

Predicting Arrhythmia Based on Machine Learning Using Improved Harris Hawk Algorithm

Nitesh Sureja¹, Nandini Chaudhari², Rocky Upaghyay³, Shivam Upadhyay⁴, Safeya Dharmajwala⁵

^{1,4,5}Department of Computer Science & Engineering
Krishna School of Emerging Technology & Applied Research
Vadodara, INDIA

^{2,3}Department of Information Technology
Krishna School of Emerging Technology & Applied Research
Vadodara, INDIA

*nmsureja@gmail.com (only corresponding Author)

Abstract—Arrhythmia disease is widely recognized as a prominent and lethal ailment on a global scale, resulting in a significant number of fatalities annually. The timely identification of this ailment is crucial for preserving individuals' lives. Machine Learning (ML), a branch of artificial intelligence (AI), has emerged as a highly efficient and cost-effective method for illness detection. The objective of this work is to develop a machine learning (ML) model capable of accurately predicting heart illness by using the Arrhythmia disease dataset, with the purpose of achieving optimal performance. The performance of the model is greatly influenced by the selection of the machine learning method and the features in the dataset for training purposes. In order to mitigate the issue of overfitting caused by the high dimensionality of the features in the Arrhythmia dataset, a reduction of the dataset to a lower dimensional subspace was performed via the improved Harris hawk optimization algorithm (iHHO). The Harris hawk algorithm exhibits a rapid convergence rate and possesses a notable degree of adaptability in its ability to identify optimal characteristics. The performance of the models created with the feature-selected dataset using various machine learning techniques was evaluated and compared. In this work, total seven classifiers like SVM, GB, GNB, RF, LR, DT, and KNN are used to classify the data produced by the iHHO algorithm. The results clearly show the improvement of 3%, 4%, 4%, 9%, 8%, 3%, and 9% with the classifiers KNN, RF, GB, SVM, LR, DT, and GNB respectively.

Keywords—Harris Hawk Algorithm; Swarm Intelligence; Machine Learning; Levy Flight; Arrhythmia.

I. INTRODUCTION

All The simplicity, adaptability, and gradient-free mechanisms of swarm intelligence algorithms (SIAs) have drawn academics' attention in recent years [1]. With the aid of a fitness function, SIAs attempt to provide efficient and ideal solutions by eliminating ineffective ones.

The following algorithms are commonly used in solving optimization problems: Dragonfly Algorithm (DFA), Whale Optimization Algorithm (WOA), Water Wave Optimization (WWA), Harris hawk algorithm (HHO), Moth Flame Optimization (MFO), Bacteria Foraging Optimization (BFO), Cuckoo Search (CS), Social Spider Optimization (SSPA), Bat Algorithm (BA), Naked Mole-Rat Algorithm (NMRA), Firefly Algorithm (FA), Salp Swarm Algorithm (SSA), Gray Wolf Optimizer (GWO), Artificial Bee Colony (ABC), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Shuffle Frog Leaping Algorithm (SFLA), and Chicken Swarm Optimization (CSO).

We propose an improved Harris hawk algorithm (i-HHO) in this research to help various classifiers with classification problems by reducing features from the Arrhythmia dataset. The HHO is proposed in 2019 [2]. In order to examine the prey,

surprise pounce, and other attack techniques of hawks in nature, HHO typically imitate the notions of Harris hawks. The literature claims that HHO outperformed other well-known metaheuristic algorithms for a number of benchmark tests [2]. HHO has the best qualities among its rivals when it comes to maintaining a steady balance between exploitation and exploration, which enables it to outshine in optimization jobs. HHO employs many search strategies in its exploitation to effectively impact local search outcomes. The HHO algorithm can be regarded as a robust technique for addressing optimization problems. The no free lunch theorem, as stated in reference [3], asserts that there exists no one solution that can effectively address all optimization problems within the domain of SIAs. As a result, many authors became interested in both creating new algorithms and enhancing those found in the literature. Our decision to propose a new improved form of HHO in this work is motivated by this. This work's significant contributions, in particular is as follows:

- A system is designed to classify the dataset using Arrhythmia dataset
- Feature selection is performed using an improved Harris Hawk Optimization algorithm (i-HHO) that helps in removing unwanted features present in the dataset
- The i-HHO is improved by incorporating levy flights, chaotic sequence and, opposite elite learning recombination mechanism.
- Classification is performed using various machine learning models, and the results are obtained.
- The system outperforms the existing state-of-art systems concerning accuracy, precision, recall, and F-measure.

II. LITERATURE REVIEW

In their study, Xie et al., introduced a ML approach that uses the salp swarm algorithm to diminishing the dimensionality of the problem for enhancing the efficacy of the classification algorithms. The classification task utilizes Random Forest, Support Vector Machine (SVM), and Kernel Extreme Learning Machine (KELM) algorithms. The KELM model demonstrated superior performance, with an accuracy rate of 84% [4].

In their study, Cristal, S. et al., suggested methodology that leverages deep learning techniques and utilization of the basic Harris Hawk optimization algorithm for the purpose of extracting features from the data. These extracted features are subsequently forwarded to the classification stage, where enhanced convolutional neural networks are employed to determine the presence or absence of heart disease in the patient. The model performed, with an accuracy rate of 99.2% [5].

