Journal of Information Hiding and Multimedia Signal Processing ©2024 ISSN 2073-4212
Ubiquitous International Volume **15**, Number **4**, December **2024**

Coronary Artery Disease Prediction by Combining Three Classifiers

Ghada Abdallah Radwan[1,3], Mohamed Helmy Khafagy[2],
Mostafa Thabet Mohamed Mabrouk[1], Marwa Hussien Mohamed[3,4]*

[1]Information System Department, Faculty of Computers and Artificial Intelligence,
Fayoum University, Fayoum 63514, Egypt.
ga1328@fayoum.edu.eg, mtm00@fayoum.edu.eg

[2]Computer Science Department, Faculty of Computers and Artificial Intelligence,
Fayoum University, Fayoum 63514, Egypt.
mhk00@fayoum.edu.eg

[3]Information System Department, Information System and Computer Science Faculty,
October 6 University, 6th of October 12573, Egypt.
ghada.abdallah.csis@o6u.edu.eg, eng_maroo1@yahoo.com

[4]Computer Technology Engineering Department, Engineering Technologies Collage,
Al-Esraa University, Baghdad, 10081 IRAQ.
maraw@esraa.edu.iq

*Corresponding author: Marwa Hussien Mohamed

Received August 24, 2024, revised September 21, 2024, accepted September 28, 2024.

Abstract. *Coronary artery disease is one of the important diseases many people have suffered from. The accuracy of the prediction needs to increase to save more people's lives from this disease. This research aims to improve accuracy by building a new proposed model using eight machine-learning models. In the Z-Alizadeh Sani Tabular dataset, an Integer, Real Feature Type, is used, and the synthetic minority sampling technique was combined with the Edited Nearest Neighbors (SMOTE-ENN) to solve the imbalance in the dataset. Using Chi-square and Recursive Feature Elimination (RFE) techniques to select Twenty-six from fifty-six features in the datasets to minimize the complexity of calculations and save processing time. Furthermore, every machine learning classifier is applied twice before and after using Smote-Enn. Then, an ensemble classification was applied based on the highest accuracy principle. Experimental results show that ensemble classification is superior to a single classifier. The fusion between logistic regression (L.R.), light gradient boosting (LightGBM), and adaptive boosting (AdaBoost) performed the highest, obtaining an accuracy of 98.9%.*
Keywords: coronary artery disease; feature analysis; machine learning; heart disease.

1. **Introduction.** CAD is one of the major diseases that cause death nowadays. According to the World Health Organization, millions of people across the world die from CAD every year. In essence, the condition of the heart that interferes with the normal functionalities of the heart is called heart disease [1]. However, the most common cause of heart disease is the narrowing or blockage of the coronary arteries, which supply blood to the heart. It is estimated that over 30 million deaths will take place in the world alone by the year 2040 [2]. Prevention of heart diseases, including CAD, is largely based on lifestyle changes and healthier habits. A good diet, regular exercise, and cessation of injurious behaviors help the arteries remain healthy, less prone to injury, and free from atherosclerotic

plaques. Cardiovascular good health may be enhanced by quitting smoking, performing regular physical activity, managing blood pressure, low cholesterol and prevention of diabetes, maintaining a stable body weight, minimizing stress, following a Mediterranean diet, reducing salt, avoiding high-fat foods, eating high-fibre foods, and limiting alcohol consumption [3,4]. Therefore, it is necessary to detect or predict the disease during an early stage in infected patients to prevent it and consequently minimize possible loss of lives. Various artificial intelligence techniques [5], such as data mining, machine learning, deep learning, and expert systems, are adopted in healthcare industries for the reasons of diagnosis, detection, and prediction of many diseases like diabetes, waterborne diseases, COVID-19, malaria, typhoid, etc.

Still, one such method useful in building models applied to the healthcare sector to diagnose diseases is machine learning (ML) [6]. ML is a concept in artificial intelligence applied to create models or systems capable of learning from the available data set to make valuable predictions and decisions related to future events [7]. Beginning with the most primitive data collection stage, ML undergoes several steps to knowledge extraction with insightful patterns [8,9]. This roadmap includes cleaning, transformation, selection, evaluation, and knowledge presentation to present the explored insight to users. It has further enabled systems to learn from diagnostic data, identify valuable patterns during the learning process, and reduce human intervention in decision-making [10,11]. These technologies frequently yield better-predicted performance.

The main contributions of the present research work are the following:

- Enhancement of the accuracy of the prediction of early CAD patients.
- Handling the class imbalance problem is performed using the SMOTE-ENN method.
- Feature selection was conducted using Chi-square and RFE techniques, with the most significant 26 features picked to minimize calculations' processing time and complexity.
- Verify eight single algorithms before and after data balance.
- Using hybrid classification up to the third level of fusion.

The rest of this article is organized as follows: Section 2 provides the Related work about the study. Section 3 describes materials and methods. Experimental results are presented in Section 4. A discussion of this study is presented in Section 5. The conclusion of this article and future studies are presented in Section 6.

2. Related work.

Previous studies on CAD prediction will be covered in this part. In [5], the authors used machine learning (ML) techniques to predict CAD using a dataset acquired from the two public hospitals in Kano State, Nigeria. In their study, the Random Forest model was the best, with an accuracy of 92.04%.

In [12], the authors propose a new model to predict CAD; they solve the problem of the logistic regression model by using feature selection and feature mixture of tree machine learning algorithms, then test their model on the Z-Alizadeh Sani dataset. Five various methods to solve the balance problem in the dataset. Accuracy result was 94.7%.

In [13], the authors write a comparative study comparing seven intelligence models to forecast coronary artery illness. Statlog and Cleveland Cardiology datasets were used to predict CAD. In their research, the deep neural network was the best, with an accuracy of 98.15%.

In [14], the authors applied eleven machine-learning models to know the main features of CAD prediction. Multiple feature combinations and classification algorithms were used. In their study, Random Forest was the best, with an accuracy of 96.28%.

In [15], the authors recommended a new machine-learning model for CAD diagnosis. The Z-Alizadeh-Sani dataset was used. Using the dataset's correlation coefficient, the

Table 1. Comparison Between CAD Prediction Methods.

