Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 4, pp. 3139–3151, DOI 10.12694/scpe.v25i4.2858

AN EXPLAINABLE AI MODEL IN HEART DISEASE CLASSIFICATION USING GREY WOLF OPTIMIZATION

VARUN G[*], JAGADEESHWARAN J[†], NITHISH K[‡], ASHICK SANJEY DS[§], VENKATESH V[¶], AND ASHOKKUMAR P[‖]

Abstract. Heart disease is one of the world's leading causes of death. It is estimated that around one-third of all deaths are caused by heart disease in the entire world. Recently many research works have focused on using machine learning models to detect and warn patients about the occurrence of heart disease at the early stage. However, machine learning models provide promising results, and the performance of the classification is affected by various reasons which include imbalanced training, and missing values. There are three main contributions of this research work. Firstly, missing values are addressed by employing a grouping of instances. Secondly, a dual filter based feature selection is introduced to pick the most effective features and lastly, we make of Grey Wolf Optimization for optimizing the hyperparameters of the machine learning models. Together, these contributions aim to improve the robustness and efficiency of machine learning applications by addressing missing data, optimizing feature selection, and fine-tuning model parameters. The accuracy of 98.41% indicates the superiority of the proposed classification which is more than 17.15% than the existing machine learning models. On the other hand, we use Explainable AI (XAI) methods to make our proposed model interpretable.

Key words: Heart Disease classification, Grey wolf optimization, Explainable AI, feature selection.

1. Introduction. Heart Disease is caused by abnormal disorders that affect the functioning of the heart and the bloodstream. Traditionally, doctors examine a patient with heart-oriented diseases by checking blood pressure, glucose level, BMI, and other attributes [1]. Several medical devices such as Electrocardiogram (ECG), X-rays, and CT scans are used to measure the above-mentioned attributes. The process of detecting heart disease is affected miserably due to a few reasons which include lack of expectations, wrong diagnosis, and delayed diagnosis. Due to these reasons, machine learning approaches are slowly adopted in the hospital domains. The information from machine learning models can be converted into live charts, and time series data, which can be used by the hospital to make timely decisions and assist the healthcare professionals in taking proper actions [2].

Classification is one of the categories of machine learning models that can detect patterns in hidden observations [3]. The World Health Organization has listed that around 17.1 million people die from heart disease every year which is the highest when compared with other diseases [4]. In recent years, many health professionals have been involved in creating massive heart disease data which can create new opportunities for machine learning classifiers to detect patterns and warn patients in their early stages.

Despite several observations of the success of the machine learning models in the healthcare industry, gaining optimal efficiency of the machine learning models is still a challenging task because of several reasons which include an imbalanced dataset, finding optimal hyperparameters, reducing the dimensions of the feature

[*]Department of Cybersecurity and Internet of Things, Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research
[†]Department of Cybersecurity and Internet of Things, Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research
[‡]Department of Cybersecurity and Internet of Things, Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research
[§]Department of Cybersecurity and Internet of Things, Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research
[¶]Department of Cybersecurity and Internet of Things, Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research
[‖]Department of Artificial Intelligence and Machine Learning, Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research

space, and so on. This research focuses on addressing the above-mentioned issues and contributing an optimal classifier that can be used for heart disease classification [5].

Normally, the medical-related datasets are imbalanced, however, this causes the machine learning model training very less efficient as the model learns too much about one class and very little about the other class (s) [6] [7]. The dataset used in this research paper is also an imbalanced dataset, where we make use of the Synthetic Minority Over-sampling TEchnique (SMOTE) to generate multiple synthetic instances of minority class to form a balanced dataset [8].

A machine learning model performance is largely dependent on its hyperparameters, for example, the Support Vector Machines (SVM) have the hyperparameters C, Kernel, Gamma, and so on [9]. These hyperparameter values have to be chosen very carefully. There are many methods such as Genetic algorithms [10] [11], Grid Search [12], and so on to find the best combination of hyperparameters. In this paper, Grey Wolf Optimization is used to pick the correct combinations of hyperparameters that produce the maximum performance.

Another important problem a machine learning classifier faces is the dimension of the data, If the number of dimension is very high, the classifier experience more difficulty in finding the pattern [13]. All the features are not required to be fed into the training process [14]. Many research works prove that selecting an optimal number of features will increase the performance of the classification. In this work, we adopt an average of ranks calculated by two feature selection methods.

The contributions of the paper are as follows:

1. Resolve the missing values by grouping instances.
2. Select the best features representing the target variable by using the average of ranks obtained by two popular filter-based feature selections, Correlation and Mutual Information.
3. Using Grey Wolf Optimization for finding the optimal hyperparameters for the machine learning models.