In their study, Haseena et al., used Moth-Flame Optimization (MFO) algorithm to identify crucial features that might improve the effectiveness of decision tree (DT), support vector machine (SVM), artificial neural network (ANN), and K-nearest neighbour (KNN) classification algorithms employed in the diagnosis of cardiac disease. The model under consideration has exceptional performance, with an accuracy rate of 99.2% [6].

In their study, Ahmad Ayid Ahmad et al., and colleagues proposed an algorithmic strategy inspired by jellyfish algorithm for the purpose of diminishing the dimensionality of the Cleveland dataset. The dataset generated by the Jellyfish method was trained using the SVM classifier, resulting in accuracies of 98.4% and 94.4% for the corresponding models [7].

In their study, Veeranjanyulu et al., introduced an improved dragonfly algorithm, in order to address the challenge of overfitting. The proposed MLP-EBMDA model, demonstrated a predictive accuracy of 84% with the random forest algorithm [8].

In their study, in 2019, T. Keerthika et al., introduced a novel approach known as the hybrid fish bee optimization algorithm (HFSBEE) for the purpose of heart disease prediction. The method under consideration aims to parallelize two distinct algorithms, namely the artificial bee colony algorithm and the fish swarm optimization algorithm. This parallelization strategy

leverages the local behavior of the artificial bee colony algorithm and the global search capabilities of the fish swarm optimization algorithm to successfully identify the ideal solution. The classification procedure involves the conversion of the medical dataset into the multi-kernel support vector machine (MKSVM). The technique that was provided yielded results with a precision of 97.68% [9].

In their study, Santoshi, D et al., introduced a novel hybrid methodology for the detection of cardiac disease. The Grey Wolf Optimizer (GWO) algorithm is employed for the purpose of optimizing the merged dataset consisting of 12-lead PTB-XL ECG data and clinical data. The GWO algorithm is employed to identify and choose the most optimal characteristics from the combined dataset. Additionally, it decreases the dimensionality of the generated dataset. Subsequently, assessments are conducted utilizing deep neural networks on the combined dataset. The findings of this study indicate that the utilization of convolutional neural networks (CNNs) employing the Visual Geometry Group (VGG-16) provide superior performance compared to other techniques. Performance is assessed using multiple metrics, including accuracy, recall, precision, and the area under the receiver operator characteristic curve (AUC) [10].

In their study, Shahrock Asadi et al. introduced a novel model that integrates particle swarm optimization (PSO) with random forest algorithms for the purpose of heart disease prediction. The aim of the study was to generate decision trees that are both diverse and accurate, while also determining the ideal number of decision trees to be used concurrently. This study utilizes an evolutionary multi-objective strategy as an alternative to the regularly employed bootstrap method, feature selection in the random forest, and random number selection. The proposed method involves generating different training sets with varying samples and features for training each tree. The purpose of this action is to improve the performance of the random forest. The efficacy of this strategy was examined by conducting a comparative analysis of its performance across six heart datasets in comparison to individual and ensemble classifiers. The findings indicate that the approach suggested, which utilizes an appropriate number of classifiers, demonstrates superior performance compared to the random forest algorithm employing various classifiers [11].

In their study, in 2012, Mohamed G. El-Shafiey et al. presented a novel approach for heart disease prediction, integrating a hybrid optimization technique combining Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), will be introduced. This approach aims to enhance the performance of a Random Forest model utilized for heart disease prediction. The initial step of this methodology involves the identification and selection of the most salient elements that have the potential to enhance the precision of cardiac disease prognosis through the utilization of Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) techniques. In this study, a modified genetic

algorithm (GA) is employed to perform global search, while a particle swarm optimization (PSO) algorithm is utilized for local search. The efficacy of the proposed GAPSO-RF methodology is assessed through the utilization of evaluation measures, specifically accuracy, specificity, sensitivity, and area under the curve. This assessment is conducted on two distinct datasets, namely Cleveland and Statlog. The experimental findings validate that the GAPSO-RF methodology achieved accuracy rates of 95.6% and 91.4% on the Cleveland and Statlog datasets, respectively [12].

In their study, G. Thippa Reddy et al., introduced an algorithm that utilizes oppositional fireflies in conjunction with BAT and rule-based fuzzy logic (RBFL) for the purpose of predicting cardiac disease. In this study, the integration of opposition-based learning (OBL) with the Firefly BAT algorithm is proposed as a means to enhance the efficacy of the FAT algorithm, while concurrently optimizing the rules of the fuzzy logic system. Subsequently, the construction of the fuzzy system is facilitated through the use of fuzzy rules and membership functions, enabling the execution of classification tasks within the framework of the fuzzy system. Finally, the experimentation is conducted using publicly accessible UCI datasets, specifically the Cleveland, Hungarian, and Swiss datasets. The experimental findings demonstrate that the RBFL prediction algorithm exhibited superior performance compared to the existing technique, with an accuracy rate of 78% [13].