Ref. NO.	Dataset	Method	The highest accuracy	Research Limitations
[5]	The dataset for CAD from the two general hospitals in Kano State, Nigeria, was used in this study.	Random Forest	92.04%	- Small dataset.
[12]	The Z Alizadeh dataset	LightGBM + LR + SMOTE Tomek	94.7%	- Feature selection methods not used.
[13]	Statlog dataset.	Deep Neural Network	98.15%	- Feature selection methods not used. - The number of features is very small.
	Cleveland dataset.	SVM	97.36%	- Small dataset. - The number of features is very small.
[14]	Heart Disease dataset retrieved from the UCI repository.	Random Forest	96.28%	- Small dataset. - The number of features is very small.
[15]	Z-Alizadeh Sani dataset.	Naïve Bayes	83%	- No method was used to s olve the imbalance problem. - Small dataset.
[16]	The dataset was collected from a medical centre and hospitals in Bangladesh.	Random Forest	90%	- No method was used to solve the imbalance problem. - The dataset used is very small and contains 59 patients.
[17]	Coronary Prediction Dataset.	Stacking + SMOTE	90.9%	- The number of features is very small - Lack of some influencing features for disease detection.
[18]	Z-Alizadeh Sani dataset.	SVM	89.5%	- No method was used to solve the imbalance problem.
[19]	Z-Alizadeh Sani dataset.	Random Trees	91.47%	- No method was used to solve the imbalance problem.
[20]	Z-Alizadeh Sani dataset.	Artificial Neural Network	93.35%	- No method was used to solve the imbalance problem. - Feature selection methods not used.

feature selection was done. The experiment was then conducted using 10-fold cross-validation partitioning techniques. In this study, Naive Bayes was the best, with an accuracy of 83%.

In [16], the authors analyzed many different components of patient data to predict CAD. The correlation-based Features Subset Evaluation Method with Optimal First Lookup was utilized to determine the most relevant attributes for CAD forecasting. The best essential factors for discovering heart illness. Many artificial intelligence techniques were applied to two types of heart disease datasets. The authors use all dataset features in their experiments and select a split of features to get the model's accuracy. The random forest algorithm has the highest results, at 90%.

In [17], the authors applied different machine-learning classifiers before using the synthetic minority oversampling approach SMOTE and after using it for CAD prediction. The stacking ensemble model achieved the best accuracy of 90.9%.

In [18], the authors utilized five supervised classification machine-learning methods operated on the Z-Alizadeh Sani dataset for CAD prediction. In their study, the SVM was the best, which achieved an accuracy of 89.5%.

In [19], the authors proposed a new method to increase the accuracy of CAD diagnosis using several machine-learning models by selecting the most effective aspects in the Z-Alizadeh Sani dataset. In their study, the random tree was the best, which achieved an accuracy of 91.47%.

In [20], the authors proposed a system using a machine learning classifier for CAD prediction. They used the Z-Alizadeh Sani dataset to identify the appropriate classifier and compare three classifiers. The SMOTE was used to resolve the class imbalance trouble. In their study, the ANN was the best, which achieved an average accuracy of 93.35% ± 2.56%.

A list of all research limitations and datasets used is shown in Table 1.

3. **Material And Methodology.** Proposed a robust method of CAD detection using the Z-Ali Zadeh Sani dataset on machine learning in Python. Now, this whole methodology has been broadly divided into parts. Each subsection shall be explained below, and a block diagram in Figure 1 shall illustrate the steps of the proposed model.

1. Importing datasets.
2. Cleaning the data: Renaming Columns and defining the replacement values for every column.
3. Scaling of data using Normalization: Minimum – Maximum Scaler to be within the range 0-1.
4. Feature Selection for Categorical Variables Using the Chi-Square Test: Conduct hypothesis testing such that hypothesis zero, H0, has no relationship between features and whether a patient develops CAD. Hypothesis one, H1, is that there will be a statistically significant relationship between features and whether a patient will develop CAD.

Alpha (p-value) = 0.05 (significance level) indicates that the probability of wrongfully rejecting the null hypothesis is 5%. So, to "reject" H0, the p-value needs to be greater than 0.05, and the value of chi-square ¡= critical value of chi-square.

One of the wrapper methods for feature selection is RFE, which selects only the continuous, numeric column features.

1. Splitting the dataset into a training set and testing set.
2. Applying classifiers before handling imbalance problems and recording these results.
3. The SMOTE-ENN was applied to solve the imbalance problem.
4. Applying classifiers again to compare results before using (SMOTE-ENN) and after using it.
5. Execute the process of learning.
6. Determine the model's performance on the test set.

3.1. **Data Preprocessing.** The dataset used in the study is the Z-Alizadeh Sani, retrieved from the UCI Machine Learning Repository [21]. This dataset includes 303 medical records from 303 patients who came to the Shaheed Rajaei Hospital complaining of chest pain. There are 55 features total in each record. These 303 samples were then divided into two groups: the normal class and the CAD class [22,23]. A sample is classified as CAD class if its lumen stenosis of the coronary arteries reaches or exceeds 50%; otherwise, it falls into the normal class. According to this definition, among 303 samples, 216 cases are classified as the CAD class and 87 instances as a normal class [12].

The following steps are applied to process the dataset:

• Data Cleaning: If any invalid and corrupted data existed in the data set, the same was removed, while incorrectly formatted data within the data set were repaired. Column names could be fixed, and define replacement values for each column in the dataset.

• Feature Encoding: The categorical variables were then label-encoded in a numerical format. This means the order should be maintained while encoding the labels. As a result, the sequence should correspond to the encoding. The label encoding approach converts every label into an integer value [21]. Features encoding would be applied to the dataset, including the use of the gender feature from female and male, obesity, chronic renal failure, cerebrovascular accident, airway, thyroid disease, congestive heart failure, dyslipidemia, weak peripheral pulse, lung rales, systolic murmur, diastolic murmur, dyspnea, atypical, nonanginal, exertional chest pain, low threshold angina, LVH, poor r progression. All these Y and N were encoded to 1 and 0, respectively. Also, the Cath columns, CAD and normal, were encoded to 1 and 0. Table 2 shows the dataset before encoding, and Table 3 shows the data after encoding.

