Influence of Class Imbalance and Resampling on Classification Accuracy of Chronic Kidney Disease Detection

Ayodeji Olalekan Salau[1,2,3*] , Elisha Didam Markus[2] , Tsehay Admassu Assegie[4] , Crescent Onyebuchi Omeje[5] , Joy Nnenna Eneh[6]

[1] Department of Electrical/Electronics and Computer Engineering, Afe Babalola University, Ado-Ekiti 360001, Nigeria
[2] Department of Electrical, Electronic and Computer Engineering, Central University of Technology, Bloemfontein 9320, Free State, South Africa
[3] Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Tamil Nadu 600124, India
[4] Department of Computer Science, Injibara University, Injibara 6040, Ethiopia
[5] Department of Electrical/Electronic Engineering, University of Port Harcourt, Rivers State, Choba 500004, Nigeria
[6] Department of Electronic Engineering, University of Nigeria, Nsukka 410001, Nigeria

Corresponding Author Email: ayodejisalau@abuad.edu.ng

ABSTRACT

Chronic kidney disease is one of the leading causes of death around the world. Early detection of chronic kidney disease is crucial to the reduction of mortality caused as a result of the disease. Machine learning methods are recently becoming popular for the detection of chronic kidney disease. This study investigates the influence of resampling for chronic kidney disease detection using an imbalanced chronic kidney disease dataset. Choosing an optimal feature subset for medical datasets is important for improving the performance of data-driven predictive models. The influence of imbalanced class distribution on predictive models has become an increasingly important topic due to the recent advances in automatic decision-making processes and the continuous expansion in the volume of the data collected by medical institutions. To address the identified research gap, an experimental evaluation of synthetic minority oversampling and near miss undersampling technique was performed on a real-world chronic kidney disease dataset using several classification methods such as decision tree, random forest, K-nearest neighbor, adaptive boosting, and support vector machine. The results demonstrate that a number of variables, including performance metrics, classification algorithm, and dataset characteristics, influence the best class distribution. The study also offers useful information about resampling methods for an imbalanced classification problem which will help improve classification accuracy.

1. INTRODUCTION

The objective of chronic kidney disease prediction is to identify whether a person is suffering from kidney disease by classifying a given observation into two classes: patient class (positive class) and non-patient class (negative class). To detect chronic kidney disease, the predictive model uses chronic kidney disease features to characterize each observation in the dataset. Automated predictive models are important in decision making especially in giving an early indication of whether a person should be tested for chronic kidney disease or not. Automated test results are used in deciding whether further tests and additional screening is required or not. If the automated test using the predictive model is positive, the patient will be subjected to additional tests, but no further test will be recommended afterwards. The emphasis, however, is not only on which class a classifier predicts for a given observation in a dataset, but also on the certainty that the correct class is predicted. In practice, different types of errors result in different outcomes in the real world. In the medical field, making a false negative prediction while testing for chronic kidney disease leads to a serious

challenge. Consequently, making a wrong prediction on the likelihood of whether a patient is suffering from chronic kidney disease or not will lead to other health issues [1], particularly in low-income countries where millions die due to lack of proper medication and treatment.

Imbalanced class distribution is a common problem in chronic kidney disease classification, where chronic kidney disease patients or the negative class could greatly outnumber the chronic kidney disease patients or the positive class. Imbalanced class distribution has become one of the major challenges in machine learning; therefore improving the predictive accuracy of kidney disease classification models is an ongoing research problem [2]. The performance of standard machine learning algorithms tends to decrease when trained on a dataset that contains imbalanced class distributions. As a result, imbalanced class distributions particularly the frequent occurrence of negative class affects the classification performance of the positive class, which is critical for disease detection. Handling imbalanced datasets and improving the classification accuracy of imbalanced datasets needs additional research effort [1, 2]. Numerous methods have been proposed to improve the predictive accuracy of imbalanced

dataset classification such as changing the class weight or class distribution [3].

