information, the subsequent stage involves conducting exploratory data analysis to gain insights into patterns and relationships among the factors. This requires carrying out descriptive statistical evaluation to acquire a thorough comprehension of the data dispersion and its elements. Data visualization helps to aid in uncovering designs that may remain obscured through only numerical examination. This strategy permits informed decision-making for subsequent steps, like data preprocessing, applying SMOTE and cost-sensitive learning. Careful selection is needed to ensure the information used maximizes what can be learned while maintaining diversity in its scope.

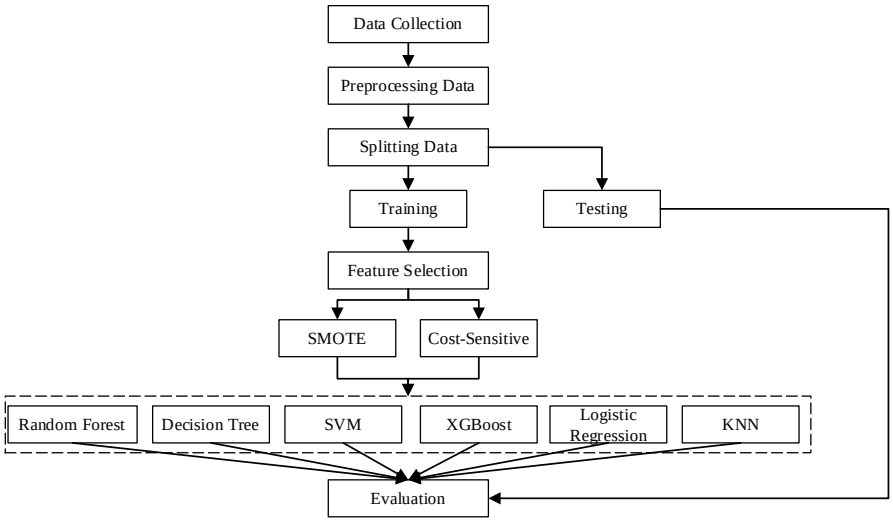

Figure 1. Research flow

Table 1. Class distribution of the CKD dataset

Class	Number of data
Positive CKD	1,524
Negative CKD	135
Total	**1,659**

2.2. Data preprocessing

Data preprocessing constituted a fundamental stage in this research to ensure that the dataset was both clean and suitably structured for model development. The initial task involved examining the dataset for missing values. As no missing entries were identified, no imputation or deletion procedures were necessary. Subsequently, column names were standardized to ensure consistency and ease of reference. This process entailed replacing any spaces with underscores using df.columns.str.replace(' ', '_'), for example converting "Cholesterol Total" to "Cholesterol_Total," followed by converting all column names to lowercase via .str.lower(), resulting in "cholesterol_total." This practice improves readability, avoids syntactic errors, and facilitates seamless processing in subsequent analytical steps.

The dataset was then separated into features (X) and target labels (y). The target variable, diagnosis, was converted into an integer format to support binary classification, where 0 represented "Negative CKD" and 1 represented "Positive CKD." All variables except diagnosis were assigned to X, while y exclusively contained the diagnosis values. To prepare the data for model training, a preprocessing pipeline was developed to address both numerical and categorical variables. StandardScaler was applied to numerical features to normalize their distribution, ensuring each had a mean of zero and a standard deviation of one, thereby preventing variables with larger magnitudes from exerting disproportionate influence on the learning process. OneHotEncoder was applied to categorical variables, transforming each category into a binary representation. These operations were coordinated through a ColumnTransformer, which applied the appropriate transformations to designated feature groups while passing any remaining columns through without alteration.

2.3. Data splitting

The dataset was partitioned into training and testing subsets to ensure objective evaluation of model performance and to minimise the risk of overfitting. The proportion of data allocated for testing was set at 25% of the total dataset, with the remaining 75% reserved for training. This division resulted in 1,244 records in the training set and 415 records in the testing set.

The random_state parameter was fixed at 42, functioning as a "seed" for the randomization process. Setting this parameter ensures that the data split remains consistent across multiple runs, thereby facilitating reproducibility of the results. Additionally, the stratify=y parameter was applied to preserve the class distribution of the target variable across both subsets, a critical consideration in classification tasks involving imbalanced datasets. Maintaining distinct training and testing datasets is essential for generating an unbiased estimate of model performance. Such separation allows for an accurate assessment of the model's capacity to generalize its predictive capabilities to unseen data, thereby ensuring the reliability of the evaluation process [18].