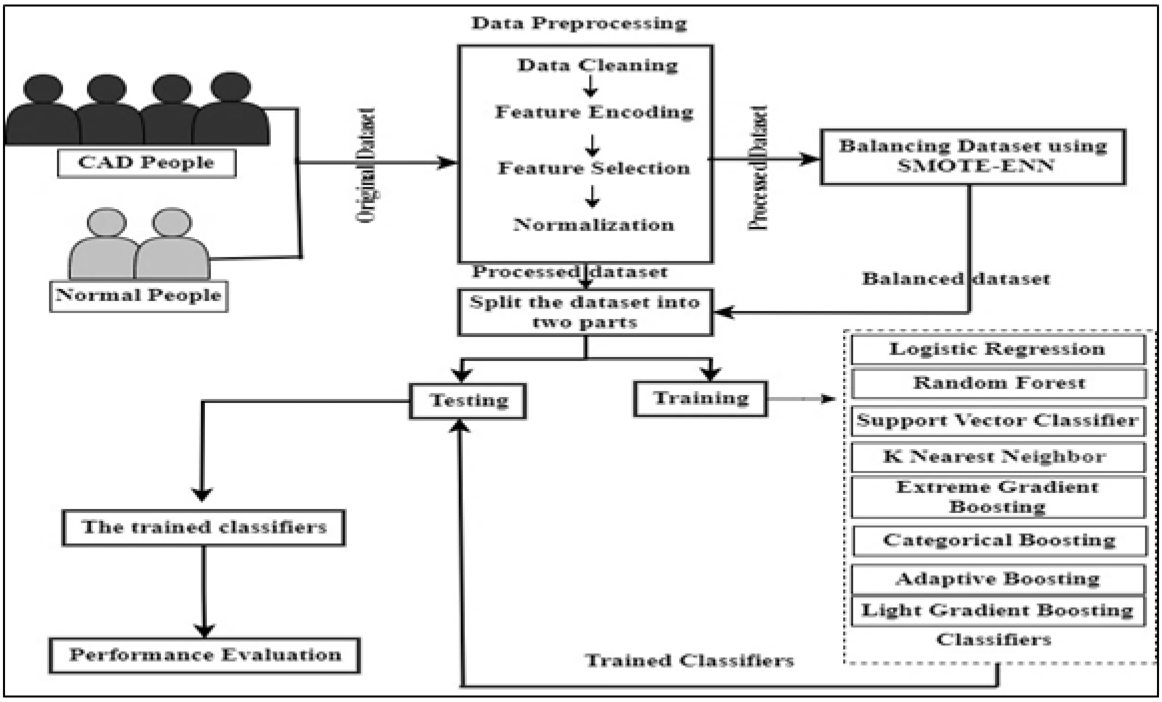

FIGURE 1. The proposed diagram of Coronary Artery Disease Prediction

TABLE 2. A sample of the dataset before applying feature encoding.

No.	Age	Weight	Length	Sex	BMI	DM	HTN
0	53	90	175	Male	29.83	0	1
1	67	70	157	Female	28.39	0	1
2	54	54	164	Male	20.07	0	0
3	66	67	158	Female	26.83	0	1
4	50	87	153	Female	37.16	0	1

TABLE 3. A sample of the dataset after applying feature encoding.

No.	Age	Weight	Length	Sex	BMI	DM	HTN
0	53	90	175	0	29.83	0	1
1	67	70	157	1	28.39	0	1
2	54	54	164	0	20.07	0	0
3	66	67	158	1	26.83	0	1
4	50	87	153	1	37.16	0	1

• Feature Selection (F.S.): F.S. techniques are useful auxiliary techniques applied at the machine learning level to achieve a valid and reliable model. The F.S. techniques propose new subsets of features and evaluate each feature's prediction potential about the target variable. The goals of the F.S. techniques are to decrease the complexity of the developed model, avoid noise, prevent overfitting, speed up the time execution of the machine learning algorithms, and improve performance results [23]. In this work, the chi-square test and RFE methods were used, and twenty-six features were selected, as shown in Table 4.

• Normalization: This method scales the values of a feature between 0 to 1 [24]. Maximum and minimum normalization techniques have been applied to these attributes to

TABLE 4. The selected features using the Chi-square and RFE methods.

Feature name	**Feature name**
Sex	Potassium (K)
Diabetes Mellitus (DM)	Function Class
Hypertension (HTN)	Airway Disease
Family History	Weak peripheral pulse
Typical Chest Pain	Lung Rales
Dyspnea	Diastolic murmur
Atypical	Low Threshold angina
Nonanginal	Left Ventricular Hypertrophy (LVH)
Q Wave	Poor R Wave Progression (Poor R Progression)
S.T. Elevation	Creatine (C.R.)
St Depression	Hemoglobin (H.B.)
T inversion	Valvular Heart Disease (VHD)
BBB_LBBB	Regional Wall Motion Abnormality

normalize them. Amongst the most common methods to process the data is maximum and minimum Normalization, which mathematically can be defined as:

$$\text{Normalized } x = \frac{x(\text{original}) - \min x}{\max x - \min x} \tag{1}$$

Here, x is the input feature, and max and min denote the values of maximum and minimum, respectively. The output value that results after Normalization is the normalized x. This approach was used in this research for feature scaling. Probable significance correlations between the features may be useful to find. See Table 5, which describes data after Normalization.

TABLE 5. A sample of the dataset after Normalization.

No	**Age**	**weight**	**Length**	**Sex**	**BMI**	**DM**	**HTN**
0	0.53	0.20	0.35	1.0	0.30	1.0	1.0
1	0.58	0.22	0.41	1.0	0.27	1.0	1.0
2	0.44	0.51	0.62	0.0	0.47	0.0	0.0
3	0.21	0.20	0.47	1.0	0.22	0.0	0.0
4	0.46	0.19	0.29	1.0	0.33	0.0	0.0

3.2. **Imbalance class problem handling.** The distribution of the target classes in the dataset is wildly skewed, with around three times as many CAD patients as healthy participants [25]. SMOTE-ENN is used to address the imbalance issue in this situation. The fundamental concept of the SMOTE method is to examine minority classes and use oversampling to create new minority classes. In this way, the instances of the dataset were distributed in a balanced way, and the noise problem caused using SMOTE was solved using the ENN method, allowing us to design powerful classification models to ensure a very accurate prediction of the occurrence of coronary artery disease.