As machine learning models seem to be powerful in decision-making, trusting the result produced by the machine learning models is not fully practical. Hence, a new field called Explainable AI (XAI) has emerged which allows any user of the classifier to understand how the logic is created and how the classifier yields the specific output for the given input combinations. In other words, an XAI-based classifier enables humans to understand the behavior by exhibiting the explanations that increase the trust in the outputs. XAI-based models are capable of producing explanations that can say what is happening in an ML model.

The use of XAI is increasing especially in the field of medical domain. Due to the explainability, the trust in using the ML models increases and thus the decisions produced by the classifier are less doubtful.

As mentioned by [15], there are many reasons to use the XAI concepts in the machine learning based classification models which are listed below

- What is the problem in training a model and why the problem has occured.
- what input features are responsible in calculating the final class?
- How does the layers affect the end result?

Shapley Additive exPlanations (SHAP) is a popular XAI technique that makes use of the Shapley values to make the models more understandable. The concept of SHAP is used in multiplayer games where a group of players plays together in winning a game, In the end, the rewards are not distributed equally to all the players, but the rewards are distributed based on the contribution of each player in the game. This contribution is calculated by the SHAP algorithm. The same logic can be used in the classification, where a collection of features is involved in the classification task, at the end, a measure has to be made which calculates the weightage of each feature towards the end classification. The SHAP should satisfy multiple conditions [16] which are listed below.

- The weightage must be shared among all the features involved.
- If two features made the same contribution, the weight assigned to both features should be the same.
- the features which did not contribute towards the target class should get zero weightage.

One of the important challenges in machine learning classifiers is the process of picking the correct hyper parameters. The optimal combination gives the perfect performance. In many research works, nature-inspired models such as grey wolf optimization. This algorithm originated from the behavioral patterns noticed in the grey wolf packs. Grey wolf optimization offers a promising solution for hyperparameter tuning which helps the

models to perform at their best.

Grey wolves live in packs that comprise an alpha wolf, beta wolf, delta wolf, and omega wolf. This hierarchical structure is the backbone of the GWO. The first and the best solution in decision-making is the alpha wolf, the second best and the third best solutions in hunting and surrounding the prey are the beta and delta wolf respectively. The last one in the hierarchy is the omega wolf which follows the order of the other wolves.

Each machine learning classifier requires a specific combination of settings known as hyperparameters which drastically influence the model's performance. For example, in neural networks, the learning rate, and number of layers are the main hyper parameters that can be customized to increase the performance. However, the search for the best combination of hyperparameters is often challenging. The traditional methods like the grid search lack in high dimensional spaces. The GWO provides flexibility in this regard.

The goal of explainable AI (XAI) is to enable AI systems to provide human-understandable explanations for their choices. Consider it a form of AI transparency, which is crucial in deep learning since the models can appear to be these intricate black boxes. By giving us a glimpse inside these models, the methods of XAI help us understand how they make decisions. In addition to increasing public confidence in AI, this openness makes it easier to identify biases, comprehend how the model functions, and even enhance its output.

XAI is now a turning point for deep learning optimisation. It's similar to using a map to find your way around a maze of intricate neural networks. Through the use of interesting techniques such as saliency maps and attention mechanisms, XAI helps us understand which features of the information being provided are most relevant to the model's results. This knowledge enables us to fine-tune and enhance the network, increasing its effectiveness and lowering its error rate. Furthermore, by identifying any irregularities in the training data, XAI enables us to correct biases and mistakes and build deeper learning models that are more accurate and dependable.

This paper is organized into 5 sections. The next section overviews the different existing related research works about heart disease classification. The full explanation of the proposed approach is discussed in section 3. The evaluation of the results and the explainability is exhibited in section 4. Finally, the conclusion and future work are presented in section 5.

2. Literature Review. Three datasets called Cleveland, Hungarian, and CH are used in a work by [17]. The authors performed four types of experiments to compare their research work with the existing models. The first type involves the use of raw datasets and measures the performance of the classification. In the second type, the authors have used a popular filter-based feature selection known as chi-square to obtain a subset of relevant features. The principal component analysis (PCA) is implemented in the reduced dataset in the third type and finally, the same PCA is implemented in the original dataset. The Irrelevant features, duplicate features, features without proper explanation, and features with lots of missing values are skipped from consideration. The Random Forest (RF) Classifier got the maximum accuracy of 98.7% for the Cleveland dataset 99% for the Hungarian dataset and 99.4% for the CH dataset.

A research work done by [18] uses Decision Trees (DT) to classify heart disease. In their work, they extracted the IF THEN rules which can be used to improve the clinical diagnosis. The dataset used in the research work is an electronic health record-based dataset, it contains 1190 instances. The feature selection is also used to improve the performance of the classification. To resolve the issue of overfitting, pruning is done to reduce the complexity of the parameters which helps to minimize the classification errors. The authors reported an accuracy of 87%.

