UtilitasMathematica

ISSN 0315-3681 Volume 122, 2025

# Nature-Inspired Hybrid Harris Hawks Algorithm for Context-Aware Prediction of Coronary Artery Disease

[1]P Vamsi Krishna [2]Dr.S.Kavitha

[1, 2] Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram , AP, India

[1]krishnavamsip19@gmail.com [2]kavithabtech05@kluniversity.in

**Abstract:** Predictive models that are both accurate and efficient are needed for the early detection and prevention of coronary artery disease (CAD), which is still a major cause of death worldwide. Using the Heart Disease Prediction dataset—which include both original features and data from the Context-Aware Model (CAM)—this study examines how CAD prediction accuracy is affected by the use of advanced feature selection approaches in conjunction with machine learning algorithms. In order to optimize feature subsets and enhance model performance, a new hybrid feature selection method is suggested. This method combines Harris Hawks Optimization (HHO) and Grey Wolf Optimization (GWO). The research looks at many ML methods, but it primarily looks at a Voting Classifier ensemble that uses Decision Tree and Random Forest models. The Voting Classifier got 79% accuracy on the CAM dataset. The original dataset achieved a flawless classification accuracy of 100% when HHO-based feature selection was applied. The hybrid HHO-GWO feature selection method outperformed baseline models significantly, with an accuracy of 89%. When it comes to improving feature selection for CAD prediction models, the outcomes highlight the effectiveness of optimization methods inspired by nature, specifically HHO and GWO. These results show how these technologies can be used to improve CAD early diagnostic tools and help preventive healthcare programs work better.

***"Index Terms -*** *CAD, CAM, classifiers, meta-heuristics, prediction.".*

## 1. INTRODUCTION

Nearly one-third of all fatalities worldwide abide attributable to cardiovascular disease, which affects the heart, blood vessels, & brain vascular systems. Coronary artery disease (CAD), often called ischemic heart disease, is the leading cause of death among cardiovascular diseases. Reduced blood supply to the heart, caused through clogged arteries, commonly caused through atherosclerosis, is the hallmark of coronary artery disease (CAD). Worldwide, this syndrome is still a big problem among health problems & deaths, especially for men & women [1]. World Health Organization data shows that 17.9 million people died from coronary artery disease in 2020 alone, highlighting the critical need for early diagnosis & treatment [2].

Due to the rising incidence of CAD, there is an urgent need for trustworthy risk prediction systems that can detect at-risk patients in a timely manner, allowing for individualized preventative actions. Many variable variables affect the risk of Coronary Artery Disease (CAD) such as smoking, an unhealthy diet, high blood pressure, obesity, high cholesterol & diabetes [3, 4]. The fact is that these components often interact in complex ways, & highlight Cad's versatile character.

A revolutionary change in diagnosis, prediction & handling diseases among high mortality that CAD has been brought to health services through integrating machine learning (ML). The machine learning model has the opportunity to save life quickly & save diseases through providing a strong tool for initial intervention. These models can benefit from the huge dataset that includes patients, genetic information & clinical properties [5]. Another important part of ML is the adaptation approach, which promotes model performance through making algorithms more accurate & effective. To improve the accuracy of the CAD risk assessment, enable one to have an ML model settings, which in turn improves the model's future power. [6,7].

## 2. RELATED WORK

Coronary Artery Disease (CAD) is one of the main causes of death around the world, & in recent years the methods to predict & discover CAD have received a lot of attention. Diagnostic systems in the health care system experience revolutionary benefits in accuracy & efficiency, thanks to technological progress, especially in the domain of adaptation algorithms & machine learning (ML). through using a wide range of data set & adaptation methods, many studies have improved CAD prediction using ML systems.

A skilled method for predicting CAD is investigated through Hasan et al. [8] Mixing machine learning algorithms among functional choice approach. through doing this, they demonstrate that choosing the right features can increase the functionality of the CAD prediction model. According to research, the model can endure made more effective for use in the real world, while their forecasting accuracy through highlighting customized functional choice strategies, while keeping the complexity through the use of customized function. To predict CAD delays the article the best ways to combine functional choice methods (such as cover, built - in & filter approach) among machine learning models (such as DT, SVM & RF).