3.3. **Utilizing single classification.** Classification is the process of designing a classifier for identifying and separating classes of data so that it can anticipate or predict the class units or entities with unknown values of class labels. The study of the training dataset will determine the presumptive model. Various techniques, such as decision trees and basic classification criteria, would characterize the derived model. There are two phases of data classification: first, a classifier representing a predefined set of concepts or data classes is built. In building the classifier at training, a classification method learns from a training dataset and the corresponding class label columns or attributes [26,27]. Then, the model is used for prediction in the next step. The prediction accuracy for the classifier is estimated based on a separate set of training cases. The best classification techniques have been tested to determine which classifier gives the best result on the dataset.

3.4. **Utilizing fusion multi-classifiers.** A merging of classifiers is combining many classifiers to attain optimum accuracy. It is a group of classifiers whose independent predictions are fused in some technique to classify new.

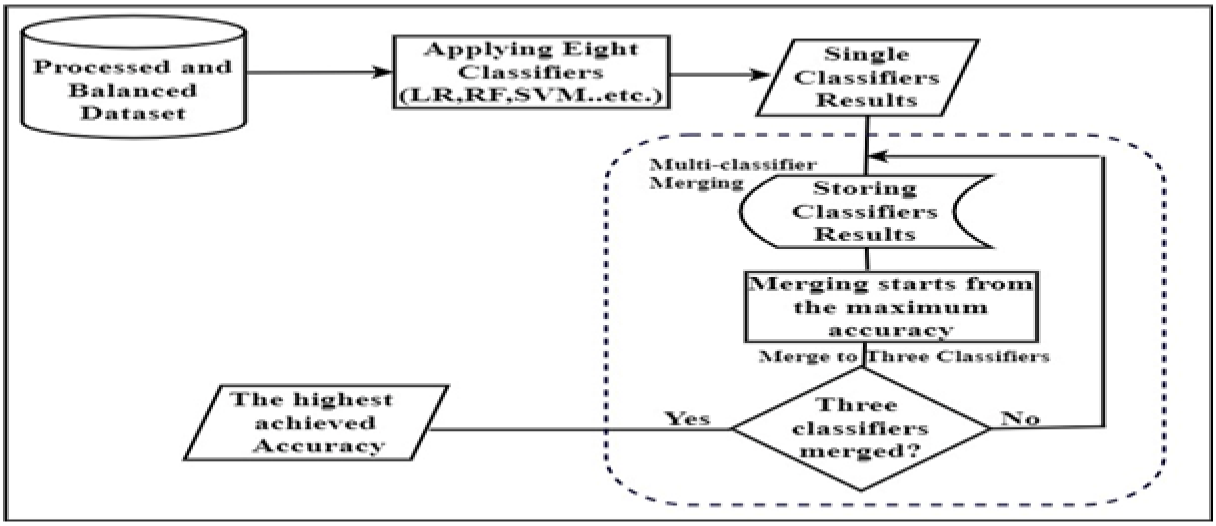

FIGURE 2. Proposed method of Multi-Classifier Fusion for Coronary Artery Disease prediction.

Voting is the class for uniting classifiers. Figure 2. Describes the proposed method of Multi-Classifier Fusion for CAD prediction.

• The multi-classifier merging procedure begins with the classifier that got the best accuracy, incorporating predictions from other single classifiers to increase accuracy based on the outcomes of utilizing a single classification.

• Selecting the best accuracy across all processes by repeating the same procedure up to the third level of merging.

4. **EXPERIMENTAL RESULTS.** Two experiments were conducted to calculate the proposed model: the first used the single classifier, while the second used the multi-classifier fusion task. The performance of the two experiments was calculated using four evaluation metrics: accuracy, precision, recall, and f-score. The elements of the confusion matrix are false positive (fp), false negative (fn), true positive (tp), and true negative (tn) [5]. These are the definitions of the previous measures.

4.1. **The Metrics of Evaluation.**

1. **Accuracy:** It computes the proportion of the correctly predicted instances to all the cases available in the data and reports how well the classification process works. The accuracy can be obtained from the following equation:

$$\text{Accuracy} = \frac{\text{tn} + \text{tp}}{\text{tn} + \text{tp} + \text{fn} + \text{fp}} \tag{2}$$

2. **Precision:** Displays the proportion of true positive subjects to false positive subjects for each variable. The following equation can be used for calculating the precision:

$$\text{Precision} = \frac{\text{tp}}{\text{fp} + \text{tp}} \tag{3}$$

3. **Recall:** The sensitivity, represented using the following equation, indicates the proportion of CAD-positive patients the models accurately predicted. The following equation can be used for calculating the recall:

$$\text{Recall} = \frac{\text{tp}}{\text{fn} + \text{tp}} \tag{4}$$

4. **F1 Score:** F1 scores fall in the range of zero to one. As the highest possible value is one and the worst possible value is zero, F1 is expressed as follows:

$$\text{F1 Score} = \frac{2 \times \text{tp}}{\text{fp} + \text{fn} + 2 \times \text{tp}} \tag{5}$$

4.2. **The Experiment of Single Classifier.** The accuracies of the eight single classifiers (L.R., SVM, R.F., KNN, XGBoost, CatBoost, LightGBM, and AdaBoost) are compared before and after using SMOTE-ENN in Table 6 and Table 7, respectively. The performance of all single classifiers is shown in Figure 3 below before implementing the SMOTE-ENN process. For instance, a comparison of single classifiers in terms of accuracy, precision, recall, and F1 measure before applying SMOTE–ENN is shown in Figure 4, Figure 5, and Figure 6, Figure 7, respectively. While applying SMOTE–ENN, the LightGBM is the best among the hall of fame in protecting the best accuracy, precision, recall, and F1 scores of 87.91%, 89.55%, and 91.60%, respectively, but CatBoost garners the highest recall at 95%.

Table 6. Single Classifiers' Best Performance Before Using SMOTE-ENN.

Machine Learning Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
LR	86.8%	88%	93.8%	90.9%
SVM	84.6%	89.1%	89.1%	89%
R.F.	81%	82.2%	93.5%	87.6%
KNN	80%	83.8%	89.1%	86.4%
XGBoost	85.7%	87%	93.1%	90.2%
CatBoost	83.5%	83.6%	95.3%	89.1%
AdaBoost	81%	82.2%	93.6%	87.6%
LightGBM	87.9%	89.6%	93.6%	91.6%

Applying SMOTE–ENN gives much better results. The results in Table 4 show that all classifiers improved significantly, whereas LightGBM increased by 9.1%. Moreover, the standard deviations of the models also indicate a much more stable performance. It is evident that the two classifiers for the highest performance improvement are the L.R. accuracy score, which improved from 86.8% to 98.5%, and the accuracy score of AdaBoost also improved from 81% to 98.5%, that is, by 11.7% and 17.5%, respectively. The models' recalls and f1-scores improved as predicted with better results. L.R. and AdaBoost had the same outcomes, achieving 96.7% and 98%, respectively.