The availability of a large number of medical datasets related to chronic kidney disease and the development of predictive models for chronic kidney disease detection with sampled data representing the overall population has become common practice in developing a predictive model for disease diagnosis. Although many research works have focused on chronic kidney disease prediction, comparing the predictive accuracy of classification models, data sample design, and the effect of imbalanced class distribution on the performance of predictive models has been neglected. Few studies have been conducted to address the issue of imbalanced class distribution on the performance of classification models for the detection of chronic kidney disease. Andrić et al. [4] investigated the effect of imbalanced class distribution on the performance of classification models. However, the authors did not conduct any experiments on any medical dataset, as the study focused on credit card assessment. Moreover, the problem of imbalanced class distribution was not empirically evaluated for chronic kidney disease detection. Furthermore, despite the importance and applicability of automated disease diagnosis, better resampling techniques were not used.

To address the perceived research gap, this paper examines the effect of imbalanced class distribution on the predictive accuracy of selected classification models. In addition, the effect of oversampling and sampling techniques on a real-world chronic kidney disease dataset was examined using gradient boosting, decision tree, random forest, and k-nearest neighbor (kNN). The performance of the proposed model is evaluated with uncertainty measures such as decision function and prediction probability of the positive and negative classes. The main contribution of this study is summarized as follows: (1) The study showed that the examined resampling techniques optimizes the prediction accuracy, (2) evaluation of gradient boosting, decision tree, random forest, and kNN was performed for the imbalanced class, and (3) investigation was performed using the minimum class distribution size for the imbalanced class distribution using the dataset.

These remaining sections of the paper are organized as follows: section 2 presents a review of related works, while the classification algorithms are discussed in section 3. The experimental framework is described in section 4, and the section also presents the experimental results, while section 5 presents the conclusion and recommendation for future work.

2. RELATED WORK

Numerous studies have been carried out on the problem of chronic kidney disease detection with automated machine learning models such as tree-based algorithms like decision tree and random forest algorithms [5] and ensemble methods such as gradient descent algorithm [6]. In recent years, there is huge research attention towards the use of machine learning for medical applications due to the wide spreading cases of chronic kidney disease. Among the recent research conducted on the problem of chronic kidney disease prediction with machine learning algorithms, the most widely used algorithms of chronic kidney disease include the support vector machine (SVM) [7], and adaptive boosting algorithm [8].

Imbalanced class distribution is a common problem for medical datasets and also for disease classification with machine learning algorithms [9]. Assegie et al. [9] analyzed

the effect of imbalanced class distribution on the performance of classifiers for disease prediction. The authors claim that the performance of ensemble classifiers is better compared to individual classifiers on an imbalanced classification problem. One of the approaches proposed to avoid the effect of class imbalance on classifier performance is random oversampling. Random oversampling technique is commonly applied to imbalanced class to produce more samples of the minority class (usually the positive class in medical datasets) to balance the class distribution.

Balancing the class distribution of a dataset tends to improve the accuracy of the classification algorithm. Bai et al. [10] proposed a synthetic minority oversampling (SMOTE) technique for balancing the class distribution in an intrusion detection dataset. The authors analyzed the performance of classification algorithms on balanced dataset with SMOTE and the results show that the classification accuracy improved with the balanced dataset as compared to a dataset with imbalanced class distribution. The class distribution was balanced by introducing synthetic samples to the dataset.

Khuat and Le [11] applied the SMOTE on a diabetes mellitus dataset for balancing the class distribution of the diabetes mellitus positive and negative class. According to Seo and Kim [12], the synthetic oversampling technique is one of the most powerful techniques widely employed in medicine for imbalanced class distribution.

The most common approach for medical diagnosis using machine learning classifiers on imbalanced class distribution is the resampling technique. Despite the advantages of oversampling techniques such as SMOTE and random sampling for improving the performance of a predictive model, studies performed for choosing a resampling technique for model evaluation are limited. Moreover, the problem of imbalanced class distribution on the performance of predictive models is rarely researched; hence, this study which proposes resampling techniques, compares the existing random under-sampling with Near Miss and synthetic minority over-sampling and in addition, compares the predictive accuracy of different classification models. This study therefore aims to address the aforementioned gaps with a view to providing insights into chronic kidney disease detection modelling with classifiers using empirical evidence on how to choose the training sample and resampling technique and which algorithm is appropriate for chronic kidney disease detection.