2.4. Feature selection

Feature selection was performed using model based feature importance. After preprocessing, a RF classifier was trained on the training data with n_estimators=100, class_weight=balanced, and random_state=42, and predictors were ranked using feature_importances_. From the 54 available variables, the top 20 features with the greatest influence on classification outcomes were retained and summarized as importance percentages in Table 2. This reduced feature set was applied consistently in subsequent model training and evaluation using an index based selector within the pipeline to ensure a fair comparison across models while reducing redundancy, computational cost, and the risk of overfitting.

Table 2. Top 20 feature ranked based on importance scores

Rank	Feature	Importance (%)
1	serumcreatinine	11.35
2	gfr	7.94
3	itching	5.43
4	proteininurine	4.73
5	musclecramps	3.96
6	bunlevels	3.63
7	fastingbloodsugar	3.5
8	hba1c	3.15
9	systolicbp	2.64
10	cholesterolhdl	2.63
11	physicalactivity	2.5
12	sleepquality	2.36
13	dietquality	2.34
14	serumelectrolytesphosphorus	2.34
15	healthliteracy	2.33
16	fatiguelevels	2.2
17	acr	2.19
18	cholesterolldl	2.18
19	hemoglobinlevels	2.17
20	cholesteroltotal	2.17

2.5. Handling class imbalance

2.5.1. Synthetic minority over-sampling technique method

This study implements the SMOTE to address the previously identified issue of class imbalance within the dataset. SMOTE generates synthetic samples for the minority class by utilizing the characteristics of existing observations [19], [20]. The process involves selecting data points from the underrepresented class and creating new synthetic points between them, with Euclidean distance serving as the basis for interpolation. By increasing the number of minority class instances, SMOTE not only balances class distribution but also enables machine learning models to capture a broader spectrum of data variability.

SMOTE was implemented using imbalanced-learn with SMOTE (random_state=42). All other parameters used default settings, including k_neighbors=5 and sampling_strategy='auto', which oversamples the minority class until a 1:1 class ratio is reached in the training data. SMOTE was applied within the cross-validation pipeline after preprocessing and feature selection, ensuring resampling was performed only on the training portion of each fold. Figure 2 shows class distribution in the training set before and after applying SMOTE. The figure illustrates the degree of imbalance observed after stratified splitting, where negative CKD (diagnosis=0) is the minority class and positive CKD (diagnosis=1) is the majority class. After SMOTE is applied within the training fold, the minority class is oversampled until an approximately 1:1 ratio is achieved.

2.5.2. Cost-sensitive learning

Cost-sensitive learning was implemented via class weighting based on class frequencies in the training set. For LR, SVM, DT, and RF, we used class_weight = 'balanced', which assigns weights inversely proportional to class prevalence. For XGBoost, cost sensitivity was implemented using scale_pos_weight computed from the training distribution as the ratio between class frequencies, and this value was fixed during cross-validation to penalize misclassification according to the observed imbalance.

Cost-sensitive learning is an imbalance-handling approach that explicitly assigns higher penalties to misclassification errors on the minority class, thereby encouraging the model to better capture underrepresented CKD cases. The objective is to minimise the number of required interactions while ensuring optimal convergence rates [21]. In the context of imbalanced classification, cost-sensitive learning has been widely adopted as an effective strategy for reducing the probability gap between majority and

minority classes. It achieves this by adjusting cost factors to place greater emphasis on correctly classifying minority class instances [22].

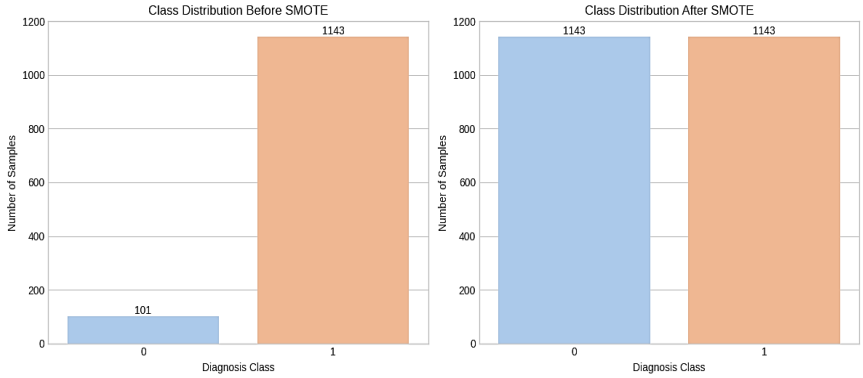

Figure 2. Class distribution in the training data before and after applying SMOTE