The ensemble model is used to achieve a promising accuracy in heart disease classification. A machine learning model identifies the patterns in the dataset to predict the target variable. Some classifiers yield satisfactory performance whereas some models predict with limited performance. Hence ensemble method [19] is used to improve the accuracy of the classifier. In this method, the weak classifier performance is increased by methods such as bagging, boosting, and so on. Feature selection is also applied to improve the accuracy of the prediction. The results show that the accuracy of the existing models was improved by 7%.

The Electrocardiography (ECG) data are used by [20] to identify the presence of heart disease. The authors have used many machine learning classifiers such as Naive Bayes, Random Forest, and Logistic Regression. The dataset used in the experiment has 302 instances. The accuracy of 92% was obtained by the authors.

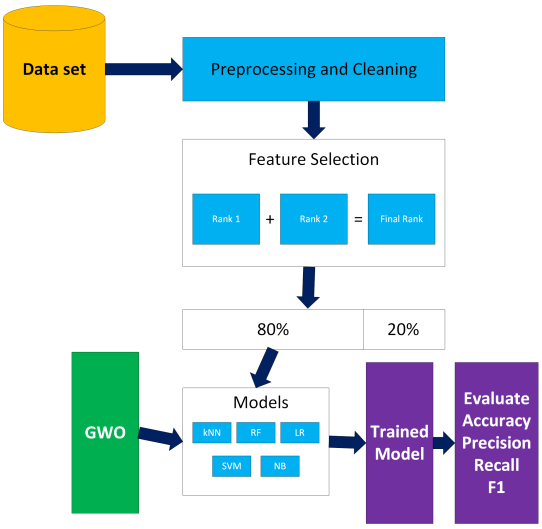

Fig. 3.1: The Methodology of the proposed classifier

The hybrid models are also used to detect the presence of heart disease like the one used in [21]. The irregular patterns are identified for the detection purpose. The paper tackles many of the common machine learning issues such as missing values, and imbalance datasets using proper preprocessing methods. Feature selection is done with the help of an extra classifier. The dataset that was used by the authors was Cleveland. The performance of the proposed model is compared and validated against multiple existing models and the proposed hybrid model produces an accuracy of 96.66%.

The imbalance problem is a major problem in the classifiers deployed in the healthcare industry. One of the solutions to solve the imbalance is undersampling, where the instances are dropped to create a balanced dataset. A work [22] mixes instance selection and under-sampling to improve the performance of the classification. Their proposed model produces an increased accuracy of 3.78% when compared with the existing methods.

A research work by [23] proposed a data mining technique called as Gini index decision tree where the the neural networks can increase the performance in the heart disease dataset. The authors have used the Cleveland dataset and found that the classifier accuracy is increased when the decision tree is used. The data is divided into supervised and unsupervised groups to equal the width, entropy, and frequency.

A hybrid method was projected by [24] which uses PSO and k means clustering for heart disease classification in Shimla, India. The clustering was made in the dataset for the construction of the model based on C4.5, neural networks, and linear regression. The number of the wrong classification and the accuracy were measured.

Different methods are used in the feature selection such as the incremental feature selection method which can increase the accuracy by reducing the false classification. The use of these methods has significantly increased the accuracy up to 93% [25].

Much research is involved in building a hybrid classification model that relies on multiple classifiers. A research work [26] uses a hybrid classifier and rough set concepts in heart disease classification for an accuracy of 92.55%. A twin SVM model has been used for effective decision-making in heart disease classification [27]. This model creates a dual non-parallel hyperplane to increase performance up to 86.75%.

3. The Proposed Heart Disease Classifier. In this section, we broadly explain the working of the proposed classifier. As mentioned in the section 1, we address some important problems such as missing values, balancing the dataset, feature selection, and hyperparameter tuning. Each of these is explained in the following subsections. Figure 3.1 exhibits the methodology of the proposed classifier models.

3.1. Missing Values resolving. A dataset is said to have a missing value when the data is absent for at least one feature in any of the instances. Missing values causes problems in the training stage and hence it is

resolved by various methods such as Deletion, Imputation, usage of ML models, and so on.

Imputation is one of the most trusted methods for handling missing values. In many cases, the statistical measure is calculated such as mean, median, min, and max, and assigned to the missing values. The problem with this method is, the actual distribution is disrupted. For example, they consider a dataset that has features called gender and height. Assume that height has many missing values, instead of taking the mean of height and assigning it to the missing values, we can group the dataset into two groups, one for males and another for females, and calculate the mean separately and assign the respective weight accordingly. In the heart disease dataset, we group the instances in the dataset according to gender and do the mean imputation respectively.

3.2. Balancing the dataset. The classifier learns all the patterns in the data during the training stage. If the number of instances for all the classes is equal, the classifier will have equal chances of learning all the patterns in all the classes equally. However, practically, it is not possible to obtain a perfectly balanced dataset. If the number of instances in any one class is higher than the other classes, then the classifier will learn more about the majority class and learn very little about the minority class.