But AdaBoost was better than LR. at precision, which achieved 95.5%, while the L.R. achieved 94%. Figure 11 shows the performance of all single classifiers after applying SMOTE – ENN.

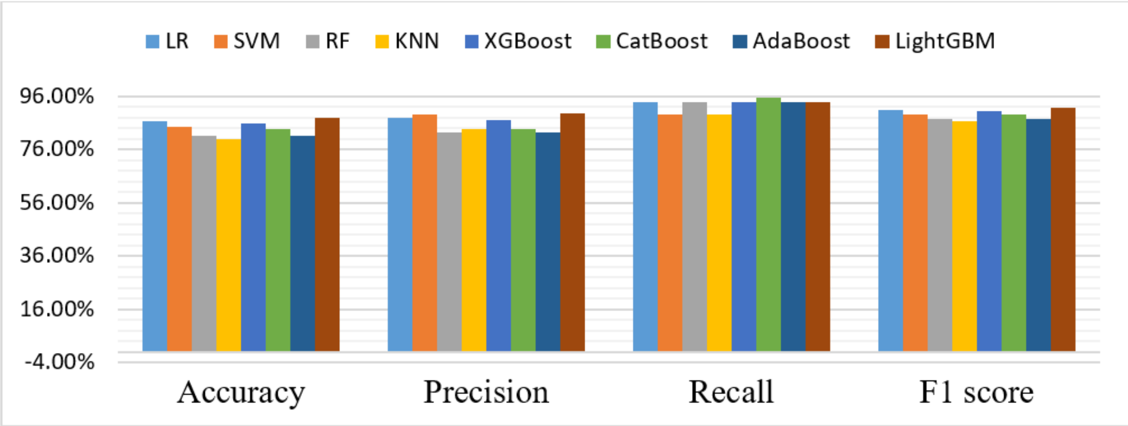

FIGURE 3. Comparison of single classifier performance before using SMOTE – ENN.

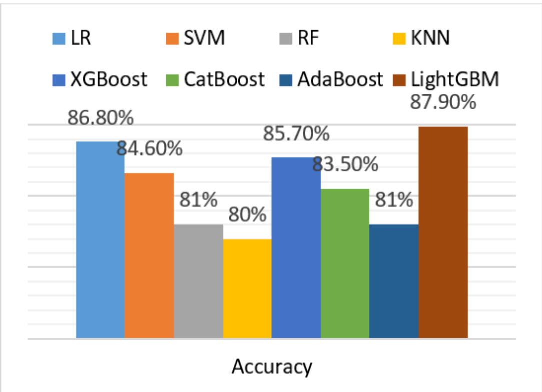

FIGURE 4. Comparison of single classifier accuracies before using SMOTE – ENN.

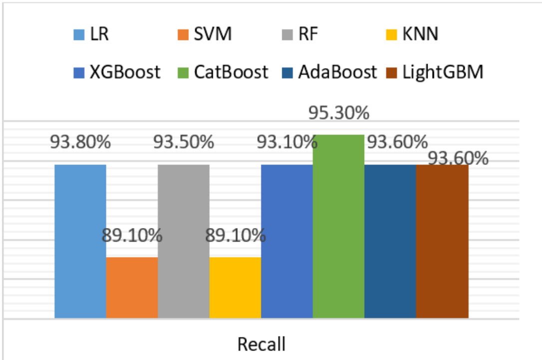

FIGURE 5. Comparison of single classifier recalls before using SMOTE – ENN.

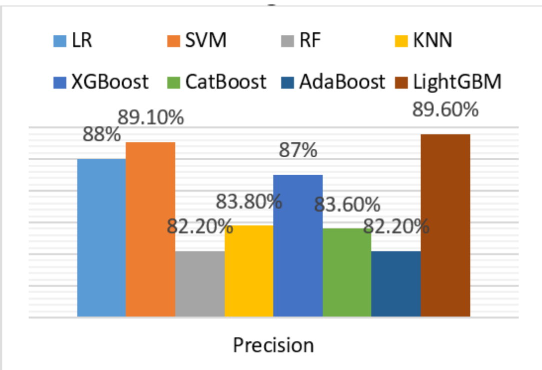

FIGURE 6. Comparison of single classifier precisions before using SMOTE – ENN.

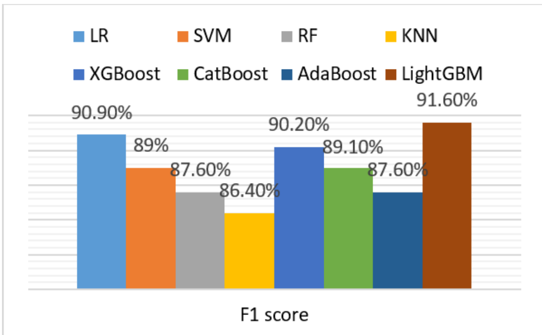

FIGURE 7. Comparison of single classifier F1 scores before using SMOTE – ENN.

The comparison between single classifiers based on accuracy, precision, recall, and f1 score after applying SMOTE – ENN are displayed in Figure 9, Figure 10, Figure 11, and Figure 12, respectively.

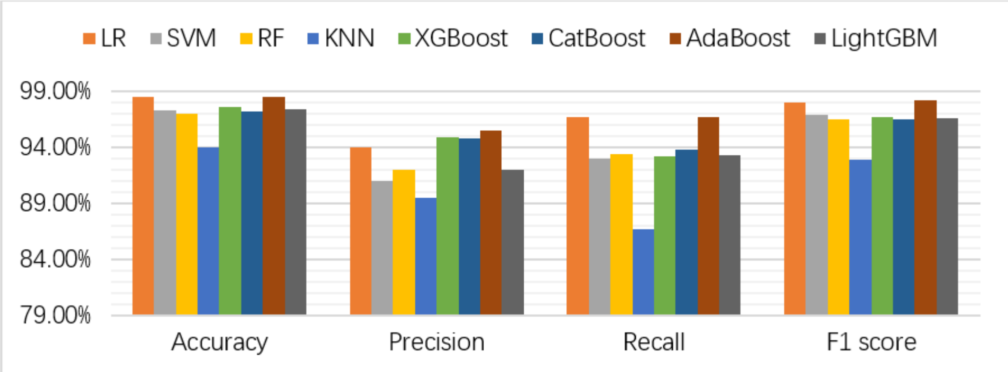

FIGURE 8. Comparison of single classifier performance after using SMOTE – ENN.