2.6. Machine learning algorithms

This study applies six machine learning algorithms to classify CKD: DT, KNN, LR, RF, SVM, and XGBoost. These algorithms were selected to encompass a range of learning paradigms for comprehensive performance evaluation under different imbalance-handling strategies. RF is an ensemble method that constructs multiple DT and aggregates predictions by majority vote [23]. It is robust for classification and regression, mitigates overfitting, and is effective for imbalanced CKD datasets. SVM is effective for high-dimensional spaces and managing class imbalance [24], [25], identifying an optimal hyperplane between classes.

DT is a non-parametric, supervised learning algorithm with a hierarchical structure of root nodes, decision nodes, branches, and leaves [26]. Internal nodes test attributes, branches show outcomes, and leaves provide final predictions. The tree grows by recursively splitting data until a stopping condition is reached. KNN is a non-parametric, instance-based method assigning labels based on the majority class among nearest neighbors [27], [28]. While simple and versatile, KNN can be computationally intensive for large datasets due to distance calculations, a limitation mitigated through clustering or selective instance reduction. Both DT and KNN are valued for interpretability and adaptability across domains.

XGBoost is well suited for clinical tabular data because it can model complex nonlinear relationships and feature interactions, includes regularization to reduce overfitting, and typically delivers strong predictive performance with efficient training. It sequentially builds DT, each iteration correcting prior errors, and has shown high accuracy in tasks such as predicting heat transfer coefficients in liquid cold plates, with predictions within ±10% and ±20% of true values for 63% and 90% of cases respectively [29]. LR models binary outcome probabilities using the logistic function and estimates parameters through maximum likelihood [30], [31]. Its interpretability and ability to handle categorical and continuous predictors make it widely adopted in medicine, finance, and social sciences. These six algorithms provide a broad basis for evaluating CKD classification performance under SMOTE and cost-sensitive learning strategies.

2.7. Evaluation

Performance assessment was conducted using four primary evaluation metrics: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly classified instances over the total number of instances, offering a general indication of predictive correctness. Precision quantifies the proportion of positive predictions that are actually correct, reflecting the model's ability to avoid false positives. Recall, also known as sensitivity, measures the proportion of actual positive cases correctly identified by the model, making it especially important in medical diagnostics where missing positive cases can have severe consequences. The F1-score, defined as the harmonic mean of precision and recall, provides a balanced measure that accounts for both false positives and false negatives, offering a comprehensive performance indicator in the presence of class imbalance.

3. RESULTS AND DISCUSSION

The experimental results demonstrate that the effect of imbalance mitigation is strongly model dependent and cannot be generalized across classifier families. Technically, SMOTE and cost-sensitive learning intervene at different stages of the learning process. SMOTE alters the empirical data distribution by

synthetically generating minority samples in feature space using interpolation between nearest neighbors [19], [20], whereas cost-sensitive learning modifies the objective function by reweighting misclassification penalties [21], [22]. These fundamentally distinct mechanisms explain the heterogeneous performance observed in Table 3.

Table 3. Classification results of the six machine learning models under SMOTE and cost-sensitive learning

No	Model	Strategy	Accuracy	Precision	Recall	F1-score
1	DT	Cost-sensitive	0.89	0.63	0.64	0.63
2	DT	SMOTE	0.83	0.58	0.64	0.59
3	KNN	Cost-sensitive	0.92	0.78	0.53	0.54
4	KNN	SMOTE	0.68	0.56	0.67	0.52
5	LR	Cost-sensitive	0.77	0.6	0.76	0.6
6	LR	SMOTE	0.78	0.6	0.73	0.6
7	RF	Cost-sensitive	0.92	0.86	0.52	0.53
8	RF	SMOTE	0.92	0.76	0.63	0.67
9	SVM	Cost-sensitive	0.89	0.64	0.66	0.65
10	SVM	SMOTE	0.9	0.64	0.61	0.62
11	XGBoost	Cost-sensitive	**0.93**	**0.92**	0.59	0.63
12	XGBoost	SMOTE	0.92	0.73	0.71	0.72