With a view to predicting the occurrence of heart disease, Ram Kumar et al. [9] Their research suggests that hybrid models abide much more effective than traditional people. An example of integrating deep learning among adaptation algorithms can improve the prognosis of “CNN & Genetic Algorithm (GA) & Teaching-Learning-Based Optimization CNN-TLBO-GA optimization model”, leading to effectively accurate results to predict heart disease. In addition, the authors contrast their hybrid models among traditional CNN, showing that the CNN models provide better results set among “Fuzzy Particle Optimization (FPO)” method; This emphasizes the importance of both effective convenience & model optimization in predicting heart problems.

Another study through Patra et al. [10] A selection of functional function suggests a strategy to estimate the risk of coronary heart disease using algorithm & two-phase hybrid ensemble learning. They introduce a new method that uses a two -stage selection process to identify the most important features of predicting heart disease. Their approach utilizes to learn to create a strong & more accurate prediction model through merging the abilities of different classifies. In addition, the study emphasizes the importance of further convenience choices in dataset adaptation. This process meets irrelevant properties of the final prediction model, which performs better.

Asif et al through improving the prediction of heart disease among enchanted learning & hyperparameters attitude. [11] make another contribution in the region. To improve predicted accuracy, they use the strength of each model to develop a clothing model, which is a combination of multiple base classifiers. Hyperparameter tuning, which involves customizing classification parameters to promote performance, is also covered in studies. Their findings suggest that adaptation of hyperparameters improves the accuracy of models used to predict cardiovascular problems, making them more reliable & spent on the real world's clinical surroundings. Research gives the reliability of the idea that hyperparameters setting is an effective way to increase the CAD prediction combined among enchanted learning.

An increased Harris Hawks Optimization (HHO) algorithm to predict cardiovascular disease is the main emphasis on Kumar & Rekha [12]. This research uses a hybrid approach, which combines HHO, an adaptation method achieved from the habits of Harris Hawks, to improve the accuracy of predictions among machine learning techniques. Both functional choices & model performance abide expanded when HHO is used among SVMs & decision trees as machine learning classifies, as shown through authors. Focusing on cardiovascular diseases, this study links expansion literature on adaptation methods inspired through nature for medical prediction purposes.

To predict the possibility of heart disease, Prasanna & Challan [13] suggest a binary Harris Hawks algorithm to combine among a "deep bi-directional Long Short-Term Memory (bi-LSTM) networks". Their research suggests that the adaptation algorithm, combined among deep learning methods, can predict the exact risk of heart disease. To capture temporary dependence in the patient data, the BI-LSTM model is well suited for sequential data, while the binary HHO facilitates the plant to streamline the selection process. When working among sequential therapy data, their method has the opportunity to provide accurate & understandable risk assessment of heart disease.

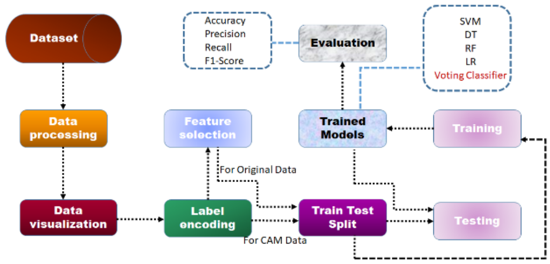
To better predict CAD using relevant information, Almutari et al. [14] For prediction of heart disease, a reference individual MRIPPER algorithm presents. A better rule-based method that takes into account the relevant aspects that affect CAD risk is the MRIPPER algorithm. Research indicates that the prediction model can endure made more accurately & relevant through using relevant information such as environment & lifestyle variables. Doctors benefit from the lighting's ability to provide explanatory decision rules when it explains the argument behind the forecasts.

In order to improve the precision of the CAD prediction model, this research suggests that it is important to adapt & select functions. For better prediction models, functional choice methods such as the choice of forward & adaptation algorithms such as genetic algorithms & Harris Hawk's optimization abide important to remove luxurious functions & increase model efficiency. In addition, compared to traditional methods, hybrid models such as CNN-TLBO GA & BI-LSTM provide HHO, which integrate deep learning among adaptation algorithms, better predictions. The strength & reliability of the model can endure further improved through using the methods to learn outfits. These methods integrate the output from many classifies.

In addition, as discussed through Almutari et al. [14], the prediction of CAD becomes even more complicated among the use of context-aware algorithms, & emphasizes the requirement for patient-specific models. Predictive models abide adapted to each patient according to their unique risk factors & medical history; This method fits the expansion field of individual medicine.