3. METHODOLOGY

This section presents the method employed for data collection and analysis. In addition, the statistical method employed for analyzing the dataset was presented. Furthermore, the performance metrics employed in the evaluation of the performance of the employed machine learning algorithms are discussed.

3.1 Data collection

To conduct the experiments, real-world chronic kidney disease data was collected from Kaggle online data repository. The characteristics of the dataset employed in the study are summarized in Table 1. The dataset has 399 observations of which 149 observations are for chronic kidney disease patients and 250 observations are not for chronic kidney disease patients (non-patient observations). Each observation has 18

features which are used for describing the class label.

Table 1. Summary of data sample

Number of observations	Number of classes	Number of positive samples	Number of negative samples	Number of features
399	2	149	250	18

The acquired dataset contains information on chronic kidney disease. Features extracted include demographic information of patient such as age and laboratory test results. The laboratory test result features include blood pressure, sugar level, albumin, blood glucose random test, potassium, sodium, blood urea, red blood cell count, and hemoglobin.

3.2 Descriptive statistics of numeric features

The descriptive statistics include the mean, median, maximum, minimum, and standard deviation for each input feature of the chronic kidney disease dataset. The descriptive statistics for each numeric input feature in the dataset are summarized in Table 2.

Table 2. Descriptive statistics for numerical input features

Input feature	Max.	Min.	STD	Mean
Age	90	2	16.97	51.48
Blood pressure	180	50	13.47	76.46
Sugar	5	0	1.02	0.45
Albumin	5	0	1..27	1.01
Blood glucose random	490	22	74.78	148.0
Blood urea	391	1.50	49.28	57.42
Hemoglobin	17.8	3.1	2.71	12.52
Sodium	163	4.50	9.20	137.5
Potassium	47	2.50	2.81	12.52

We observe from Table 2, that the maximum age value is 90 and the minimum age in the observation is 2. This is because, the dataset includes all of the age groups ranging from children to the elderly.

3.3 Feature selection

To reduce the dimensionality and remove irrelevant features that negatively affects the performance of the classification model, a sequential feature selection algorithm was employed. The sequential feature selection algorithm is widely used for selecting the optimal number of feature subsets that are relevant to train the model. The sequential feature selection algorithm removes or adds one feature at a time based on classifier performance. The number of features added is determined a priori; thus, the algorithm terminates when the pre-defined number of feature subsets to be selected is reached.

3.3.1 Optimal feature selection with sequential feature selection

A sequential feature selection algorithm is employed for finding the optimal set of input features from the 18 features characterizing each of the observations in the dataset. The sequential feature selection algorithm returned the feature subset described in Table 3 as the optimal feature set. Hence, the classification algorithms are trained on the optimal feature set summarized in Table 3. A classification accuracy of 99.3%

is achieved with the sequential feature selection algorithm when the selected features are used for training as depicted in Figure 1. The best combination is given as: 0, 1, 2, 3, 4, 6, 8, 9, 10, 11, 13, 14, 15 achieving an accuracy of 99.3%.

Table 3. Optimal input features selected by the sequential feature selection algorithm

No.	Index	Selected feature
1	0	Age
2	1	Blood pressure
3	2	Specific gravity
4	3	Albumin
5	4	Sugar
6	6	Blood urea
7	8	Sodium
8	9	Potassium
9	10	Hemoglobin
10	13	Bacteria
11	14	Pus Cell clumps
12	15	Hypertension
13	13	Bacteria