TABLE 7. Single Classifiers' Best Performance After Using SMOTE-ENN.

Machine learning model	**Accuracy**	**Precision**	**Recall**	**F1 score**
LR	98.5%	94%	96.7%	98%
SVM	97.3%	91%	93%	96.9%
R.F.	97%	92%	93.4%	96.5%
KNN	94%	89.5%	86.7%	92.9%
XGBoost	97.6%	94.9%	93.2%	96.7%
CatBoost	97.2%	94.8%	93.8%	96.5%
AdaBoost	98.5%	95.5%	96.7%	98.2%
LightGBM	97.4%	92%	93.3%	96.6%

4.3. **Fusion Classification process for multi-classifiers.** When comparing the performance of the classifiers before and after solving the problem of data imbalance, it was found that the classifiers achieved the best results after solving the problem using the SMOTE-ENN, so the merging of the multi-classifier process was applied after balancing the dataset using SMOTE-ENN.

TABLE 8. The second level of merging accuracies.

2^{nd} level of merging	**Accuracy**
L.R. + SVM	93.3%
L.R. + RF	96.7%
L.R. + KNN	90.0%
L.R. + XGBoost	94.5%
L.R. + CatBoost	96.7%
L.R. + AdaBoost	86.7%
L.R. + LightGBM	97.5%

Table 9 displays the comparison of the best accuracies for the third level of the multi-classifier merging process.

Table 10 displays the comparison of the best accuracies for different levels of the multi-classifier merging process.

In the second level of fusion, the best accuracy of merging two classifiers was between L.R. and LightGBM, which achieved an accuracy of 97.55%.

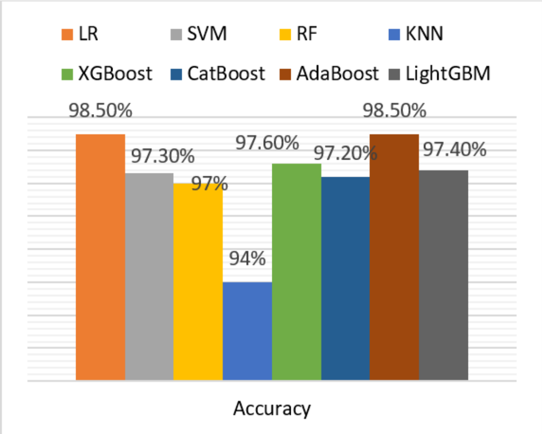

Figure 9. Comparison of single classifier accuracies after using SMOTE – ENN.

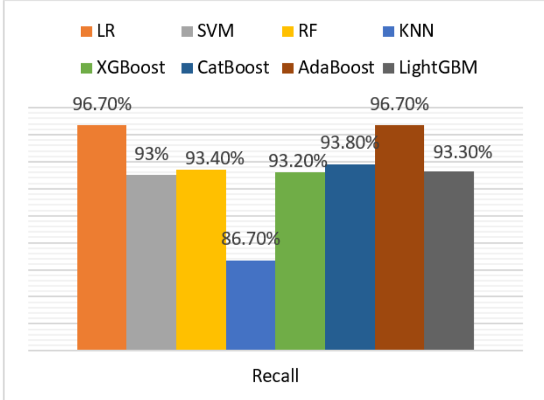

Figure 10. Comparison of single classifier recalls after using SMOTE – ENN.

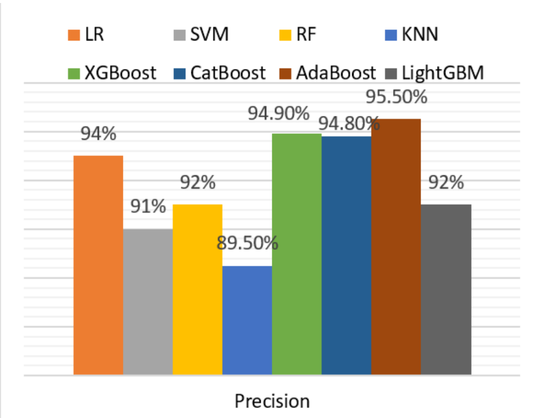

Figure 11. Comparison of single classifier precisions after using SMOTE – ENN.

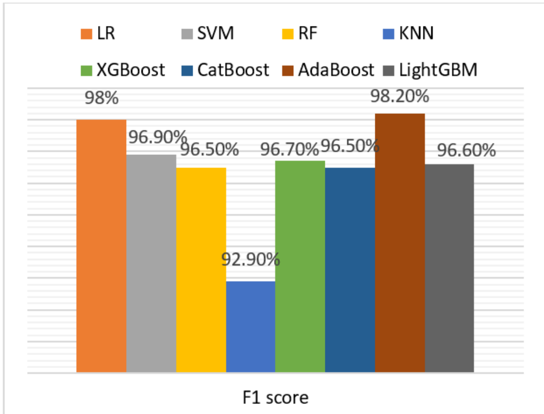

Figure 12. Comparison of single classifier F1 scores after using SMOTE – ENN.

In the third level of merging, the best accuracy of the three merging classifiers was between LR, LightGBM, and AdaBoost, which achieved an accuracy (of 98.86%).

Table 9. The third level of merging accuracies.

3$^{\text{rd}}$ level of merging	**Accuracy**
LR + LightGBM + AdaBoost	98.9%
LR + LightGBM + SVM	95.7%
LR + LightGBM + XGBoost	92.6%
LR + LightGBM + CatBoost	94.4%

As shown in Table 6, the accuracy at the first level is better than the second level. LR and AdaBoost achieved an accuracy of 98.5%, while accuracy at the third level exceeded accuracy at the first and second levels where the merging of LR, LightGBM, and AdaBoost achieved 98.9% of accuracy.

Table 10. Multi-classifier merging best accuracies.

Merging level	Best Classifier	Accuracy
1st level	L.R. / AdaBoost	98.5%
2nd level	LR + LightGBM	97.5%
3rd level	LR + LightGBM + AdaBoost	98.9%

Table 11 compares the proposed method with the five recent related works using the same dataset but with different numbers of features.