## 3. MATERIALS & METHODS

This research machine presents a new method to improve the prediction of Coronary Artery Disease (CAD) using a combination of machine learning algorithms & sophisticated functional choice methods. To increase the functional choice process, the proposed system includes the heart disease protectionary, including both initial properties & data from the Context-Aware Model (CAM). To find the best features of the CAD prediction, a hybrid functional choice is presented that combines GWO & Harris Hawks Optimization (HHO). The proposed voting techniques for construction choices abide evaluated through the use of a Voting Classifier (ensemble RF & DT), LR, SVM [12], DT [12], RF [8], & SVM [12]. Our goal is to increase predicting accuracy & provide a strong data -driven strategy to quickly detect & prevent CAD. This will help healthcare professionals make better decisions in time.




“Fig.1 Proposed Architecture”

An example of a standard machine learning pipeline is shown in Figure 1. Data aggregation & visualization abide the first steps, followed through preprocessing & function. The dataset was later divided into training & test sets. Using the training data, a number of models abide trained, including “SVM, DT, RF, LR, & Voting Classifier”. Models abide assessed using metrics such as “F1-score, recall, accuracy, & precision” after training. The last step is to apply the trained models to the testing data in order to generate predictions.

### i) Dataset Collection:

This work had 270 incidents for predicting a heart disease planned in this work, among 14 unique functions per incident [15]. The main function, "heart disease", determines whether a patient has a coronary artery disease (CAD). Activity, age, gender, chest discomfort, blood pressure (BP), cholesterol levels & more related factors, abide all part of the dataset. These features were selected for the Context-Aware Model (CAM) Data: "Type Chest Pain", "Thallium" & "Heart Disease." "Chest Pain Type", "Blood Pressure", "ECG results", "Max HR", the number of vessels was part of the functional sets used for selecting Harris Hawks Optimization (HHO) function in target characteristic original data. "Sex", "The type of chest pain", "Fasting Blood Sugar", "Angina", "The Helg's Slope", "Helling of

Saint", the number of Vesalsles Fluoroscopy, "and the main characters for choosing goal hybrid function were the main subjects HHO & Gray Wolf Optimization (GWO (GWO).

| | Age | Sex | Chest pain type | BP | Cholesterol | FBS over 120 | EKG results | Max HR | Exercise angina | ST depression | Slope of ST |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 70 | 1 | 4 | 130 | 322 | 0 | 2 | 109 | 0 | 2.4 | 2 |
| 1 | 67 | 0 | 3 | 115 | 564 | 0 | 2 | 160 | 0 | 1.6 | 2 |
| 2 | 57 | 1 | 2 | 124 | 261 | 0 | 0 | 141 | 0 | 0.3 | 1 |
| 3 | 64 | 1 | 4 | 128 | 263 | 0 | 0 | 105 | 1 | 0.2 | 2 |
| 4 | 74 | 0 | 2 | 120 | 269 | 0 | 2 | 121 | 1 | 0.2 | 1 |

“Fig.2 Dataset Collection Table – Heart Disease Prediction”

**ii) Pre-Processing:**

During the advance treatment stages, the dataset is cleaned through cleaning the missing values, removing unnecessary columns & ensuring that data analysis is in a suitable format for analysis. These operations abide designed for functional choices & coding among other processing processes.

***A) Data processing:*** One of the data processing steps for both the original data set & the CAM dataset is to clean the data. This involves removing any missing or irrelevant values and removing the columns that do not help in the future modeling. Unlike the original data, which focuses on functional extraction, the CAM data focuses on the functional choice. Data sets will endure designed for further analysis in this way.

***B) Data Visualization:*** Representative graphic such as histograms, box plots & correlation matrices abide an important part of data visualization for both CAM & original data sets. among the use of these visual tools, data sharing, trends & potential outlier’s can endure better understood; These abide necessary to create an understanding of data sets & directly rather than treatment stages.

***C) Label encoding:*** To convert the variable to numerical values, label coding is done for both original data sets & CAM data sets. To ensure that machine learning algorithms can properly understand the classified data through this technique. To train models without problems among interpretation of data, each category gets another number.