As presented in Table 3, the features such as age, blood pressure, and so on, are selected as the optimal feature set. The optimal features are used as input features for training the classification algorithms to achieve a better predictive performance. The features presented in Table 3, consists of only 13 features out of the total 18 features which characterize the dataset and are used when the classification algorithm is trained on the training set. As observed from Table 4, the classification accuracy of the model tends to improve with additional features. However, the highest accuracy was achieved with 13 features out of the 18 features. Figure 1 demonstrates the cross-validation score of the predictive model for chronic kidney disease detection.
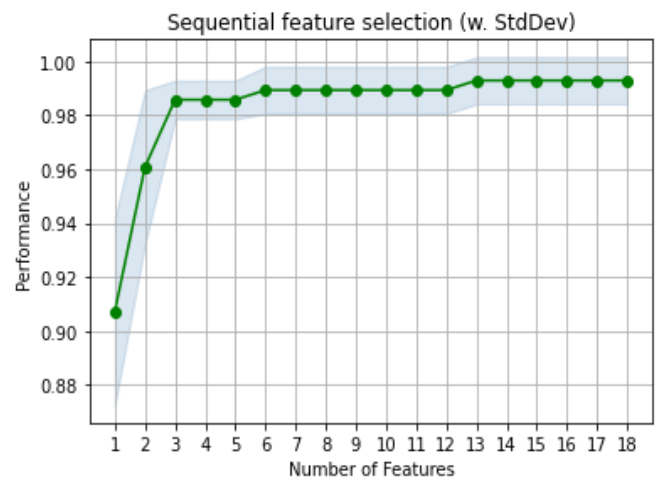

Figure 1. Cross-validation score of different feature combination selected using sequential feature selection

It was observed that the maximum cross-validation accuracy is achieved with 13 input features for characterizing each observation of the dataset as depicted in Figure 1. Moreover, the cross-validation accuracy remained constant for feature subsets 14 to 18. Hence, the maximum number of input features that maximizes the predictive accuracy of the classification model is 13 and the highest cross-validation score is 99.3%. The best combination of the 13 input feature subset are as follows: Best combination: 0, 1, 2, 3, 4, 6, 8, 9,

10, 11, 13, 14, 15. This combination gives an accuracy of 99.3%.

3.4 Synthetic minority oversampling (SMOTE)

Training a standard algorithm on a dataset with imbalanced class distribution leads to performance problems of the minority class in the training set [13]. Therefore, the synthetic minority oversampling technique (SMOTE) is commonly applied for imbalanced classification problems to alleviate the effect of the class imbalance on the performance of the predictive model. The synthetic oversampling technique introduces new instances of the minority class to balance the distribution.

Real-world datasets such as telecommunication, fraud detection, and medical diagnosis datasets usually have a higher number of observations of a given class under-sampled compared to the other classes [14]. An imbalanced dataset substantially compromises the machine-learning algorithm since most machine-learning algorithms expect balanced class distribution or an equal miss classification cost. For this reason, as mentioned earlier, numerous approaches have been proposed to handle imbalanced datasets. In this study, we present a synthetic minority oversampling approach. The synthetic minority oversampling approach is mathematically described as follows.

$$B_r = \left| \frac{\alpha}{y} \right| \qquad (1)$$

where, Br is the balancing ratio given by Eq. (1), B represents the imbalanced dataset, alpha indicates the subset of the sample belonging to the minority class, and y indicates the subset of the sample belonging to the majority class. The balancing process is equivalent to resampling r to a new dataset rnew, where B_r> rnew.

3.5 Random majority under sampling

The majority under-sampling is the process of balancing the dataset by randomly reducing samples from the majority class. A number of practical classification problems contain imbalanced class distributions [14]. There are two resampling techniques to handle imbalanced datasets: oversampling and undersampling techniques. Oversampling increases the number of minority class samples until the class distribution is balanced. The most employed oversampling technique is the synthetic minority over-sampling technique (SMOTE). In contrast, majority oversampling reduces the samples from the majority class until the class distribution is balanced. The most common under-sampling technique is random sampling, where the samples of the majority class are randomly removed until the class distribution is balanced.

3.6 Evaluation criteria

The most important predictive performance evaluation metric for evaluating a classifier is the uncertainty estimate of predictions [15]. To test the uncertainty of the classifier we employed prediction probability as performance metrics. Other useful metric used for the evaluation of the model is the accuracy score and receiver operating characteristics curve (ROC). The accuracy and ROC curve of a classifier is defined by Eq. (2).