III. MATERIAL AND METHODS

The efficacy of machine learning models is reliant upon the quality of the features as input. As the quantity of features within the datasets escalates, the predictive efficacy of the model diminishes, while the computational expenses escalate. Certain features within the datasets exhibit negligible significance towards the outcome and consequently augment the computational intricacy of the model. The elimination of superfluous attributes mitigates interference and enhances the model's performance. Moreover, the process of feature selection holds significant importance in comprehending the underlying characteristics of the dataset. The careful selection of attributes enhances individuals' comprehension of the material. The present work employed the Harris hawk optimization algorithm to identify the optimal features from the dataset. Machine learning models are constructed based on the data utilized during the training phase. The process of selecting optimal features enhances the generalizability of the features acquired by the model. Therefore, this enhances the performance of the model when applied to novel data.

A. Basic Harris Hawk algorithm

The Harris hawk algorithm (HHO) is a swarm-based method that was first introduced by Heidari et al. in 2019 [2]. The Harris hawk algorithm is designed to replicate the foraging behavior

observed in Harris hawks. The algorithm replicates the tactics utilized in investigating prey, launching attacks, and executing surprise pounces. In the presented algorithm, hawks are utilized to represent potential solutions, while the prey symbolizes the ultimate best option. The Harris hawk utilizes its formidable visual acuity to diligently monitor and execute ambush tactics in order to apprehend detected prey. The optimizer employs a process wherein the search agents undergo two phases of exploration and four phases of exploitation for their updates. The system employs many time-varying methods in conjunction with a greedy scheme in order to improve the quality of the outcomes. The symbols utilized in this context are delineated in Table 1.

TABLE I. SYMBOLS USED IN MODELING OF HHO

Description	Symbol
Location of nth hawk, Position vector of hawks,	X_n, X
Rabbit position (best agent)	X_r
A random hawk position	X_k
Hawks Mean position, Jump strength	X_m, J
Maximum iterations, iteration counter, Swarm size	T, t, N
Random numbers in [0,1]	$r1, r2, r3, r4, r5, q, u, v, rand$
Upper and Lower bounds of variables	ub, lb
Initial state of energy, Escaping energy	$E0, E,$
Difference between prey position and current hawk, A Random Vector	$\Delta X, \alpha$
Transfer functions	$T(X)$
Dimension and Complement of X	$D, \neg X$

1) Exploration Phase

HHO works in two phases to perform exploration using Eq. (1) and Eq. (2) [2] [14].

$$X(t+1) = \begin{cases} X_k(t) - r_1 |X_k(t) - 2r_2 X(t)| & q \geq 0.5 \\ X_r(t) - X_m(t) - r_3(lb + r_4(ub - lb)) & q < 0.5 \end{cases} \quad (1)$$

$$X_m(t) = \frac{1}{N} \sum_{n=1}^N X_n(t) \quad (2)$$

2) Exploitation to Exploration Phase

$$E = 2E_0 \left(1 - \frac{t}{T}\right) \quad (3)$$

$$E_0 = 2r - 1 \quad (4)$$

Here E and E0 represents escaping energy and the initial state of energy.

3) Exploitation Phase

a) *Soft Besiege*

The soft besiege operation is performed using Eq. (5), Eq. (6) and Eq. (7) [2] [14].

$$X(t+1) = \Delta X(t) - E|JX_r(t) - X(t)| \quad (5)$$

$$\Delta X(t) = (X_r(t) - X(t)) \quad (6)$$

$$J = 2(1 - r_5) \quad (7)$$

b) *Hard Besiege*

The hard besiege operation is performed using Eq. (8). [2] [14]:

$$X(t+1) = X_r(t) - E|\Delta X(t)| \quad (8)$$

c) *Soft Besiege (Progressive Rapid Dives)*

The hawk chooses possibly best dive for the new position using Eq. (9) and Eq. (10) while $r < 0.5$ and $|E| \geq 0.5$ [2] [14]:

$$Y = X_r(t) - E|JX_r(t) - X(t)| \quad (9)$$

$$Z = Y + \alpha \times \text{levy}(D) \quad (10)$$

Also, as an improvement to the algorithm, the levy flight based on a Levy function calculated using Eq. (11) and Eq. (12) is also employed here [2] [14].

$$\text{Levy}(x) = 0.01 \times \frac{\mu \times \sigma}{|v|^{1/\beta}} \quad (11)$$

$$\sigma = \left(\frac{r(1+\beta) \times \sin(\frac{\pi\beta}{2})}{r(\frac{1+\beta}{2}) \times \beta \times 2^{\frac{(\beta-1)}{2}}} \right)^{\frac{1}{\beta}} \quad (12)$$

In this context, the symbol β represents a predetermined constant value, denoted as 1.5. During this stage, the location of the hawk is revised according to Eq. (13), Eq. (9) and Eq. (10) are utilized to compute the variables Y and Z.

$$X(t+1) = \begin{cases} Y & \text{if } F(Y) < F(X(t)) \\ Z & \text{if } F(Z) < F(X(t)) \end{cases} \quad (13)$$

d) *Hard Besiege with Progressive Rapid Dives*

The last stage, hard besiege is performed using Eq. (14) with progressive rapid dives.

$$X(t+1) = \begin{cases} Y & \text{if } F(Y) < F(X(t)) \\ Z & \text{if } F(Z) < F(X(t)) \end{cases} \quad (14)$$