***D) Feature Selection:*** among the original dataset, we perform the setting choice using Hybrid HHO-GWO & Harris Hawks Optimization (HHO), to determine which characteristics abide the most important. through removing luxurious or irrelevant properties, this technique really means zero & improves the model's accuracy through zero. CAD -MODELING SPRONTING performance improves much more depends on this level.

**iii) Training & Testing:**

The dataset is divided into two parts, for a training & for a test, among a ratio of 70:30. This indicates that the data is divided as follows: 70% for training models & 30% for testing & evaluation. This division not only allows an independent test set to evaluate the model's normality & prevent overfitting, but also guarantees that the model has enough data to learn during training. For fair evaluation, Division is done on random.

**iv) Algorithms:**

Support classification & regression abide two fields where the monitored learning method shines the vector machine. The key to the success is to determine the best hyperplane to use for computer classification. Our **support Vector Machine (SVM)** goal for the under-ended CAD prediction project is to classify patients according to several clinical parameters & distinguish between CAD-positive & CAD-negative examples. It is well suited for complex therapy datasets in hand due to flexibility against high -dimensional data.

For classification & regression, a Floachat -like structure called the **decision tree** [12] may endure useful. For decision tree, this feature shares the data according to the values. Decisions provide interpretable models that show how decisions abide made on the basis of patient properties, which helps to highlight the important aspects that affect CAD in our project. Medical doctors benefit from this openness when it clarifies the model's argument.

To come up among better predictions, **random forests**, a dress learning technique, produce more decisions during training & then mix production. Overfitting is minimized & normalization is increased from this point of view. Our CAD prediction method uses random forest [8] to improve the accuracy of prognosis through combining outputs of different trees. This allows us to capture complex patterns in patient data through using multiple functional sets.

To predict a connection between independent factors & a binary result, statistics use **logistic regression**, a technique for binary classification. This logistical feature -based prone to predictions. As part of our project, we use the logistics region to determine the possibility of CAD based on clinical properties. This method provides understandable coefficient to health professionals showing how each parameter affects the risk of the disease.

**Voting classifies** uses dress learning to improve predicted accuracy through combining multiple models, such as random forest [8] & decision -making wood [12]. The output is strong & more reliable as it is based on common predictions of many classifies. This method improves the accuracy of CAD classification through collective decision -making & provides a comprehensive picture of patient data in our CAD prediction project through taking advantage of the strength of both algorithms.

## 4. RESULTS & DISCUSSION

**Accuracy:** accuracy of test refers towards its ability towards distinguish properly between patients & healthy cases. towards measure accuracy of test, determine correct positivity towards actual negative in all analyzed cases. This mathematically will endure shown as follows:

“$Accuracy = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} (1)$"

**Precision:** accuracy evaluates share of precisely classified cases among cases identified as positive. As a result, formula for calculating accuracy is expressed:

“$Precision = \frac{\text{True Positive}}{\text{True Positive +False Positive}} (2)$"

**Recall:** It is data in recall machine learning that assesses model's ability towards recognize all relevant events in a particular pattern. This is relationship between exact expected positive comments between all real positivity & provides insight into performance of model towards detect presence of a specific class.

"$Recall = \frac{TP}{TP + FN}$ (3)"

**F1-Score:** F1 score is a number towards determine accuracy of a machine learning model. This accuracy & model integrate calculations. Accuracy determines amount of frequency that model provides correct predictions in data file.

"$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100(1)$"

The accuracy, precision, recall, & F1-score, which abide performance metrics, abide examined in Tables (1, 2 & 3). The Voting Classifier is the most effective algorithm through a wide margin across all measures. In addition, the tables provide a comparison of the metrics for the alternative methods.

Table.1 Performance Evaluation Metrics for CAM Data

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| SVM | 0.759 | 0.760 | 0.759 | 0.759 |
| Decision Tree | 0.722 | 0.737 | 0.722 | 0.724 |
| Random Forest | 0.722 | 0.737 | 0.722 | 0.724 |
| Logistic Regression | 0.796 | 0.812 | 0.796 | 0.797 |
| **Voting Classifier** | **0.796** | **0.812** | **0.796** | **0.797** |