$$Predicitve\ Accuracy = \frac{T}{N} \qquad (2)$$

where, T is the number of correct predictions and N is the total number of predictions (correct and incorrect) made by a given classification model. In an imbalanced dataset, the type of errors is more due to a more frequent number of one class than the other class. This is very common in practice specializing in medical datasets where the data points in the positive class are lower than the negative [16]. In such cases, accuracy might not sufficiently describe the performance score of a classification model. Accuracy is an inadequate measure of performance to quantify the predictive model performance on imbalanced chronic kidney disease classification [17]. Thus, the receiver operating characteristics (ROC) curve is employed for evaluating the classification performance of the model on chronic kidney disease detection. In addition, the confusion matrix is employed as an alternative performance measure to further investigate the model performance for imbalanced chronic kidney disease dataset classification. One of the most comprehensive ways to present the result of evaluating a binary classification model on an imbalanced classification task is a confusion matrix [18]. The confusion matrix is a two-by-two array where the rows correspond to the true class (ground truth) and the column corresponds to the predicted class.

The observations correctly predicted as chronic kidney disease patients are true positive (TP) and the observations correctly predicted as not chronic kidney patients are true negative (TN). In contrast, the observations incorrectly classified as chronic kidney disease patients are false positive (FP), and observations incorrectly predicted as not chronic kidney disease patients are true negative (TN). The TPR shows the fraction of true positives among true positive predictions given by Eq. (3).

$$TPR = \frac{TP}{TP + FN} \qquad (3)$$

In contrast, the FPR shows the fraction of false positive among the false negative predictions. The FPR is given by Eq. (4).

$$FPR = \frac{FP}{TP + TN} \qquad (4)$$

Precision: By precision defines the percentage of true positives, number of true positive (NTP), among all examples that the classifier has labeled as positive: NTP and number of false positive (NFP). The value is thus obtained by using the formula given in Eq. (5).

$$Precsion = \frac{NTP}{NTP + NFP} \qquad (5)$$

4. RESULTS AND DISCUSSION

In this study, a comprehensive experimental study on the effect of oversampling and imbalanced class distribution for chronic kidney disease prediction was presented. The SVM, random forest, k-nearest neighbor (kNN) and adaptive boosting algorithm are applied to a chronic kidney disease dataset from Kaggle online data repository.

4.1 Effect of imbalanced class distribution

In imbalanced classification problems, the use of probability provides the required level of confidence in the model outcome. Moreover, probability is an important performance measurement evaluation index which plays a role in comparing classification models [19]. The severity of skewed class distribution present in imbalanced classification takes more bias in the predicted probability as they favor the majority class. As such, it's a good idea to calibrate the predicated probability for machine learning models before evaluating their performance. Figure 2 demonstrates the predicted probabilities for imbalanced classification. In Figure 2, we observe that the decision tree model is well-calibrated as compared to other models. The calibrated probability of the classifiers is required to get the most out of the models for imbalanced classification [20]. Thus, grid searching of different probability calibration and visualization of probability calibration methods on a dataset with skewed class distribution is important for getting insights of the model performance, especially for imbalanced chronic kidney disease classification.

The probability is used as a measure of uncertainty in an imbalanced classification problem. In imbalanced classification, class labels are often insufficient in terms of selecting and evaluating a model. The predicted probability provides the basis for better model evaluation and selection. As such, using a machine learning model that predicts probability is generally preferred for imbalanced classification tasks [21-26].

In calibrated probabilities, the probability reflects the likely hood of true events. For disease classification problems, the truth of the predicted value is important because the classification outcome is used for making a high-risk medical decision. The calibrated probability matches the true likelihood of positive class observation classified as the positive class. Machine learning algorithms that need to be calibrated before use include: SVM, Decision tree, kNN, and Random forest. Machine learning algorithms that do not need to be calibrated before use include: Logistic regression, linear discriminant analysis, Naïve Bayes, and Artificial neural network. A bias in the training dataset, such as a skew in class distribution implies that the model will naturally predict a higher probability for the majority class than the minority class on average. The problem with skewed class distribution is that the model overcompensates and gives too much focus on the majority class.
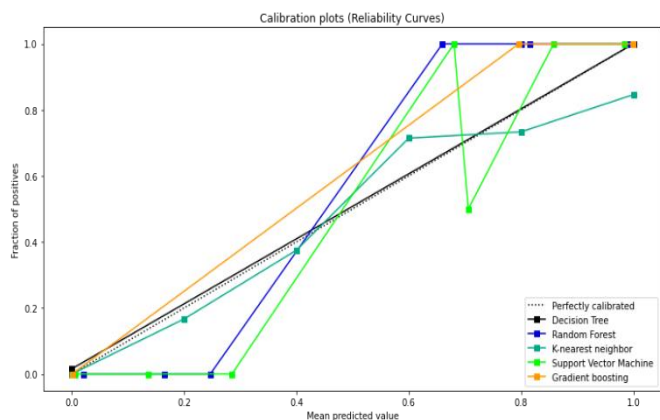