Here, Y, Z, and $X_m(t)$ are calculated using Eq. (15), Eq. (16) and Eq. (2) respectively [2] [14]:

$$Y = X_r(t) - E|JX_r(t) - X_m(t)| \quad (15)$$

$$Z = Y + \alpha \times \text{levy}(D) \quad (16)$$

Finally, calculation of fitness function involves classification error rate and a minimum number of selected features, which can be mathematically formulated as:

$$\downarrow \text{Fit} = \beta \frac{|F|}{|N|} + \alpha \gamma(R) \quad (17)$$

Here,

- $\gamma(R)$ is the classifier error rate,
- $|F|$ is the number of selected features, and
- $|N|$ represents the total number of features.
- $\alpha \in [0, 1]$ and $\beta = (1-\alpha)$ are two factors.

The pseudo code of the basic HHO is given in the Fig. 1.

Begin:

Inputs: Number of Hawks (N) and MaxIter T,

Initialize the population of Hawks (X)

Output: Best Solution (X_r)

While (stopping condition is not met)

Compute the fitness of each hawks

Set a random X_r as the best solution

for (each hawk (X_i))

Update jump strength (J) and (E_0)

Update value of E using Eq. (3)

if ($|E| \geq 1$) then

Perform location vector updating using Eq. (1)

if ($|E| < 1$)

if ($r \geq 0.5$ and $|E| \geq 0.5$)

Perform location vector updating using Eq. (5)

else if ($r \geq 0.5$ and $|E| < 0.5$)

Perform location vector updating using Eq. (8)

else if ($r < 0.5$ and $|E| \geq 0.5$)

Perform location vector updating using Eq. (13)

else if ($r < 0.5$ and $|E| < 0.5$)

Perform location vector updating using Eq. (14)

Return X_r

End

Figure 1 Basic HHO Algorithm

B. *Chaotic sequence recombination mechanism*

If the running duration is indefinite, it is possible for every state to be achieved in a chaotic system due to random fluctuations. This suggests that the utilization of chaotic maps in the development of optimization techniques or their integration into underdeveloped optimization algorithms can effectively boost the exploration capabilities of such algorithms [15][16]. The utilization of chaotic sequences in optimization algorithms is a common practice aimed at mitigating the risk of premature convergence due to the sensitivity of initial conditions, the randomness inherent in chaotic sequences, and their ergodic properties [17][18]. As evidenced in the existing literature, the utilization of a chaotic sequence has been shown to significantly enhance the efficacy of the HHO algorithm by replacing the conventional random values. Consequently, the utilization of the

logistic mapping chaotic sequence is employed for the purpose of constructing the initial solutions inside the framework of the HHO algorithm. The logistic mapping can be represented using the following mathematical model.

$$\begin{aligned} u_{i+1} &= c \cdot u_i(1 - u_i), \\ i &= 1, 2, \dots, k; u_i \in (0, 1) \\ u_i &\neq 0.25, 0.5 \text{ and } 0.75 \end{aligned} \quad (18)$$

Here, k is the number of iterations, u_i denotes the chaotic variable in the i th iteration, and c denotes the control parameter, which is set to four. The chaotic search around the initial candidate solutions during the HHO beginning phase can increase population diversity and subsequently increase the algorithm's capacity for exploration. The following equation determines how to create the initial population P :

$$x_i = LB = rand() * (UB - LB), i = 1, 2, \dots, N_p \quad (19)$$

Here, the function $rand()$ generates a random number between $[0, 1]$ and x_i signifies the i th candidate solution. The chaotic sequence u_i is then added to P to create an updated population P_c .

$$P_c = P + u_i \cdot P \quad (20)$$

Recombining P_c and P creates the recombination mechanism; a new population is then created by choosing the solutions that match the initial NP fitness values. After carrying out these actions k times, a fresh beginning population is produced. The CSR method improves the distribution of the initialized solutions in the search space in addition to the random distribution utilized by the standard HHO, hastening the convergence of the HHO algorithm.

C. Opposite elite learning recombination mechanism

The objective of the machine intelligence technique referred to as opposition-based learning (OBL), initially proposed by Tizhoosh in 2005 [19], is to augment the capabilities of swarm intelligence (SI) algorithms. The central concept is upon the identification of a superior alternative between the existing individual and its corresponding opposite choice, with the determination being guided by fitness values. The OBL technique has demonstrated a greater probability of converging towards the global optimum of the objective function. Consequently, the utilization of the OBL technique has been widely adopted by researchers as a means to enhance the functionalities of SI algorithms. If x_i represents the hypothetical subject, the computation of the generalized opposite answer \bar{x}_i can be described as follows:

$$\bar{x}_i = LB = rand() * (UB + LB) - x_i, i = 1, 2, \dots, N_p \quad (21)$$

The parent generation is then represented by a population P made up of x_i ($i = 1, 2, \dots, NP$), whereas the offspring is represented by a population P_o made up of \bar{x}_i ($i = 1, 2, \dots, NP$). By recombining P_o and P , the recombination mechanism is completed, and a new population will be created by choosing the solutions that match to the first NP fitness values.