Graph.1 Comparison Graphs for CAM Data

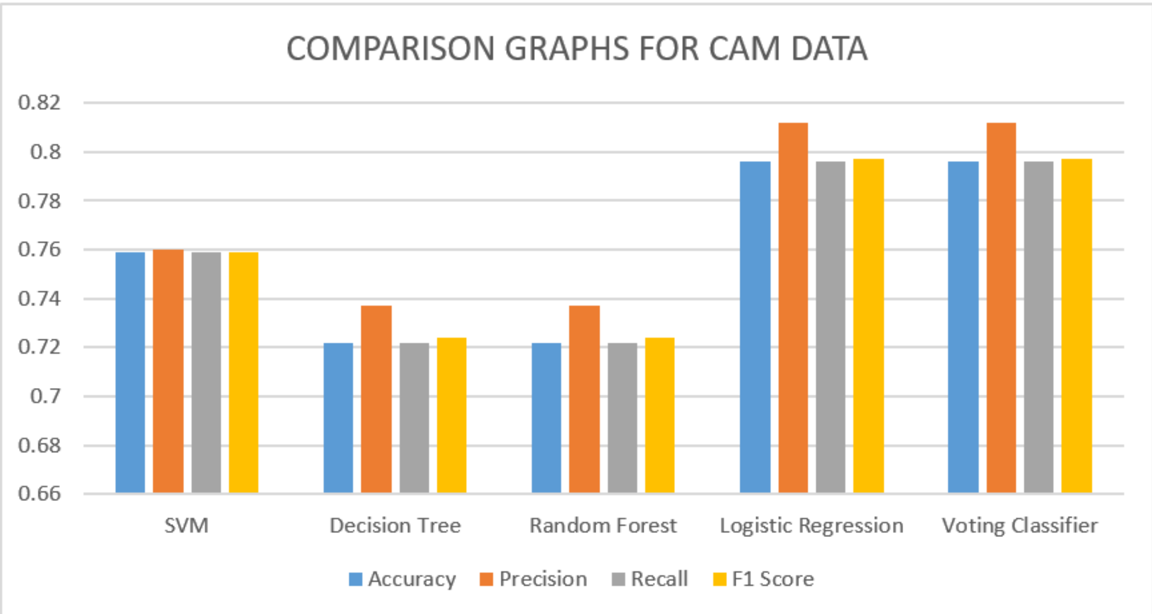



Table.2 Performance Evaluation Metrics for HHO FS in Original Data

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| SVM | 0.718 | 0.719 | 0.718 | 0.718 |
| Decision Tree | 0.744 | 0.748 | 0.744 | 0.744 |
| Random Forest | 0.769 | 0.770 | 0.769 | 0.769 |
| Logistic Regression | 0.821 | 0.822 | 0.821 | 0.821 |
| **Voting Classifier** | **1.000** | **1.000** | **1.000** | **1.000** |