Figure 2. Probability calibration curver of classifiers

4.2 Effect of the resampling technique

The SVM, decision tree, random forest, and gradient boosting classifiers are evaluated with the receiver operating characteristics curve. The classification algorithms are tested on synthetic minority oversampling and random minority oversampling with the Near Miss algorithm. We have tested the performance of the algorithms on the imbalanced dataset and a balanced dataset with SMOTE and random oversampling approach. We observe from Figure 3, that the receiver operating characteristic curve of gradient booting, SVM, and random forest classifiers are better as compared to the decision tree, random forest, and kNN classifiers.

The area under the curve is better for SVM and random forest classifier as compared to other classifiers. The receiver operating characteristics curve for k-nearest neighbor (KNN) is lower than the other classifiers as shown in Figure 3. Thus, the KNN algorithm is severely affected by imbalanced class distribution compared to the other classifiers.

4.3 Effect of input feature quality

In this study, a sequential feature selection algorithm was used for finding the optimal set of input features from the 18 features characterizing each of the observations in the dataset. The sequential feature selection algorithm returned the optimal feature subset described in Table 4 as the best feature subset. Hence, the classification algorithm was trained on the optimal feature set selected by sequential feature selection, summarized in Table 4. Classification accuracy of 99.3% was achieved with the classification model trained on the optimal feature subset selected by the sequential feature selection algorithm. The optimal feature subset selected among the 18 features in the chronic kidney disease dataset is summarized in Table 4.

Table 4. Optimal input features selected by sequential feature selection algorithm

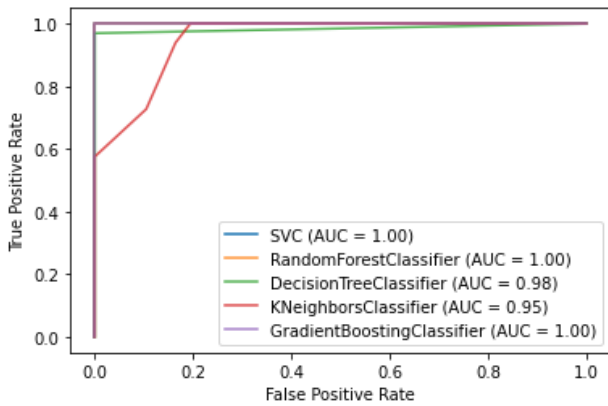| Classifier	Feature subset	Accuracy (%)	Precision	F-Score
SVM	4	98.1	1.00	0.93
	5	98.2	1.00	0.94
	6	99.3	0.99	0.93
	7	99.3	0.98	0.98
	4	98	1.00	0.91
RF	5	98.1	0.97	0.92
	6	98.2	0.93	0.91
	7	98.3	0.98	0.94
	8	98.2	0.98	0.91
	4	98.1	1.00	0.94
DT	5	98.2	0.95	0.91
	6	98.3	09.3	0.92
	7	98.3	0.97	0.92
	8	98.3	0.98	0.93
	4	98.1	1.00	1.00
GB	5	98.1	0.98	0.99
	6	98.2	1.00	1.00
	7	98.2	0.96	0.92
	8	98.2	0.98	0.99
	4	97.02	1.00	0.91
KNN	5	97	0.93	0.91
	6	93.1	0.97	0.92
	7	95	0.96	0.94
	8	98.1	0.92	0.93

Figure 3. ROC of classifiers

5. CONCLUSIONS

In this study, a comprehensive experimental study on the effect of resampling and imbalanced class distribution of chronic kidney disease prediction was presented. Support vector machine (SVM), random forest, k-nearest neighbor (kNN), and adaptive boosting algorithms are applied to a real-world chronic kidney disease dataset obtained from Kaggle online data repository. In addition, examined how the input feature quality affects the models performance for classification.