For tree-based ensemble methods such as RF and XGBoost, the relatively stable performance under both strategies can be attributed to their structural robustness. RF aggregates multiple decorrelated DT [23], reducing variance and mitigating overfitting introduced by synthetic instances. In the case of XGBoost, the boosting framework sequentially corrects residual errors while incorporating regularization [29], which helps control model complexity even when class weighting is applied. The superior precision achieved by XGBoost under cost-sensitive learning suggests that penalizing minority class misclassification directly influences gradient updates in a more controlled manner than synthetic resampling, preserving clearer decision boundaries while still addressing imbalance.

In contrast, instance-based methods such as KNN exhibit higher sensitivity to distributional shifts. KNN relies explicitly on local neighborhood structure for classification [27], [28]. When SMOTE generates synthetic samples through linear interpolation in feature space [19], [20], the resulting artificial neighbors may distort local density estimation, especially in high-dimensional or sparsely populated regions. This explains the substantial decline in KNN accuracy under SMOTE. Cost-sensitive learning, by comparison, does not modify the geometry of the feature space, allowing KNN to maintain more stable neighborhood relationships.

SVM demonstrates comparatively balanced behavior under both strategies. Because SVM seeks an optimal separating hyperplane with maximum margin [24], [25], oversampling can improve minority class representation near the margin, potentially enhancing recall. However, cost-sensitive weighting adjusts the penalty parameter for misclassified samples, effectively shifting the decision boundary toward the majority class. The moderate performance differences observed suggest that SVM benefits from improved minority representation but remains sensitive to over-representation if synthetic samples introduce noise.

LR, as a linear probabilistic classifier [30], [31], assumes a linear decision boundary in transformed feature space. The relatively modest performance gains under both imbalance strategies indicate that class imbalance is not the sole limiting factor. Instead, the CKD dataset likely contains nonlinear interactions among features that cannot be fully captured by a linear log-odds formulation. Therefore, imbalance mitigation alone does not substantially enhance its discriminative capacity.

From a broader methodological perspective, the results demonstrate that imbalance handling should not be treated as a universally beneficial preprocessing step. Instead, its effectiveness depends on the interplay between sampling strategy, model complexity, and decision boundary formulation. Oversampling methods such as SMOTE reshape the training distribution and may improve recall, yet they risk introducing synthetic overlap between classes. Cost-sensitive learning preserves the original data structure but modifies the optimization landscape, often leading to improved precision. These trade-offs are consistent with prior imbalance learning theory [21], [22], which emphasizes the role of misclassification cost adjustment in narrowing probability gaps between classes.

Compared with earlier CKD classification work by Azizah and Paramitha [32], which used a Naïve Bayes algorithm and reported an accuracy of 89.9%, the best performing configuration in this study, XGBoost with cost sensitive learning, achieved 93% accuracy and 92% precision. Recent studies on larger and more clinically representative cohorts further contextualize these results. Chen *et al.* [33] developed and externally validated an ensemble model (RF, XGBoost, and LightGBM) using multi center EHR data from

five sites (1,200 patients), reporting an AUC of approximately 0.89 with external accuracy around 0.79 to 0.80. Iftikhar *et al.* [34] reported RF accuracy of approximately 0.90 to 0.91 using a clinically focused dataset. Haider *et al.* [35] achieved an accuracy of 0.9237 using a hard voting ensemble. While ensemble performance across studies appears competitive, the present study contributes additional technical insight by isolating the effect of imbalance mitigation under a unified experimental pipeline, clarifying how performance variation is linked not only to model architecture but also to the interaction between learning algorithm and imbalance strategy.

4. CONCLUSION

This study systematically compared SMOTE and cost-sensitive learning across six machine learning classifiers for early CKD detection using a public CKD dataset of 1,659 records and 54 features. The results demonstrate that imbalance mitigation significantly influences model behavior, where cost-sensitive learning generally improves precision through loss reweighting, while SMOTE more frequently enhances recall and F1-score by modifying class distribution. XGBoost combined with cost-sensitive learning achieved the best overall performance with 93% accuracy and 92% precision, and RF showed stable performance under both strategies. The main contribution of this work is the controlled, unified evaluation of oversampling and penalty-based approaches across diverse classifier families within a consistent preprocessing and validation pipeline. However, the study is limited by the use of a single public dataset without external validation, moderate dataset size under severe imbalance, and the absence of exhaustive hyperparameter optimization and calibration analysis. Future research should therefore include external multi-center validation, systematic hyperparameter tuning integrated with cross-validation, exploration of hybrid or adaptive imbalance strategies, and incorporation of probability calibration and explainability methods to strengthen clinical robustness and deployment readiness.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Kemdiktisaintek for providing financial support for the publication of this paper. The authors also would like to express their sincere gratitude to Universitas Muhammadiyah Surakarta for their support in providing the facilities and resources necessary for conducting this research.