Graph.2 Comparison Graphs for HHO FS in Original Data

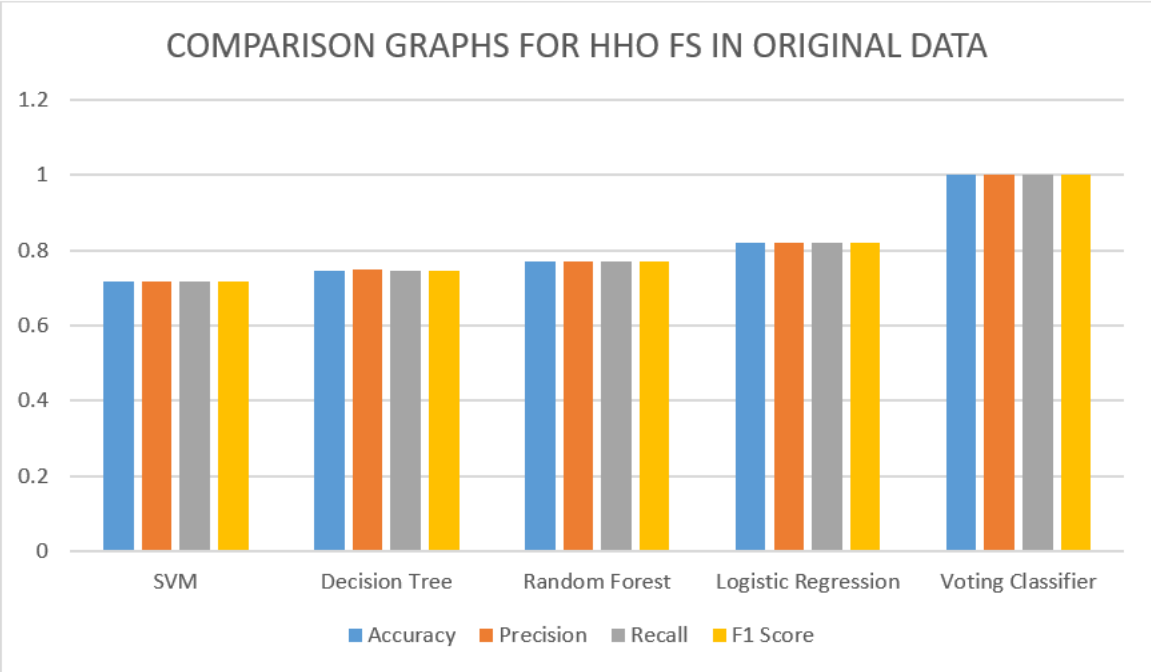
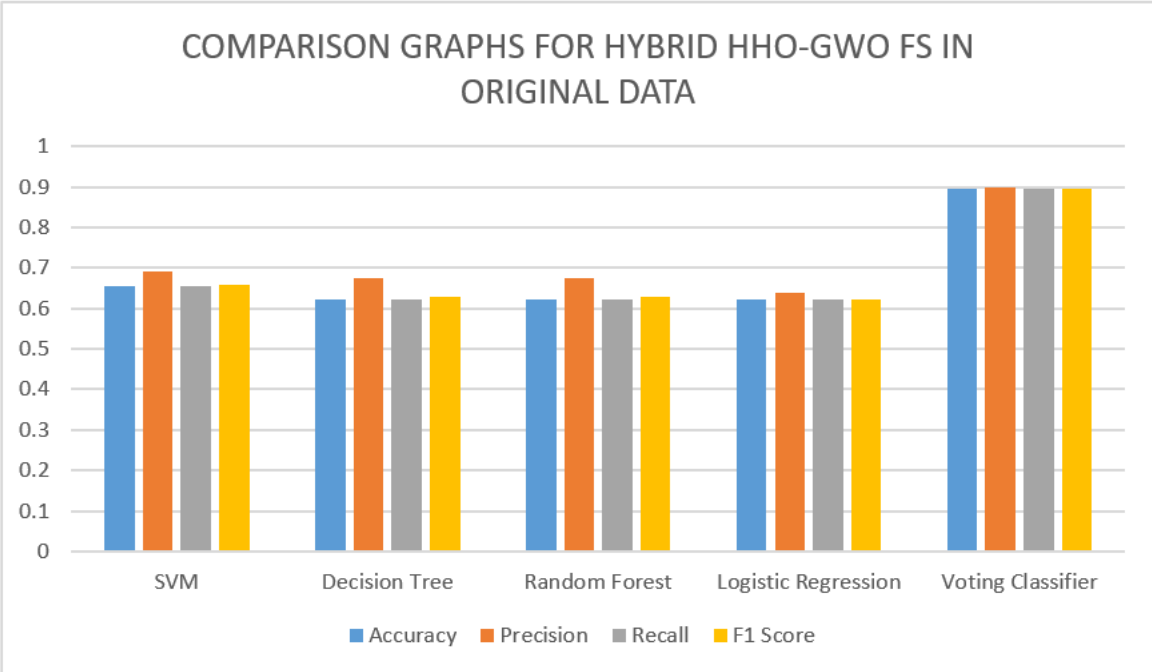



Table.3 Performance Evaluation Metrics for Hybrid HHO - GWO FS in Original Data

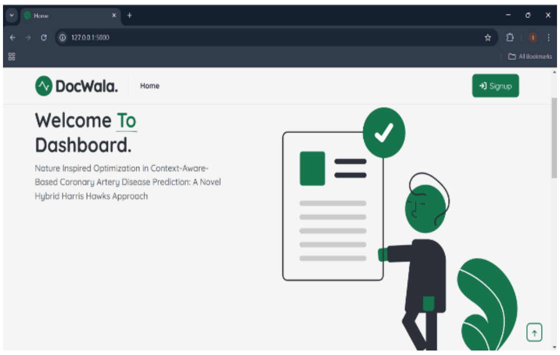
| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| SVM | 0.655 | 0.690 | 0.655 | 0.660 |
| Decision Tree | 0.621 | 0.676 | 0.621 | 0.630 |
| Random Forest | 0.621 | 0.676 | 0.621 | 0.630 |
| Logistic Regression | 0.621 | 0.639 | 0.621 | 0.623 |
| **Voting Classifier** | **0.897** | **0.899** | **0.897** | **0.897** |