The simulation result shows that the SVM, decision tree, random forest, and kNN models have better performance when trained on optimal features selected through a sequential feature selection algorithm. Furthermore, we have explored the effect of skewness of class distribution on the performance of standard machine learning models. Finally, we have implemented the state-of-the-art chronic kidney disease prediction model which achieved a classification accuracy of 99.3%. Furthermore, we have analyzed the models performance for risky medical decision-making with probability calibration. The results show that the model is robust and can be applied in medical decision support systems for the diagnosis of chronic kidney disease.

The limitation of this study is that the study was conducted only on a chronic kidney disease dataset. In addition, the study employed five supervised learning algorithms for comparison. In future work, the authors will test the resampling techniques with other supervised learning methods such as logistic regression, Naïve Bayes, and deep learning methods to investigate the effectiveness of the resampling technique.

REFERENCES

[1] Gupta, R., Koli, N., Mahor, N., Tejashri, N. (2020). Performance analysis of machine learning classifier for predicting chronic kidney disease. In 2020 International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1-4. https://doi.org/10.1109/INCET49848.2020.9154147

[2] Gameiro, J., Branco, T., Lopes, J.A. (2020). Artificial intelligence in acute kidney injury risk prediction. Journal of Clinical Medicine, 9(3): 678. https://doi.org/10.3390/jcm9030678

[3] Cao, L., Shen, H. (2017). Imbalanced data classification based on hybrid resampling and twin support vector machine. Computer Science and Information Systems, 14(3): 579-595. https://doi.org/10.2298/CSIS161221017L

[4] Andrić, K., Kalpić, D., Bohaček, Z. (2019). An insight into the effects of class imbalance and sampling on classification accuracy in credit risk assessment. Computer Science and Information Systems, 16(1): 155-178. https://doi.org/10.2298/CSIS180110037A

[5] Komal, K.N., Tulasi, R.L., Vigneswari, D. (2019). An ensemble multi-model technique for predicting chronic kidney disease. International Journal of Electrical and Computer Engineering, 9(2): 1321-1326. https://doi.org/10.11591/ijece.v9i2.pp1321-1326

[6] Polat, H., Mehr, H.D., Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. Journal of Medical Systems, 41: 1-11. https://doi.org/10.1007/s10916-017-0703-x

[7] Dritsas, E., Trigka, M. (2022). Machine learning techniques for chronic kidney disease risk prediction. Big Data and Cognitive Computing, 6(3): 98. https://doi.org/10.3390/bdcc6030098

[8] Rustam, Z., Amalia, Y., Hartini, S., Saragih, G.S. (2021). Linear discriminant analysis and support vector machines for classifying breast cancer. IAES International Journal of Artificial Intelligence, 10(1): 253-257. https://doi.org/10.11591/ijai.v10.i1.pp253-256

[9] Assegie, T.A., Tulasi, R.L., Kumar, N.K. (2021). Breast cancer prediction model with decision tree and adaptive boosting. IAES International Journal of Artificial Intelligence, 10(1): 184-190. https://doi.org/10.11591/ijai.v10.i1.pp184-190

[10] Bai, Q., Su, C., Tang, W., Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. Scientific Reports, 12(1): 1-8. https://doi.org/10.1038/s41598-022-12316-z

[11] Khuat, T.T., Le, M.H. (2019). Ensemble learning for software fault prediction problem with imbalanced data. International Journal of Electrical & Computer Engineering (2088-8708), 9(4): 3241-3246. https://doi.org/10.11591/ijece.v9i4.pp3241-3246

[12] Seo, J.H., Kim, Y.H. (2018). Machine-learning approach to optimize SMOTE ratio in class imbalance dataset for intrusion detection. Computational Intelligence and Neuroscience, 2018: 9704672. https://doi.org/10.1155/2018/9704672