Several improvements have been implemented to enhance the performance of the algorithm. The starting population is widely considered as a critical aspect that significantly impacts the performance of the swarm algorithm. The OBL method is employed to determine the opposite solutions. Next, the fitness is computed for both x and \bar{x}_i , and the optimal N is selected from the union of x and \bar{x}_i . In every iteration, the opposition based solution is computed relative to a preset jumping rate (J) of 0.2. This is done to prevent an excessive diversity overflow that might potentially lead to a misleading search. A self-adaptive term is incorporated into the HHO framework and can be updated using the following equation.

$$Y_i(t+1) = X_i(t+1) + S_R(X_{cBest} - X_i)(t) \quad (22)$$

Here, X_{cBest} refers to the best solution obtained, $Y_i(t+1)$ is the new updated solution in the proposed algorithm, and S_R refers to the jumping rate or perturbation. In Eq. (22), the first term (social component) $X_i(t+1)$ is the output of the classical HHO obtained from updating Eq. 1. The cognitive component $S_R(X_{cBest} - X_i)(t)$ which is the self-adaption vector is added as a local search operator to be able to exploit areas around the best obtained solutions. Both social and cognitive components described in Eq. (22) provides a promising and good direction towards the current individual by combining individual best directions and the global best ones.

IV. RESULTS AND DISCUSSION

A. Dataset

The proposed approach is evaluated using Arrhythmia dataset which is achieved from the UCI [20]. The dataset used comprises a total of 279 attributes, with 206 of them being linear valued and the other attributes falling under the nominal category. Regarding the research conducted by H. Altay Guvenir, the objective is to differentiate between the existence and non-existence of cardiac arrhythmia, and afterwards categorize it into one of the sixteen separate classes. Class 01 denotes the standard electrocardiogram (ECG) pattern. Classes 02 to 15 encompass distinct classes of arrhythmia, while class 16 pertains to the remaining arrhythmias that have not been identified.

Begin:

Inputs: Number of Hawks (N) and MaxIter T ,

Initialize the population of Hawks (X)

Compute the fitness for each X_i , Calculate \bar{X} ,
Determine N solutions from $X \cup \bar{X}$
Output: X_r
While (stopping condition is not met)
Compute the fitness of each hawks
Set a random X_r as the best solution
for (each hawk (X_i))
Update jump strength (J) and (E_0)
Update value of E using Eq. (3)
if ($|E| \geq 1$) **then**
Perform location vector updating using Eq. (1)
if ($|E| < 1$)
if ($r \geq 0.5$ and $|E| \geq 0.5$)
Perform location vector updating using Eq. (5)
else if ($r \geq 0.5$ and $|E| < 0.5$)
Perform location vector updating using Eq. (8)
else if ($r < 0.5$ and $|E| \geq 0.5$)
Perform location vector updating using Eq. (13)
else
Perform location vector updating using Eq. (14)
end if
end if
Perform location vector updating using Eq. (22)
If (rand < OP)
Compute \bar{x}_{i+1} and its fitness
 $x_{i+1} = \bar{x}_{i+1}$ if $f(\bar{x}_{i+1}) < f(x_{i+1})$
end if
end for
Update X_r
Perform chaotic map sequencing
End
Return X_r

Figure 2 Proposed iHHO Algorithm

TABLE II. THE DESCRIPTION OF THE ARRHYTHMIA DATASET [20]

Class	Class Description	Instances
1	Normal	245
2	Ischemic changes (Coronary Artery Disease)	44
3	Old Anterior Myocardial Infarction	15
4	Old Inferior Myocardial Infarction	15
5	Sinus tachycardia	13
6	Sinus bradycardia	25
7	Ventricular Premature Contraction (PVC)	3
8	Supraventricular Premature Contraction	2
9	Left bundle branch block	9
10	Right bundle branch block	50
11	1. degree AtrioVentricular block	0
12	2. degree AV block	0
13	3. degree AV block	0
14	Left ventricle hypertrophy	4
15	Atrial Fibrillation or Flutter	5
16	Others	22

B. Evaluation Measures

The evaluation process utilizes various metrics, including specificity, sensitivity, accuracy, precision, and F-Measure, as outlined by Karalolis et al. in their 2010 publication [21]. The following defined rules are employed by them.

- **True Positive (TP):** A true positive (TP) refers to the accurate classification of patients with heart disease.
- **True Negative (TN):** The count of patients who do not have heart disease and are accurately identified as such.
- **False Positive (FP):** A false positive (FP) refers to the erroneous classification of healthy patients as individuals with cardiac disease.
- **False Negative (FN):** A False Negative (FN) refers to the count of individuals who are deemed healthy but are incorrectly labeled as having cardiac disease.