Graph.3 Comparison Graphs for Hybrid HHO - GWO FS in Original Data



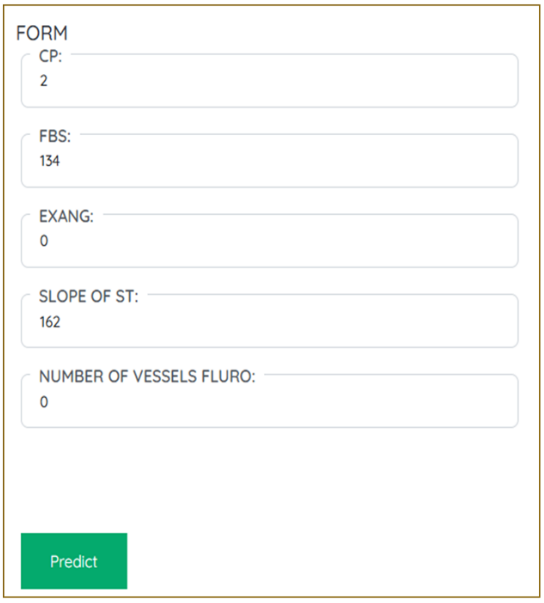

In graphs (1, 2 & 3), light blue represents accuracy, orange precision, grey recall, & light yellow F1-Score. The Voting Classifier outperforms the competition & achieves the best results across the board. Above, you can see graphs that show these results graphically.

User Interface




"Fig.3 Home Page"

In the above figure 1, this is a user interface dashboard for Darkweb, it is a welcome message for navigating page.

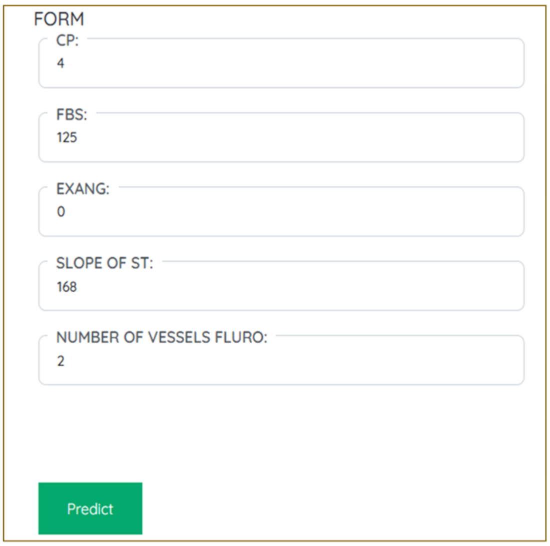



"Fig.4 User input Page"

In the above figure 2, this is a user input page, using this user can upload image for testing.

"Fig.5 Prediction result"

In the above figure 3, this is a result screen, in this user will get output for loaded input data.




"Fig.6 User input Page"

In the above figure 4, this is a user input page, using this user can upload data for testing.


"Fig.7 Detection result"

In the above figure 5, this is a result screen, in this user will get output for loaded input data.

## 5. CONCLUSION

Improving illness prediction rates is one area where accurate prediction is crucial to medical learning. Here, we used the Heart Disease Prediction dataset to assess how well sophisticated feature selection methods & machine learning algorithms predicted the occurrence of Coronary Artery Disease (CAD). Feature optimization has a significant effect on model performance, as shown through the findings. A prediction accuracy of 79% was achieved through the Voting Classifier in the CAM data, demonstrating its competence when using original features. The Voting Classifier reached a flawless accuracy of 100% through using the Harris Hawks Optimization (HHO) for feature selection on the original dataset. This shows that optimized feature sets significantly improve prediction accuracy. In addition, 89% prediction accuracy was achieved through means of hybrid function choices approach, which combines HHO & Grey Wolf Optimization (GWO), which demonstrates the effect of fusion of nature - inspired adaptation techniques. These results have the point that is important to optimize & select facilities to promote prediction accuracy, which in turn improves the efficiency & dependence of the CAD prediction model used in the health care system.