The dataset that is used in the research work has 274456 instances in the negative class and 27261 instances in the positive class, which means the positive class has only 9.035% of the data. So to increase the instances of the minority class, we use the SMOTE technique and create a balanced dataset.

3.3. Feature Selection. The processed Heart Disease dataset contains 17 attributes, using all the 17 attributes will prevent the classifier from learning the complete pattern. Hence the correct subset of features should be selected from the original dataset so that the dimension of the input space gets reduced also the classifier easily learns about the relationship between input features and the target variable. In this paper, we use the two feature selection methods known as Correlation and Mutual Information which can pick the top 10 relevant features.

Correlation is a statistical measure that reveals the association between two variables, signifying their relationship from -1 to +1. A positive value suggest a direct relationship, while negative values indicate an inverse one. In feature selection, correlation aids in pinpointing redundant or closely linked features. Excessive correlation between features may cause problems, so opting for features with weaker correlation can potentially improve both the model's performance and simplicity.

A statistical metric called mutual information quantifies the degree of information that one variable tells about another, thereby establishing the relationship between the two. It measures how dependent one variable is on the others, regardless of the kind of link. When it comes to feature selection, mutual information evaluates each feature's relevance by calculating the degree to which one feature's information can be used to predict another. High mutual information features with the target variable are usually used for modelling since they are thought to be more informative.

$$r = \frac{\sum(X_i - \bar{X}_i)(Y_i - \bar{Y}_i)}{\sqrt{(X_i - \bar{X}_i)^2(Y_i - \bar{Y}_i)^2}} \tag{3.1}$$

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \tag{3.2}$$

In the process of selecting features, we first rank features based on the equation 3.1 and 3.2, then the average rank is found based on the two features as per the formula mentioned in 3.3.

$$Rank_i = \frac{Rank_{Corr_i} + Rank_{MI_i}}{2} \tag{3.3}$$

3.4. Hyperparameter Tuning. The parameters that can be changed are called a hyperparameter of a machine learning model. Choosing the best combination of hyperparameters can increase the performance of the classification. Grey Wolf Optimization (GWO) [28] is an algorithm that is inspired by how the gray wolf hunts its prey using the social hierarchy.

According to GWO, there are four levels of hierarchy Alpha, Beta, Delta, and Omega. The alpha wolf is the supreme commander of the wolf army. This wolf is responsible for making the most important decisions such as where to hunt. The second level in the hierarchy is the beta wolf where these wolves pass the decisions of the alpha wolf to other wolves and sometimes help the alpha wolf in making decisions. The next level of wolves is known as the delta wolf which takes care of the injured wolves and assists in the hunting process. The last level in the hierarchy is the omega wolf which eats the prey at the end. Omega wolves should be moved forward towards the prey and encircle it so that the prey becomes very easy to hunt. When a prey is found, the equations 3.4 and 3.5 are used to update the location of the wolf whenever a prey is found.

$$D = |\tilde{C}.\tilde{X}_p(t) - \tilde{X}(t)| \tag{3.4}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{\mathrm{A}}.\vec{D} \tag{3.5}$$

$$\vec{\mathrm{A}} = 2\vec{a}.\vec{r_1} - \vec{a} \tag{3.6}$$

$$\vec{C} = 2.\vec{r_2} \tag{3.7}$$

$$D_\alpha = |C_1 * X_\alpha - X| \tag{3.8}$$

$$D_\beta = |C_2 * X_\beta - X| \tag{3.9}$$

$$D_\delta = |C_3 * X_\delta - X| \tag{3.10}$$

$$X_1 = X_\alpha - A_1 D_\alpha \tag{3.11}$$

$$X_2 = X_\beta - A_2 D_\beta \tag{3.12}$$

$$X_3 = X_\delta - A_3 D_\delta \tag{3.13}$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{3.14}$$

In the GWO equations 3.6 and 3.7, t represents the current iteration, and r_1, and r_2 are random vectors between 0 to 1. The equations 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, and 3.14 are used to update the wolf positions until the global solutions are found. These equations are used to find the best hyperparameter based on the fitness function which calculates the accuracy at each combination of hyperparameter values.

The working of GWO is displayed in algorithm 1. The algorithm outlines how hyperparameter tuning is done using Grey Wolf Optimizer. The parameters of machine learning model is initialized and the fitness function is found for three best solutions: X_α, X_β, and X_ω. Until the maximum number of iterations is reached, each search agent position is updated on the defined equations. The values of a, A and C is also updated subsequently on the current search state. The fitness function is also gets reevaluated to obtain the three best solutions. The X_α gives the best possible combinations of hyperparameters at the end of all iterations. The algorithm allows to explore over the hyperparameter space efficiently and provide the best solution.