The measure of accuracy is determined by dividing the number of correct predictions by the total number of forecasts. The computation is performed with Equation 23.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (23)$$

Recall measures the number of actual instances rightly predicted. It is computed using Eq. 24.

$$Recall = \frac{TP}{TP+FN} * 100\% \quad (24)$$

F1-Score measures number of unreal instances rightly predicted. It is computed using Eq. 25.

$$F1 - Score = \frac{(2*Precision*Recall)}{(Precision+Recall)} * 100\% \quad (25)$$

The proportion of rightly predicted actual instances is known as precision and computed using Eq. 26.

$$Precision = \frac{TP}{TP+FP} * 100\% \quad (26)$$

The results obtained with the methods using the Arrhythmia dataset are elaborated in tables 3 and 4. The graphical comparison is depicted in Figures 11 and 12. Additionally, figures 3 to 10 depict the resulting confusion matrices. Table 1 presents the dataset, including the count of characteristics, instances, and classes associated with each dataset. The proposed methodologies are evaluated in comparison to the obtained outcomes. The studies were conducted on a PC equipped with an Intel i3 processor and 8 GB of RAM, operating on the Windows 10 64-bit operating system. All of the algorithms have been implemented using the Python programming language. The parameters for the maximum

number of iterations and the population size in the iHHO algorithm were defined as 50 and 100, respectively. This study employs a total of seven classifiers, including Support Vector Machine (SVM), Gradient Boosting (GB), Gaussian Naive Bayes (GNB), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), and K-Nearest Neighbours (KNN), to categorize the data generated by the iHHO. The evaluation measures included in this study are accuracy, precision, recall, and F-measure. Upon analysing the outcomes presented in Tables 3 and 4, it becomes evident that the suggested methodology outperforms the results achieved by the classifiers without iHHO algorithms, exhibiting a substantial disparity.