In [12], the authors applied LightGBM with the LR models only after the imbalance problem had already been solved using the SMOTE Tomek method, where all features were used, and the feature selection techniques were not applied to choose the most effective features. Otherwise, this would have given better accuracy than 94.7%.

In [15], the authors applied it to the dataset without solving the imbalance problem, affecting the model's performance. On the other hand, if this problem had been solved before the application of the model, it would achieve higher accuracy than 83%.

In [18], the authors applied the SVM model directly to the dataset, affecting the model performance because of the imbalance problem. If this problem had been solved before the model was used, it would have achieved more than 89.5%.

In [19], the authors applied the random trees model to the dataset without solving the imbalance problem, which affected the model's performance; if this problem had been solved before using the model, it would have achieved higher accuracy than 91.47%.

In [20], the dataset was implemented using an Artificial Neural Network algorithm without solving the imbalance problem. Feature selection techniques have not been applied to select the most effective features, which decreased the model's performance since all the features had been utilized. If the model had solved these problems before applying, it would have reached higher than 93.35% accuracy.

Table 11. Comparison between the proposed method and the recent related work on the Z Ali Zadeh Dataset

Ref. No.	Method	features selected	Accuracy
[12]	LightGBM + LR + SMOTE Tomek	55	94.7%
[15]	Naïve Bayes	13	83.0%
[18]	SVM	15	89.5%
[19]	Random Trees	40	91.47%
[20]	Artificial Neural Network	55	93.35%
Proposed Method	LR + LightGBM + AdaBoost + SMOTE-ENN	26	98.9%

In the proposed method, the L.R. combined the AdaBoost and LightGBM models into the dataset after solving the imbalance problem using SMOTE-ENN and concurrently created new minority classes in SMOTE. Thereby, the instances of the dataset have been equally distributed to solve the problem of noise caused by SMOTE through the ENN method, allowing us to create powerful classification models for providing very accurate predictions for the occurrence of CAD. Chi-square and RFE methods had selected twenty-six features. As mentioned, the proposed method achieved the highest accuracy compared to previous studies, where it achieved an accuracy of 98.9%, as shown in Table 11.

5. Discussion of Results.

5.1. Impacts of Applying Classifiers on Imbalanced and Balanced Datasets.

LightGBM outperforms the other eight algorithms, shown by its highest accuracy, precision, recall, and F1 score when the models are applied to the dataset of imbalanced classes. CatBoost had the best recall. That has been shown in Table 6 and Figures 4, 5, 6, and 7.

SMOTE–ENN application results in much better outcomes. The results of Table 7 show that all classifier performances improved drastically. Particularly, the performance of LightGBM has been increased by 9.1%. The two classifiers that improved the most in performance were the L.R. accuracy score, which grew from 86.8% to 98.5%, and the AdaBoost accuracy score, which improved from 81% to 98.5%. These represent 11.7% and 17.5% increases, respectively. As expected, the recall and F1-scores for the models also improved with better results. L.R. and AdaBoost had the same outcomes, 96.7% and 98%, respectively.

But AdaBoost was better than L.R. at precision, which realized 95.5%, while L.R. realized 94%. Table 7, Figures 8, 9, 10, 11, and 12 explain these results. The results show that each classifier trained on the balanced dataset performs better. Excessive imbalance in class distribution is a worrying issue and is rather widespread in some fields, such as fraud detection, healthcare diagnostics, and risk management. What ML techniques do is try to reduce errors. Because of the much higher chances of cases in the majority class in imbalanced datasets, the models are more likely to label fresh observations to that class. Also, even though the penalty of a false negative in real life is normally far higher than that of a false positive, machine learning algorithms penalize both equally.

5.2. Performance of Multi-Classifiers on Balanced Dataset.

The performance of the classifiers in this research was compared before and after solving the data imbalance problem; it was found that the best performance was obtained after solving the problem with SMOTE–ENN. For this reason, the multi-classifier process was merged and applied after balancing the dataset by SMOTE–ENN.

5.2.1. *Comparison between Single Classifiers, Combination of Two and Three Classifiers Based on Accuracy.*

The high accuracy achieved from the combination of two classifiers is less than the single classifier, whereas, in a single-classifier experiment, the accuracy achieved by the L.R. and AdaBoost models was 98.5%. Still, in combining two classifiers, the best-achieved accuracy using LightGBM with L.R. was only 97.5%, which had fallen by 1.5% compared to a single classifier, as shown in Tables 7 and 8.

The accuracy achieved by combining three classifiers is better than that achieved by a single or two combined classifiers. In the experiment of a three-classifier, the achieved accuracy of L.R. combined with Adaboost and LightGBM was 98.9%, which increased by 0.4% compared to a single classifier and exceeded two combined classifiers by 1.4%, as shown in Table 9.

Based on the results, it was concluded that a single classifier performs better than merging two classifiers, as it consumes time without any noticeable performance improvement.

In comparison, the result was better in three combined classifiers than in the single and two combined classifiers, as shown in Table 10.0.4% may look like a small percentage, but for the medical field, especially when predicting a serious disease such as coronary artery disease, it saves the lives of many people.

6. **Conclusion.** In this work, CAD is, therefore, one of the most important widespread diseases in the world. Researchers, academics, and medical professionals worldwide are working towards rapid and high accuracy in CAD diagnosis.

Comparing the results by applying single classifiers before and after using SMOTE-ENN to balance the dataset showed that the best results of the classifiers were when the problem was solved with SMOTE-ENN. Before balancing, LightGBM performed better, with an accuracy of 87.9%. After balancing, L.R. performed well, followed closely by AdaBoost, achieving an accuracy of 98.5%. Therefore, the dataset should be balanced before applying any classifier to enhance the models' performance. Therefore, the multi-classifier process merging was applied after the dataset was balanced using SMOTE-ENN. In this way, features of less importance can be discarded by the chi-square test and RFE methods in feature selection, minimizing the complexity of the calculation and saving processing time.