Algorithm 1 Hyperparameter tuning

procedure GWO(n)
    Initialize the value of a,A,c
    find the values of fitness function and obtain three best solutions.
    Let X_α be the first best solution.
    Let X_β be the second best solution.
    Let X_ω be the third best solution.
    for i=1 to maxIterations **do**
        for each Search Agent s **do**
            update the position of s based on 3.4 and 3.5.
        end for
        Update the values of a, A and C based on the present search state
        find the values of fitness function and obtain three best solutions for the oresent search state.
        Let X_α be the first best solution.
        Let X_β be the second best solution.
        Let X_ω be the third best solution.
    end for
    return X_α
end procedure

Table 4.1: Test bed details

Dataset	heart disease dataset downloaded at - [29]
Hardware Configuration	16GB RAM, i9 Processor, NVIDIA RTX 3060
Software Configuration	Windows 11, Python (with sklearn lib)
Experimental Design	80 (training)-20(testing) Split.
Evaluation Metrics	Four Standard metrics, accuracy, precision, recall and F1
Hyperparameters	Tuned using GWO (explained in Table 4.2)
Models	SVM, NB, kNN, LR and RF

Table 4.2: Hyper parameters used in the experiment

ML Model	Hyper Parameters
LR	penalty = 'l2'
kNN	n_neighbors= 3 weights= 'uniform'
SVM	C=0.366 kernel ='rbf'
RF	criterion= 'gini' max_depth= 5 min_samples_split= 3 n_estimators= 120
DT	criterion= 'entropy' max_depth= 5 min_samples_split= 3

4. Results and Discussion. To validate our proposed classifier, we use the heart disease dataset provided by from Behavioral Risk Factor Surveillance System (BRFSS) [29]. The processed dataset is available publically in kaggle [30]. We split the dataset into 80% for training and 20% for testing. The table 4.3 exhibits the feature information. The test bed details of the experiment is shown in table 4.1

Table 4.3: Feature Descriptions

Feature Name	Min	Max	Mean
Heart Disease	0	1	0.0856
BMI	12.02	94.85	28.325
Smoking	0	1	0.4125
Alcohol Drinking	0	1	0.0681
Stroke	0	1	0.0377
Physical Health	0	30	3.3717
Mental Health	0	30	3.8984
Diff Walking	0	30	3.8984
Sex	0	1	0.5247
Age Category	0	12	6.5145
Race	0	5	0.6851
Diabetic	0	3	0.42
Physical Activity	0	1	0.7754
Gen Health	0	1	1.405
Sleep Time	1	24	7.0971
Asthma	0	1	0.1341
Kidney Disease	0	1	0.1341
Skin Cancer	0	1	0.0932

Table 4.4: Accuracy values of the baseline and the proposed model

ML Model	Baseline	Proposed
SVM	0.84	0.98
NB	0.72	0.95
kNN	0.79	0.88
LR	0.91	0.78
RF	0.67	0.92

The machine learning models used to evaluate the proposed approach are listed below.

- Support Vector Machines (SVM).
- Naive Bayes.
- k Nearest Neighbours.
- Random Forest.
- Logistic Regression.

The hyperparameters for the machine learning models is shown in table 4.2.

The parameters used in the classification are explained in the upcoming subsections.

4.1. Accuracy. Accuracy is the most popular method to measure the correctness of the classification. It determines the correctly predicted samples among all the samples in the testing set. Accuracy is a metric used to measure the overall classification accuracy of the model on the entire dataset. The accuracy is calculated as mentioned in eq 4.1. The accuracy values are mentioned in the fig 4.1 and the table 4.4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

4.2. Precision. The percentage of correctly classified instances is known as precision. The formula for the precision is mentioned in eq 4.2. The table 4.5 and fig 4.2 displays the precision values of the baseline and

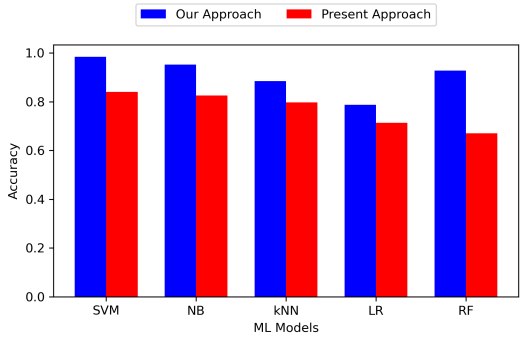

Fig. 4.1: Accuracy Comparision

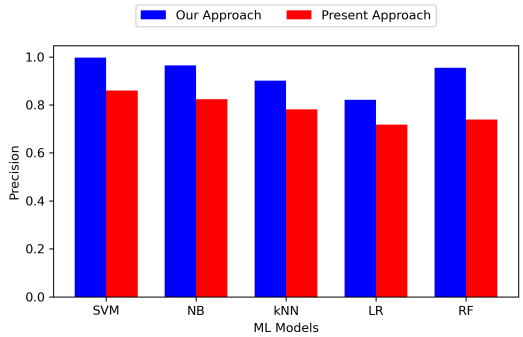

Fig. 4.2: Precision Comparision

Table 4.5: Precision values of the baseline and the proposed model

ML Model	Baseline	Proposed
SVM	0.85	0.99
NB	0.82	0.96
kNN	0.78	0.9
LR	0.71	0.82
RF	0.73	0.95

the proposed classifier.