FUNDING INFORMATION

This research was supported by an internal research grant under the PID scheme from Universitas Muhammadiyah Surakarta. In addition, the authors received publication assistance funding from the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (Kemdiktisaintek) to support the publication of this article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Helmi Imaduddin	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Siti Agrippina Alodia Yusuf		✓				✓		✓	✓	✓	✓	✓		
Muhammad Syahriandi Adhantoro	✓		✓	✓			✓			✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this research.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/rabieelkharoua/chronic-kidney-disease-dataset-analysis/data.

REFERENCES

[1] A. K. Bello, I. G. Okpechi, A. Levin, and D. W. Johnson, "Variations in kidney care management and access: regional assessments of the 2023 International Society of Nephrology Global Kidney Health Atlas (ISN-GKHA)," *Kidney International Supplements*, vol. 13, no. 1, pp. 1–5, Apr. 2024, doi: 10.1016/j.kisu.2023.12.001.

[2] A. Francis *et al*., "Chronic kidney disease and the global public health agenda: an international consensus," *Nature Reviews Nephrology*, vol. 20, no. 7, pp. 473–485, Jul. 2024, doi: 10.1038/s41581-024-00820-6.

[3] T. Sajana and K. V. S. N. R. Rao, "Machine Learning Algorithms for Health Care Data Analytics Handling Imbalanced Datasets," in *Handbook of Artificial Intelligence*, Singapore: Bentham Science Publishers, 2023, pp. 75–96, doi: 10.2174/9789815124514123010006.

[4] Y. Zhang, L. Deng, and B. Wei, "Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation," *Mathematics*, vol. 12, no. 11, pp. 1–17, May 2024, doi: 10.3390/math12111709.

[5] W. Rahayu *et al*., "Synthetic Minority Oversampling Technique (SMOTE) for Boosting the Accuracy of C4.5 Algorithm Model," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 3, no. 3, pp. 624–630, Jun. 2024, doi: 10.59934/jaiea.v3i3.469.

[6] S. Wadhawan, R. Maini, and B. Singh, "Analyzing the Impact of Oversampling on Classifier Performance for Cardiac Disease Classification," in *Lecture Notes in Networks and Systems*, vol. 915, 2024, pp. 723–739, doi: 10.1007/978-981-97-0700-3_54.

[7] H. Sug, "An Oversampling Technique with Descriptive Statistics," *WSEAS Transactions on Information Science and Applications*, vol. 21, pp. 318–332, Jul. 2024, doi: 10.37394/23209.2024.21.31.

[8] G. Dharmarathne, M. Bogahawaththa, M. McAfee, U. Rathnayake, and D. P. P. Meddage, "On the diagnosis of chronic kidney disease using a machine learning-based interface with explainable artificial intelligence," *Intelligent Systems with Applications*, vol. 22, pp. 1–13, Jun. 2024, doi: 10.1016/j.iswa.2024.200397.

[9] M. Haris *et al*., "Prediction of incident chronic kidney disease in community-based electronic health records: a systematic review and meta-analysis," *Clinical Kidney Journal*, vol. 17, no. 5, May 2024, doi: 10.1093/ckj/sfae098.

[10] F. Sanmarchi, C. Fanconi, D. Golinelli, D. Gori, T. Hernandez-Boussard, and A. Capodici, "Predict, diagnose, and treat chronic kidney disease with machine learning: a systematic literature review," *Journal of Nephrology*, vol. 36, no. 4, pp. 1101–1117, Feb. 2023, doi: 10.1007/s40620-023-01573-4.

[11] N. Alturki *et al*., "Improving Prediction of Chronic Kidney Disease Using KNN Imputed SMOTE Features and TrioNet Model," *CMES - Computer Modeling in Engineering and Sciences*, vol. 139, no. 3, pp. 3513–3534, 2024, doi: 10.32604/cmes.2023.045868.