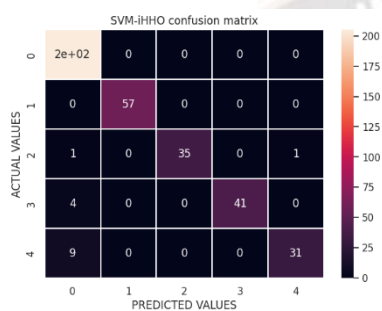


Figure 3 Confusion Matrix for SVM-iHHO

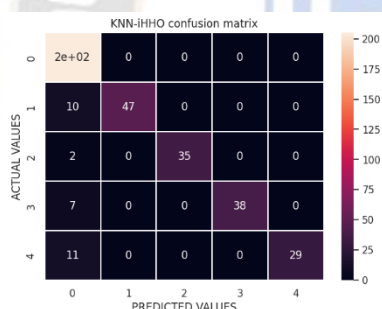


Figure 4 Confusion Matrix for KNN-iHHO

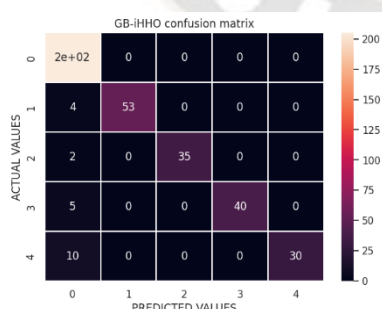


Figure 5 Confusion Matrix for GB-iHHO

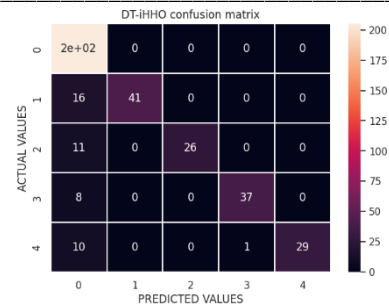


Figure 6 Confusion Matrix for DT-iHHO

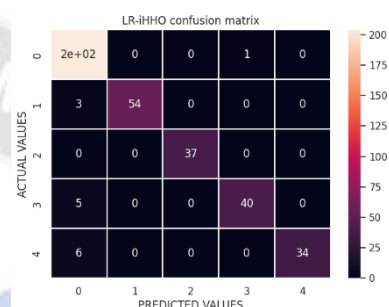


Figure 7 Confusion Matrix for LR-iHHO

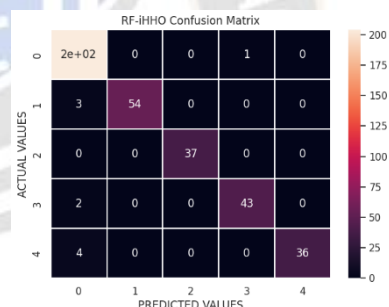


Figure 8 Confusion Matrix for RF-iHHO

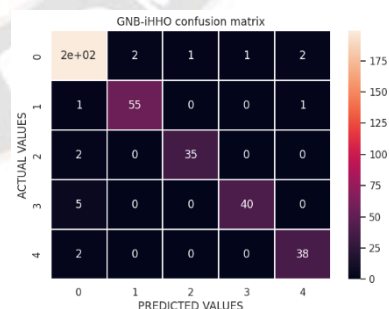


Figure. 9 Confusion Matrix for GNB-iHHO

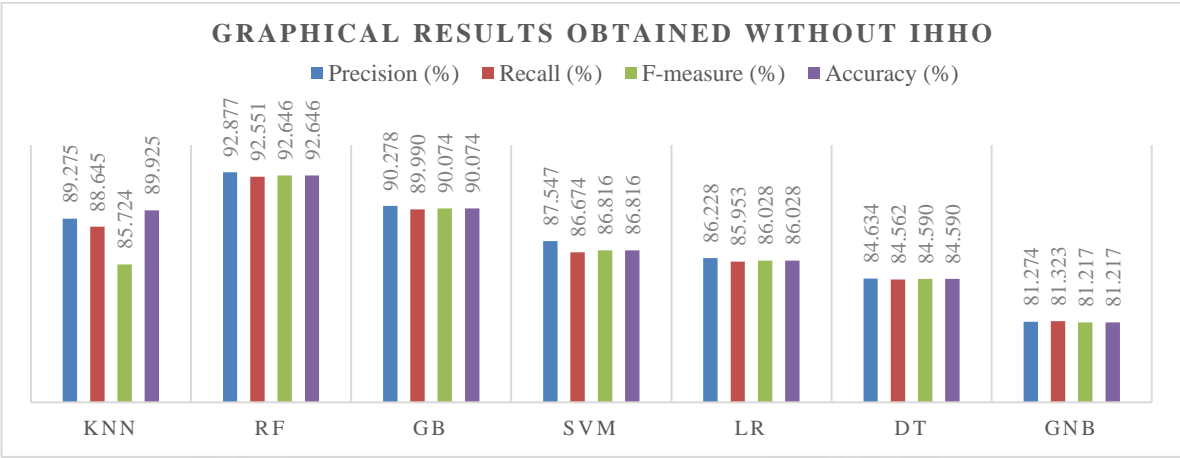


Figure 3. Graphical Results with Different ML Approaches without HHO.

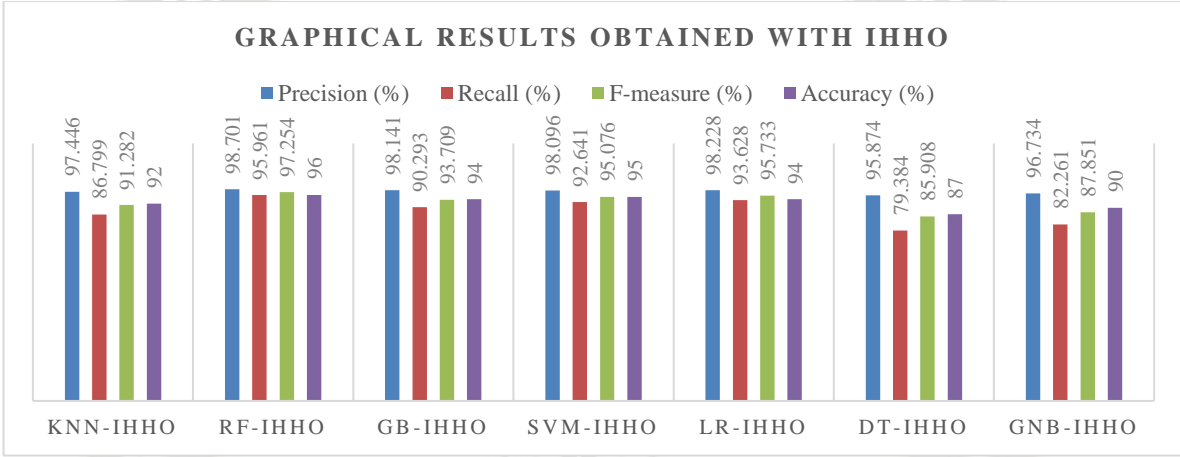


Figure 4. Graphical Results with different ML Approaches with iHHO.

TABLE III. RESULTS OBTAINED WITHOUT IMPROVED HHO

Method	Prec (%)	Rec (%)	FM (%)	Acc (%)
KNN	89.275	88.645	85.724	89.0
RF	92.877	92.551	92.646	92.0
GB	90.278	89.990	90.074	90.0
SVM	87.547	86.674	86.816	86.0
LR	86.228	85.953	86.028	86.0
DT	84.634	84.562	84.590	84.0
GNB	81.274	81.323	81.217	81.0

TABLE IV. RESULTS OBTAINED WITH IMPROVED HHO

Method	Prec (%)	Rec (%)	FM (%)	Acc (%)
KNN-iHHO	97.446	86.799	91.282	92.0
RF-iHHO	98.701	95.961	97.254	96.0
GB-iHHO	98.141	90.293	93.709	94.0
SVM-iHHO	98.096	92.641	95.076	95.0
LR-iHHO	98.228	93.628	95.733	94.0
DT-iHHO	95.874	79.384	85.908	87.0
GNB-iHHO	96.734	82.261	87.851	90.0

V. CONCLUSION

This research presents a study on the classification of Arrhythmia disease using seven ML classifiers: SVM, GB, GNB, RF, LR, DT, and KNN. We used an improved Harris-Hawk algorithm (iHHO) to perform the dimensionality reduction on the Arrhythmia dataset. Then, the reduced dataset is used by all classifiers with and without using iHHO and the results are collected, stored and compared. The comparison findings demonstrate a notable enhancement through the utilization of dimensionality reduction using iHHO. The results are improved in the range of 1% to 16%. GNB, SVM, and LR show very significant improvements of 10%, 10%, and 16%, respectively. Undoubtedly, there remains ample opportunity for enhancement in future endeavors. For instance, there is a need to enhance the dataset coverage in terms of both quantity and variety. The investigation of various strategies is necessary in the

context of dimensional reduction in order to ascertain their impact on the performance of machine learning.