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

4.3. Recall. Another metric to measure the performance of a classifier is recall. Out of all positive classified instances, how much did the classified classify as positive? This is also called as hit rate of the classifier. Equation 4.3 shows how the recall is calculated. The table 4.6 and fig 4.3 displays the recall values of the baseline and the proposed classifier.

$$Recall = \frac{TP}{TP + FP} \tag{4.3}$$

Table 4.6: Recall values of the baseline and the proposed model

ML Model	Baseline	Proposed
SVM	0.9	0.98
NB	0.91	0.96
kNN	0.9	0.92
LR	0.83	0.86
RF	0.76	0.94

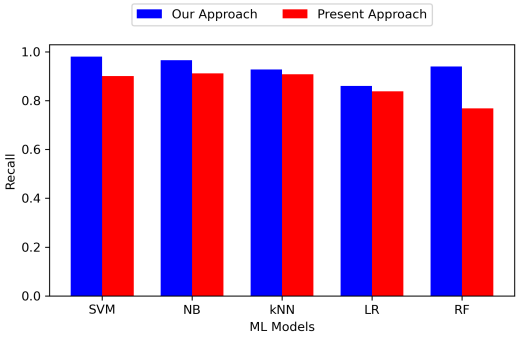

Fig. 4.3: Recall Comparision

Table 4.7: F1 scores of the baseline and the proposed model

ML Model	Baseline	Proposed
SVM	0.87	0.98
NB	0.86	0.96
kNN	0.83	0.91
LR	0.77	0.84
RF	0.75	0.94

4.4. F1 Score. F1 is the combination of Precision and Recall. The equation 4.4 displays the formula for the F1. The table 4.7 and fig 4.4 displays the F1 values of the baseline and the proposed classifier.

$$F1 = \frac{TP}{TP + FN} \tag{4.4}$$

The terms used in the equations are explained as follows

- True Positive (TP): If the classifier classifies a heart disease as heart disease.
- True Negative (TN): If the classifier classifies a normal patient as a normal patient.
- False Positive (FP): If the classifier classifies a normal patient as heart disease.
- False Negative (FN): If the classifier classifies a heart disease as a normal patient.

4.5. Model Explainable. Explainable Artificial Intelligence is one of the important research fields in recent years. An explanation is made so that the end user can understand how the machine learning classifications are made, i.e., it tries to convert the black box classification model into the interpretable model so that the decisions made by the classifiers are understandable by all.

Model explainations can be done in two ways, one is using the model as a whole and the other way is to explain only one instance. We demonstrate both the explainable methods in this section. Figure 4.5 exhibits

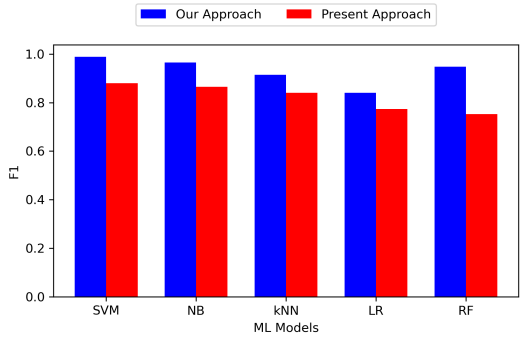

Fig. 4.4: F1 Comparision

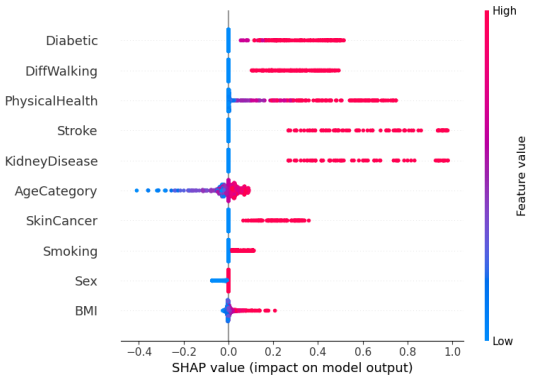

Fig. 4.5: SHAP Summary Values

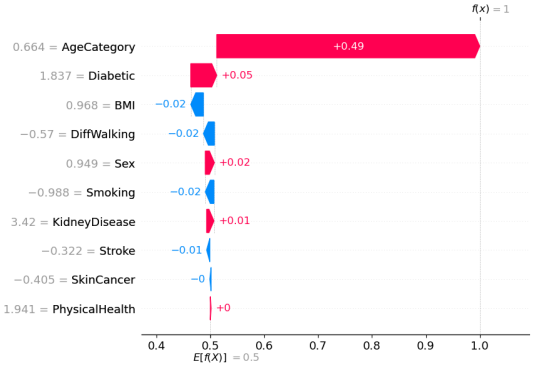

Fig. 4.6: SHAP values of RF

the summary of all features used in the classification. Figure 4.6, 4.7, 4.8 shows the SHAP values for RF, kNN, and NB respectively.