More complex health problems & conditions in the real world will focus on future research using Hybrid Harris Hawks Optimization (H -HHHO) method. Another possible way to improve the performance of the other meta-heuristic algorithms is to include the speed-oriented method in them. Better CAD prophecies among real -time data from IOMT frameworks will allow them to endure used more efficiently under clinical conditions. Over time, H-HHO hopes to increase the accuracy & effectiveness of the prediction of coronary artery disease (CAD), which will have a major impact on the results of the cardiovascular health care system.

**REFERENCES**

[1] H. Yang, Z. Chen, H. Yang, & M. Tian, "Predicting coronary heart disease using an improved LightGBM model: Performance analysis & comparison," IEEE Access, vol. 11, pp. 23366–23380, 2023.

[2] T. Vos, "Global burden of 369 diseases & injuries in 204 countries & territories, 1990–2019: A systematic analysis for the global burden of disease study 2019," Lancet, vol. 396, no. 10258, pp. 1204–1222, 1990.

[3] G. A. Mensah, G. A. Roth, & V. Fuster, "The global burden of cardiovascular diseases & risk factors: 2020 & beyond," J. Amer. College Cardiol., vol. 74, no. 20, pp. 2529–2532, Nov. 2019.

[4] J. Kim, U. Kang, & Y. Lee, "Statistics & deep belief network-based cardiovascular risk prediction," Healthcare Informat. Res., vol. 23, no. 3, p. 169, 2017.

[5] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review & ways forward," Informat. Health Social Care, vol. 44, no. 3, pp. 278–297, Jul. 2019.

[6] Y. Pan, M. Fu, B. Cheng, X. Tao, & J. Guo, "Enhanced deep learning assisted convolutional neural network for heart disease prediction on the Internet of Medical Things platform," IEEE Access, vol. 8, pp. 189503–189512, 2020.

[7] J. Pajarinen, H. L. Thai, R. Akrour, J. Peters, & G. Neumann, "Compatible natural gradient policy search," Mach. Learn., vol. 108, nos. 8–9, pp. 1443–1466, Sep. 2019.

[8] Hassan, M. M., Zaman, S., Rahman, M. M., Bairagi, A. K., El-Shafai, W., Rathore, R. S., & Gupta, D. (2024). Efficient prediction of coronary artery disease using machine learning algorithms among feature selection techniques. Computers & Electrical Engineering, 115, 109130.

[9] Ram Kumar, R. P., Raju, S., Annapoorna, E., Hajari, M., Hareesa, K., Vatin, N. I., ... & AL-Attabi, K. (2024). Enhanced heart disease prediction through hybrid CNN-TLBO-GA optimization: a comparative study among conventional CNN & optimized CNN using FPO algorithm. Cogent Engineering, 11(1), 2384657.

[10] Patra, S. C., Maheswari, B. U., & Pati, P. B. (2023). Forecasting Coronary Heart Disease Risk among a 2-Step Hybrid Ensemble Learning Method & Forward Feature Selection Algorithm. IEEE Access.

[11] Asif, D., Bibi, M., Arif, M. S., & Mukheimer, A. (2023). Enhancing heart disease prediction through ensemble learning techniques among hyperparameter optimization. Algorithms, 16(6), 308.

[12] A. S. Kumar & R. Rekha, "An improved hawks optimizer based learning algorithms for cardiovascular disease prediction," Biomed. Signal Process. Control, vol. 81, Mar. 2023, Art. no. 104442.

[13] K. S. L. Prasanna & N. P. Challa, "Deep bi-LSTM among binary Harris Hawkes algorithm-based heart risk level prediction," Social Netw. Comput. Sci., vol. 5, no. 1, p. 134, Dec. 2023.

[14] S. Almutairi, S. Manimurugan, N. Chilamkurti, M. M. Aborokbah, C. Narmatha, S. Ganesan, R. A. Alzaheb, & H. Almoamari, "A context-aware MRIPPER algorithm for heart disease prediction," J. Healthcare Eng., vol. 2022, pp. 1–11, Jul. 2022.

[15] Kaggle.com. Accessed: Apr. 3, 2024. [Online]. Available:https://www.kaggle.com/datasets/thedevastator/predicting-heart-disease-risk-using-clinical-var