[13] Alghamdi, M., Al-Mallah, M., Keteyian, S., Brawner, C., Ehrman, J., Sakr, S. (2017). Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project. PloS One, 12(7): e0179805. https://doi.org/10.1371/journal.pone.0179805

[14] Sushma, S.J., Assegie, T.A., Vinutha, D.C., Padmashree, S. (2021). An improved feature selection approach for chronic heart disease detection. Bulletin of Electrical Engineering and Informatics, 10(6): 3501-3506. https://doi.org/10.11591/eei.v10i6.3001

[15] Elhoseny, M., Shankar, K., Uthayakumar, J. (2019). Intelligent diagnostic prediction and classification system for chronic kidney disease. Scientific Reports, 9(1): 9583. https://doi.org/10.1038/s41598-019-46074-2

[16] Rustam, Z., Sudarsono, E., Sarwinda, D. (2019). Random-forest (RF) and support vector machine (SVM) implementation for analysis of gene expression data in

chronic kidney disease (CKD). In IOP Conference Series: Materials Science and Engineering, 546(5): 052066. https://doi.org/10.1088/1757-899X/546/5/052066

[17] Almustafa, K.M. (2021). Prediction of chronic kidney disease using different classification algorithms. Informatics in Medicine Unlocked, 24: 100631. https://doi.org/10.1016/j.imu.2021.100631

[18] Sisodia, D.S., Verma, A. (2017). Prediction performance of individual and ensemble learners for chronic kidney disease. In 2017 International Conference on Inventive Computing and Informatics (ICICI), pp. 1027-1031. https://doi.org/10.1109/ICICI.2017.8365295

[19] Zhang, H., Hung, C.L., Chu, W.C.C., Chiu, P.F., Tang, C.Y. (2018). Chronic kidney disease survival prediction with artificial neural networks. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1351-1356. https://doi.org/10.1109/BIBM.2018.8621294

[20] Salekin, A., Stankovic, J. (2016). Detection of chronic kidney disease and selecting important predictive attributes. In 2016 IEEE International Conference on Healthcare Informatics (ICHI), pp. 262-270. https://doi.org/10.1109/ICHI.2016.36

[21] Dai, Y., Xu, B., Yan, S., Xu, J. (2020). Study of cardiac arrhythmia classification based on convolutional neural network. Computer Science and Information Systems, 17(2): 445-458. https://doi.org/10.2298/CSIS191229011D

[22] Abdel-Fattah, M.A., Othman, N.A., Goher, N. (2022). Predicting chronic kidney disease using hybrid machine learning based on Apache spark. Computational Intelligence and Neuroscience, 2022: 9898831. https://doi.org/10.1155/2022/9898831

[23] Tamyalew, Y., Salau, A.O., Ayalew, A.M. (2023). Detection and classification of large bowel obstruction from X-ray images using machine learning algorithms. International Journal of Imaging Systems and Technology, 33(1): 1-17. https://doi.org/10.1002/ima.22800

[24] Kiflie, A., Tufa, G.T., Salau, A.O. (2023). Sputum smears quality inspection using an ensemble feature extraction approach. Frontiers in Public Health, 10: 1-11. https://doi.org/10.3389/fpubh.2022.1032467

[25] Assegie, T.A., Salau, A.O., Omeje, C.O., Braide, S.L. (2023). Multivariate sample similarity measure for feature selection with a resemblance model. International Journal of Electrical and Computer Engineering, 13(3): 3359-3366. https://doi.org/10.11591/ijece.v13i3.pp3359-3366

[26] Haile, M.B., Salau, A.O., Enyew, B., Belay, A.J. (2022). Detection and classification of gastrointestinal disease using convolutional neural network and SVM. Cogent Engineering, 9(1): 2084878. https://doi.org/10.1080/23311916.2022.2084878

NOMENCLATURE

SVM	Support Vector Machine
DT	Decision Tree
RF	Random Forest
KNN	K-Nearest Neighbor
GB	Gradient Boosting
CKD	Chronic Kidney Disease
ROC	Receiver Operating Curve
Max	Maximum
Min	Minimum
STD	Standard Deviation
TPR	True Positive Rate
FPR	False Positive Rate