The random forest classifier is an ensemble model of multiple decision trees. The accuracy of the random forest using the proposed approach is 92.75% whereas without using our approach, the accuracy is just 67%.

The number of neighbors considered for kNN is 13. The accuracy of the kNN with and without using the

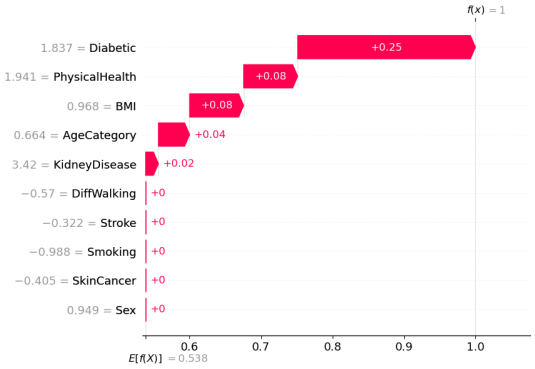

Fig. 4.7: SHAP values of kNN

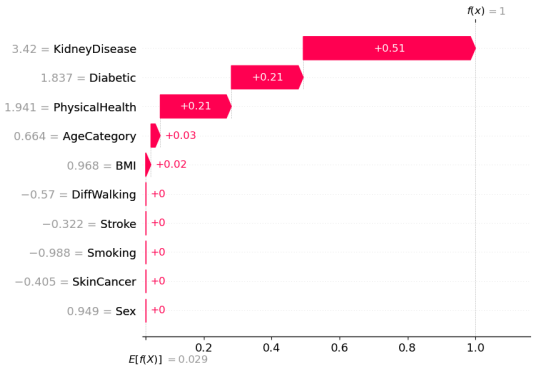

Fig. 4.8: SHAP values of NB

proposed approach is 88.5% and 79.68% respectively.

The NB is a probability-based classifier. NB produces the second-highest accuracy in our experiment next to SVM. The accuracy using our approach is 95.24%. When the baseline NB model was implemented in the same dataset, an accuracy value of 82.56% was observed.

5. Conclusion. In this research paper, we proposed a classification approach that makes use of average ranking of correlation and mutual information. Hence the complete features may not produce an optimal classification performance, we have selected 10 top features and performed the classification to detect whether a patient is having a heart disease or not. Along with the feature selection, we used SMOTE to balance the dataset and also GWO was used to pick the hyperparameters for the classification. The maximum accuracy of the proposed approach was produced by the support vector machine which is 98.41%. We have archived a 17.15% increase in accuracy.

In future work, we aim to develop a web app that can be used by all people, also we will implement the classification into other medical domain datasets.

REFERENCES

[1] Li, J.P.; Haq, A.U.; Din, S.U.; Khan, J.; Khan, A.; Saboor, A. Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. *IEEE Access* **2020**, *8*, 107562–107582.

[2] Tuppad, A.; Devi Patil, S. An Efficient Type 2 Diabetes Classification Framework Incorporating Feature Interactions. *Expert Systems with Applications* **2023**, p. 122138.

[3] Melillo, P.; De Luca, N.; Bracale, M.; Pecchia, L. Classification Tree for Risk Assessment in Patients Suffering From