[12] S. K. Dey, K. M. M. Uddin, H. M. H. Babu, M. M. Rahman, A. Howlader, and K. M. A. Uddin, "Chi2-MI: A hybrid feature selection based machine learning approach in diagnosis of chronic kidney disease," *Intelligent Systems with Applications*, vol. 16, pp. 1–13, Nov. 2022, doi: 10.1016/j.iswa.2022.200144.

[13] K. Ramu *et al*., "Hybrid CNN-SVM model for enhanced early detection of Chronic kidney disease," *Biomedical Signal Processing and Control*, vol. 100, p. 107084, Feb. 2025, doi: 10.1016/j.bspc.2024.107084.

[14] B. Sowan, L. Zhang, E. H. Houssein, H. Qattous, M. Azzeh, and B. Massad, "DOVE-FELM: A fusion-optimized feature selection and heterogeneous ensemble learning framework for early prediction of chronic kidney disease risk," *Array*, vol. 28, pp. 1–37, Dec. 2025, doi: 10.1016/j.array.2025.100613.

[15] M. F. Hossain, S. T. Diya, and R. Khan, "ACD-ML: Advanced CKD detection using machine learning: A tri-phase ensemble and multi-layered stacking and blending approach," *Computer Methods and Programs in Biomedicine Update*, vol. 7, pp. 1–20, 2025, doi: 10.1016/j.cmpbup.2024.100173.

[16] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, "Enhancing SMOTE for imbalanced data with abnormal minority instances," *Machine Learning with Applications*, vol. 18, pp. 1–31, Dec. 2024, doi: 10.1016/j.mlwa.2024.100597.

[17] R. El Kharoua, "Chronic Kidney Disease Dataset," *Kaggle*, 2024, doi: 10.34740/KAGGLE/DSV/8658224.

[18] I. Rubachev, N. Kartashev, Y. Gorishniy, and A. Babenko, "TabReD: A Benchmark of Tabular Machine Learning in-the-Wild," *arXiv preprint*, 2024, doi: 10.48550/arXiv.2406.19380.

[19] P. Sun, Z. Wang, L. Jia, and Z. Xu, "SMOTE-kTLNN: A hybrid re-sampling method based on SMOTE and a two-layer nearest neighbor classifier," *Expert Systems with Applications*, vol. 238, p. 121848, Mar. 2024, doi: 10.1016/j.eswa.2023.121848.

[20] H. Kaur and A. Kaur, "SMOTE-Based Homogeneous Prediction for Aging-Related Bugs in Cloud-Oriented Software," *Journal of Information and Knowledge Management*, vol. 22, no. 5, Oct. 2023, doi: 10.1142/S0219649223500478.

[21] B. N. Njike and X. Siebert, "Nonparametric active learning for cost-sensitive classification," *arXiv preprint*, doi: 10.48550/arXiv.2310.00511.

[22] W. Zheng and H. Zhao, "Cost-sensitive hierarchical classification for imbalance classes," *Applied Intelligence*, vol. 50, no. 8, pp. 2328–2338, Aug. 2020, doi: 10.1007/s10489-019-01624-z.

[23] Z. S. Priyambudi and Y. S. Nugroho, "Which algorithm is better? An implementation of normalization to predict student performance," in *AIP Conference Proceedings*, vol. 2926, no. 1, p. 030009, 2024, doi: 10.1063/5.0182879.

[24] A. J. Latipah, G. Ariyanto, R. Hasudungan, and N. N. Fatihah, "Classification Of Butterfly Species Based On Venasi Using Support Vector Machine," *International Journal of Engineering & Technology*, vol. 8, no. 11, pp. 173–176, 2019, doi: 10.14419/ijet.v8i1.1.25921.

[25] S. S. Mawarni, M. Murinto, and S. Sunardi, "Medical External Wound Image C[lassification Using Support Vector Machine Technique," *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 9, no. 2, pp. 98–103, Oct. 2023, doi: 10.23917/khif.v9i2.22541.