REFERENCES

- [1] Y. Zhang, P. Agarwal, V. Bhatnagar, S. Balochian, and J. Yan, "Swarm Intelligence and Its Applications," *The Scientific World Journal*, vol. 2013, pp. 1–3, 2013, doi: 10.1155/2013/528069.
- [2] A. A. Heidari et al., "Harris Hawks optimization: Algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019. doi:10.1016/j.future.2019.02.028
- [3] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, Apr. 1997, doi: 10.1109/4235.585893.
- [4] W. He, Y. Xie, H. Lu, M. Wang, and H. Chen, "Predicting Coronary Atherosclerotic Heart Disease: An Extreme Learning Machine with Improved Salp Swarm Algorithm," *Symmetry*, vol. 12, no. 10, p. 1651, Oct. 2020, doi: 10.3390/sym12101651
- [5] S. Christal, G. Uma Maheswari, P. Kaur, and A. Kaushik, "Heart Diseases Diagnosis Using Chaotic Harris Hawk Optimization with E-CNN for IoMT Framework," *Information Technology and Control*, vol. 52, no. 2, pp. 500–514, Jul. 2023, doi: 10.5755/j01.itc.52.2.32778.
- [6] S. Haseena, S. K. Priya, S. Saroja, R. Madavan, M. Muhibullah, and U. Subramaniam, "Moth-Flame Optimization for Early Prediction of Heart Diseases," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–10, Sep. 2022, doi: 10.1155/2022/9178302.
- [7] A. A. Ahmad and H. Polat, "Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm," *Diagnostics*, vol. 13, no. 14, p. 2392, Jul. 2023, doi: 10.3390/diagnostics13142392.
- [8] T. Keerthika, "A Hybrid Fish – Bee Optimization Algorithm for Heart Disease Prediction using Multiple Kernel SVM Classifier," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9S2, pp. 729–737, Aug. 2019, doi: 10.35940/ijtee.i1152.0789s219.
- [9] R. Veeranjanyulu, et al., "Identification of Heart Diseases using Novel Machine Learning Method," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1–6, doi: 10.1109/ACCAI58221.2023.10200215.
- [10] S. Asadi, S. Roshan, and M. W. Kattan, "Random forest swarm optimization-based for heart diseases diagnosis," *Journal of Biomedical Informatics*, vol. 115, p. 103690, Mar. 2021, doi: 10.1016/j.jbi.2021.103690.
- [11] D. Santoshi, S. Chaudhari, and N. Pulgam, "Advanced approach for heart disease diagnosis with grey wolf optimization and deep learning techniques," *Data Intelligence and Cognitive Informatics*, pp. 631–647, 2022. doi:10.1007/978-981-19-6004-8_49
- [12] M. G. El-Shafiey, A. Hagag, E.-S. A. El-Dahshan, and M. A. Ismail, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest," *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 18155–18179, Mar. 2022, doi: 10.1007/s11042-022-12425-x.
- [13] G. T. Reddy and N. Khare, "An Efficient System for Heart Disease Prediction Using Hybrid OFBAT with Rule-Based Fuzzy Logic Model," *Journal of Circuits, Systems and Computers*, vol. 26, no. 04, p. 1750061, Dec. 2016, doi: 10.1142/s021812661750061x.
- [14] Too, Abdullah, and Mohd Saad, "A New Quadratic Binary Harris Hawk Optimization for Feature Selection," *Electronics*, vol. 8, no. 10, p. 1130, 2019, doi: 10.3390/electronics8101130.
- [15] Y. Xu et al., "An efficient chaotic mutative moth-flame-inspired optimizer for global optimization tasks," *Expert Systems with Applications*, vol. 129, pp. 135–155, Sep. 2019, doi: 10.1016/j.eswa.2019.03.043.
- [16] Q. Zhang et al., "Chaos-Induced and Mutation-Driven Schemes Boosting Salp Chains-Inspired Optimizers," *IEEE Access*, no. 7, p. 31243, 2019, doi: 10.1109/access.2019.2902306.
- [17] M. A. El-Shorbagy, A. A. Mousa, and S. M. Nasr, "A chaos-based evolutionary algorithm for general nonlinear programming problems," *Chaos, Solitons & Fractals*, vol. 85, pp. 8–21, Apr. 2016, doi: 10.1016/j.chaos.2016.01.007.
- [18] A. A. Ewees, M. A. El Aziz, and A. E. Hassanien, "Chaotic multi-verse optimizer-based feature selection," *Neural Computing and Applications*, vol. 31, no. 4, pp. 991–1006, Jun. 2017, doi: 10.1007/s00521-017-3131-4.
- [19] H. R. Tizhoosh, "Opposition-Based Reinforcement Learning," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 10, no. 4, pp. 578–585, Jul. 2006, doi: 10.20965/jaciii.2006.p0578.
- [20] Guvenir, H., Acar, Burak, Muderrisoglu, Haldun, and Quinlan, R. (1998). Arrhythmia. UCI Machine Learning Repository. <https://doi.org/10.24432/C5BS32>.
- [21] M. A. Karaolis, J. A. Moutiris, D. Hadjipanayi, and C. S. Pattichis, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 559–566, May 2010, doi: 10.1109/titb.2009.2038906.