Congestive Heart Failure via Long-Term Heart Rate Variability. *IEEE Journal of Biomedical and Health Informatics* **2013**, *17*, 727–733.
[4] Yuan, X.; Chen, J.; Zhang, K.; Wu, Y.; Yang, T. A Stable AI-Based Binary and Multiple Class Heart Disease Prediction Model for IoMT. *IEEE Transactions on Industrial Informatics* **2022**, *18*, 2032–2040.
[5] Palanivinayagam, A.; El-Bayeh, C.Z.; Damaševičius, R. Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review. *Algorithms* **2023**, *16*.
[6] Wang, Y.; Zhang, J.; Yan, W. An Enhanced Dynamic Ensemble Selection Classifier for Imbalance Classification With Application to China Corporation Bond Default Prediction. *IEEE Access* **2023**, *11*, 32082–32094.
[7] Paoletti, M.E.; Mogollon-Gutierrez, O.; Moreno-Álvarez, S.; Sancho, J.C.; Haut, J.M. A Comprehensive Survey of Imbalance Correction Techniques for Hyperspectral Data Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2023**, *16*, 5297–5314.
[8] Chen, C.; Shen, W.; Yang, C.; Fan, W.; Liu, X.; Li, Y. A New Safe-Level Enabled Borderline-SMOTE for Condition Recognition of Imbalanced Dataset. *IEEE Transactions on Instrumentation and Measurement* **2023**, *72*, 1–10.
[9] Bao, Y.; Yang, S. Two Novel SMOTE Methods for Solving Imbalanced Classification Problems. *IEEE Access* **2023**, *11*, 5816–5823.
[10] Zhou, R.; Liu, Y.; Zhang, K.; Yang, O. Genetic Algorithm-Based Challenging Scenarios Generation for Autonomous Vehicle Testing. *IEEE Journal of Radio Frequency Identification* **2022**, *6*, 928–933.
[11] Souza, M.G.; Vallejo, E.E.; Estrada, K. Detecting Clustered Independent Rare Variant Associations Using Genetic Algorithms. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2021**, *18*, 932–939.
[12] Lin, J.; Zhang, J. A Fast Parameters Selection Method of Support Vector Machine Based on Coarse Grid Search and Pattern Search. In Proceedings of the 2013 Fourth Global Congress on Intelligent Systems, 2013, pp. 77–81.
[13] Liu, X.; Tang, J. Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method. *IEEE Systems Journal* **2014**, *8*, 910–920.
[14] Ashour, A.S.; Nour, M.K.A.; Polat, K.; Guo, Y.; Alsaggaf, W.; El-Attar, A. A Novel Framework of Two Successive Feature Selection Levels Using Weight-Based Procedure for Voice-Loss Detection in Parkinson's Disease. *IEEE Access* **2020**, *8*, 76193–76203.
[15] Love, P.E.; Fang, W.; Matthews, J.; Porter, S.; Luo, H.; Ding, L. Explainable artificial intelligence (XAI): Precepts, models, and opportunities for research in construction. *Advanced Engineering Informatics* **2023**, *57*, 102024.
[16] Loh, H.W.; Ooi, C.P.; Seoni, S.; Barua, P.D.; Molinari, F.; Acharya, U.R. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine* **2022**, *226*, 107161.
[17] Gárate-Escamila, A.K.; Hajjam El Hassani, A.; Andrès, E. Classification models for heart disease prediction using feature selection and PCA. *Informatics in Medicine Unlocked* **2020**, *19*, 100330.
[18] Ozcan, M.; Peker, S. A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics* **2023**, *3*, 100130.
[19] Latha, C.B.C.; Jeeva, S.C. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked* **2019**, *16*, 100203.
[20] Matin Malakouti, S. Heart disease classification based on ECG using machine learning models. *Biomedical Signal Processing and Control* **2023**, *84*, 104796.
[21] Shrivastava, P.K.; Sharma, M.; sharma, P.; Kumar, A. HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms. *Measurement: Sensors* **2023**, *25*, 100657.
[22] Li, D.; Zheng, C.; Zhao, J.; Liu, Y. Diagnosis of heart failure from imbalance datasets using multi-level classification. *Biomedical Signal Processing and Control* **2023**, *81*, 104538.
[23] Pecchia, L.; Melillo, P.; Bracale, M. Remote Health Monitoring of Heart Failure With Data Mining via CART Method on HRV Features. *IEEE Transactions on Biomedical Engineering* **2011**, *58*, 800–804.
[24] Verma, L.; Srivastava, S.; Negi, P.C. A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. *Journal of Medical Systems* **2016**, *40*.
[25] Sandhiya, S.; Palani, U. An effective disease prediction system using incremental feature selection and temporal convolutional neural network. *Journal of Ambient Intelligence and Humanized Computing* **2020**, *11*, 5547–5560.
[26] Liu, X.; Wang, X.; Su, Q.; Zhang, M.; Zhu, Y.; Wang, Q.; Wang, Q. A hybrid classification system for heart disease diagnosis based on the RFRS method. *Computational and Mathematical Methods in Medicine* **2017**, *2017*, 1–11.
[27] Aditya, L.; Kumar, M. Heart disease prediction and classification usingmachine learning algorithms optimized byparticle swarm optimization and ant colonyoptimization. *International Journal for Modern Trends in Science and Technology* **2020**, *6*, 426–435.
[28] Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Advances in Engineering Software* **2014**, *69*, 46–61.
[29] CDC BRFSS Dataset, 2023.
[30] Pytlak, K. Indicators of heart disease (2022 update), 2023.

Edited by: Dhilip Kumar V
Special issue on: Unleashing the Power of Edge AI for Scalable Image and Video Processing
Received: Nov 15, 2023
Accepted: Mar 15, 2024