The experimental results have shown that the multi-classifier fusion is much better than that of the single classifiers, where it has achieved an accuracy of 98.9%, far better than the second level of fusion, which reached an accuracy of 98.5%. The performance of classifiers after handling the problem of class imbalance using SMOTE-ENN is way better than that of applying classifiers without solving it for any dataset of coronary artery diseases, and better accuracies are being acquired. Apply more experiments with multi-classifiers in the future with different datasets to find the highest accuracy of the machine learning model and save more people's lives before they become CAD patients. The same models can be used to predict different diseases, and more intelligent methods will be used to predict CAD [28]. A background of varied resources will enhance performance in knowledge extraction and achieve a deep understanding of the problems associated with data collection and measurement. A real-time application for predicting CAD will be developed, and new features will be incrementally updated with the real-time application. Other feature selection techniques will provide the opportunity to select the best input features that will enhance the performance of the prediction system.

In future work, we need to use deep learning methods to predict the CAD with high accuracy and different machine learning algorithms. Apply multi-classifier experiments with various datasets and find the highest accuracy with machine learning models to save more people's lives before being a CAD patient.

REFERENCES

[1] Coronary Artery Disease. https://my.clevelandclinic.org/health/diseases/16898-coronary-artery-disease.

[2] Ayon, S. I., et al., "Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques." *IETE Journal of Research*, vol. 68, no. 4, 2020, pp. 2488-2507, https://doi.org/10.1080/03772063.2020.1713916.

[3] Ambrose, J. A.; Najafi, A., "Strategies for the prevention of coronary artery disease complications: Can we do better?" *Am. J. Med.*, vol. 131, 2018, pp. 1003–1009.

[4] Mohamed, M. H., et al., "Adaptive Learning Systems based on ILOs of Courses." *WSEAS Transactions on Systems and Control*, vol. 18, 2023, pp. 1–17, `https://doi.org/10.37394/23203.2022.1`.

[5] Muhammad, L. J., et al., "Machine learning predictive models for coronary artery disease." *S.N. Computer Science*, vol. 2, no. 5, 2021, p. 350.

[6] Thabtah, F., "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward." *Inform Health Soc Care*, vol. 44, no. 3, 2019, pp. 278–97.

[7] Muhammad, L. J., et al., "Predictive supervised machine learning models for diabetes mellitus." *Springer Nat Comput Sci*, vol. 1, 2020, p. 240.

[8] Mohamed, M., et al., "Sparsity and Cold Start Recommendation System Challenges Solved by Hybrid Feedback." *WSEAS Transactions*, vol. 13, 2019, pp. 2734-2741.

[9] Nusairat, N., et al., "Game-based student e-learning experience: Empirical evidence from private universities in Jordan." *International Journal of Data and Network Science*, vol. 8, 2024, pp. 1285-1292, https://doi.org/10.5267/j.ijdns.2023.11.006.

[10] Sayed, A., et al., "Predict student learning styles and suitable assessment methods using click stream." *Egyptian Informatics Journal*, vol. 26, 2024, https://doi.org/10.1016/j.eij.2024.100469.

[11] Debabrata, P., et al., "Fuzzy expert system approach for coronary artery disease screening using clinical parameters." *J Knowl Based Syst*, vol. 36, 2012, pp. 162–74.

[12] Zhang, S., et al., "Coronary Artery Disease Detection Model Based on Class Balancing Methods and LIGHTGBM Algorithm." *Electronics*, vol. 11, no. 9, 2022, p. 1495, https://doi.org/10.3390/electronics11091495.

[13] Ayon, S. I., et al., "Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques." *IETE Journal of Research*, vol. 68, no. 4, 2020, pp. 2488-2507, https://doi.org/10.1080/03772063.2020.1713916.

[14] Hassan, C. A., et al., "Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers." *Sensors*, vol. 22, no. 19, 2022, p. 7227, https://doi.org/10.3390/s22197227.

[15] Alotaibi, S. S., et al., "Automated prediction of coronary artery disease using Random Forest and Naïve Bayes." *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2020, pp. 109-114, https://doi.org/10.1109/ICACSIS51025.2020.9263159.

[16] Hossain, M. I., et al., "Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison." *Iran Journal of Computer Science*, vol. 6, no. 4, 2023, pp. 397-417.

[17] Trigka, M., and Dritsas, E., "Long-term coronary artery disease risk prediction with machine learning models." *Sensors*, vol. 23, no. 3, 2023, p. 1193.

[18] Dahal, K. R., et al., "Argumentative Comparative Analysis of Machine Learning on Coronary Artery Disease." *Open J. Stat.*, vol. 10, 2020, pp. 694–705.

[19] Joloudari, J. H., et al., "Coronary artery disease diagnosis; ranking the significant features using a random trees model." *Int. J. Environ. Res. Public Health*, vol. 17, 2020, p. 731.

[20] Dipto, I. C., et al., "Comparison of Different Machine Learning Algorithms for the Prediction of Coronary Artery Disease." *J. Data Anal. Inf. Process*, vol. 8, 2020, pp. 41–68.

[21] Alizadehsani, R., et al., "Extension of the Z-Alizadeh Sani dataset." *UCI Machine Learning Repository*, 2017, https://doi.org/10.24432/C5461K.

[22] Mohamed, M., et al., "Diabetic mellitus prediction with brfss data sets." *Journal of Theoretical and Applied Information Technology*, vol. 102, 2024, pp. 883-897.

[23] Jain, D., and Singh, V., "Feature selection and classification systems for chronic disease prediction: A review." *Egyptian Informatics Journal*, vol. 19, no. 3, 2018, pp. 179-89.

[24] Bikku, T., et al., "A contemporary feature selection and classification framework for imbalanced biomedical datasets." *Egyptian Informatics Journal*, vol. 19, no. 3, 2018, pp. 191-8.

[25] Mohamed, M., et al., "Music recommendation system used emotions to track and change negative users' mood." *Journal of Theoretical and Applied Information Technology*, vol. 99, 2021, pp. 4358-4376.

[26] Mohamed, M., et al., "Two recommendation system algorithms used SVD and association rule on implicit and explicit data sets." *International Journal of Scientific & Technology Research*, vol. 9, 2020, pp. 17-24.

[27] Dritsas, E., and Trigka, M., "Supervised Machine Learning Models for Liver Disease Risk Prediction." *Computers*, vol. 12, 2023, p. 19.

[28] Abdelhafeez, A., et al., "A novel approach toward skin cancer classification through fused deep features and neutrosophic environment." *Frontiers in Public Health*, vol. 11, 2023, https://doi.org/10.3389/fpubh.2023.1123581.