[26] T. Thomas, A. P. Vijayaraghavan, and S. Emmanuel, "Applications of Decision Trees," in *Machine Learning Approaches in Cyber Security Analytics*, Singapore: Springer Singapore, 2020, pp. 157–184, doi: 10.1007/978-981-15-1706-8_9.
[27] V. B and R. Gangula, "Exploring the Power and Practical Applications of K-Nearest Neighbours (KNN) in Machine Learning," *Journal of Computer Allied Intelligence*, vol. 2, no. 1, pp. 8–15, Feb. 2024, doi: 10.69996/jcai.2024002.
[28] M. T. R. A. Wijaya, I. Widaningrum, A. Prasetyo, and D. Mustikasari, "Using SVM and KNN for Predicting Customer Response Sentiment of M-PAJAK Application," *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 11, no. 1, pp. 33–39, Jul. 2025, doi: 10.23917/khif.v11i1.4528.
[29] M. R. Shaeri, M. C. Ellis, and A. M. Randriambololona, "Xgboost-Based Model for Prediction of Heat Transfer Coefficients in Liquid Cold Plates," in *8th Thermal and Fluids Engineering Conference (TFEC)*, Connecticut: Begellhouse, 2023, pp. 539–542, doi: 10.1615/tfec2023.cmd.045483.
[30] F. Acito, "Logistic Regression," in *Predictive Analytics with KNIME: Analytics for Citizen Data Scientists*, Cham: Springer Nature Switzerland, 2023, pp. 125–167, doi: 10.1007/978-3-031-45630-5_7.
[31] J. A. M. Pérez and P. S. P. Martín, "Regresión logística," *Semergen - Medicina de Familia*, vol. 50, no. 1, p. 102086, Jan. 2024, doi: 10.1016/j.semerg.2023.102086.
[32] M. F. Azizah and A. T. Paramitha, "Predictive Modelling of Chronic Kidney Disease Using Gaussian Naive Bayes Algorithm," *International Journal of Artificial Intelligence in Medical Issues*, vol. 2, no. 2, pp. 125–135, Nov. 2024, doi: 10.56705/ijaimi.v2i2.160.
[33] H. Chen, Y. Huang, and L. Chen, "Ensemble machine learning for predicting renal function decline in chronic kidney disease: development and external validation," *Frontiers in Medicine*, vol. 12, p. 1598065, Oct. 2025, doi: 10.3389/fmed.2025.1598065.
[34] H. Iftikhar, A. F. Hashem, M. Qureshi, and P. C. Rodrigues, "Clinical Application of Machine Learning Models for Early-Stage Chronic Kidney Disease Detection," *Diagnostics*, vol. 15, no. 20, pp. 1–18, Oct. 2025, doi: 10.3390/diagnostics15202610.
[35] H. A. Haider, M. Hussain, and I. L. Kharisma, "Predictive Analysis of Chronic Kidney Disease in Machine Learning," *Engineering Proceedings*, vol. 107, no. 1, pp. 1–7, Sep. 2025, doi: 10.3390/engproc2025107118.

BIOGRAPHIES OF AUTHORS

Helmi Imaduddin received the Bachelor degree in Informatics Engineering from Universitas Muhammadiyah Surakarta in 2015, and the Master degree in Information Technology from Gadjah Mada University, Yogyakarta in 2020. Currently, he is a Lecturer at the Faculty of Communication and Informatics, Universitas Muhammadiyah Surakarta, Indonesia. His research interests include natural language processing, image processing, and classification. He can be contacted at email: hi776@ums.ac.id.

Siti Agrippina Alodia Yusuf received Bachelor of Informatic in Universitas Bumigora, Mataram, Indonesia, Master of Engineering from Universitas Gadjah Mada, Yogyakarta, Indonesia. She is a lecturer in information and system technology department in Universitas Muhammadiyah Mataram. Her research interests include pattern recognition, speech processing, biomedical signal processing, pathological speech processing, and machine learning. She can be contacted at email: siti.agrippina@ummat.ac.id.

Muhammad Syahriandi Adhantoro has completed his academic education by earning both Bachelor's and Master's degrees in Informatics from Universitas Muhammadiyah Surakarta. With a strong educational background in Informatics, he is currently an active lecturer in the Information Systems Study Program at Universitas Muhammadiyah Surakarta. His research interests are currently focused on several fields that are highly relevant to the current development of information technology, namely artificial intelligence (AI), machine learning, and educational games. He has produced many high-quality research works in these fields, which not only contribute to the advancement of science and technology but also have a positive impact on society. He is committed to continuously developing research and teaching in the field of information technology. He can be contacted at email: m.syahriandi@ums.ac.id.