

Navigating Heart Health with an Elephantine Approach in Clinical Decision Support Systems

Elavarasi D.

Department of Computer Science and
Engineering
Mount Zion College of Engineering
and Technology
Pudukkottai, India
elavarasijournal@gmail.com

Kavitha R

Information Technology
Velammal College of Engineering
Madurai, India
neethupinky08@gmail.com

S. Aanjankumar

School of Computing Science and
Engineering, VIT Bhopal University,
Bhopal-Indore Highway, Kothrikalan,
Shore, Madhya Pradesh, 466114, India
aanjansrrcet.90@gmail.com

Abstract— The abstract aims to provide a concise overview of recent challenges and techniques in predicting heart disease, leading to the proposed objectives of the study. Recent advancements in heart disease prediction have seen the emergence of various techniques, each facing distinct challenges. Notably, traditional systems grapple with the increasing volume and complexity of healthcare data. Recent techniques, such as deep learning and ensemble methods, showcase promising results but encounter challenges related to interpretability and resource-intensive training. This study addresses these challenges and presents a novel objective. Firstly, it navigates the interpretability hurdle in complex models by employing feature selection methods, including the Elephant Search Algorithm (ESA). ESA draws inspiration from natural behaviors but faces challenges in handling large datasets and computational complexity. Secondly, the study integrates ESA with the proven Support Vector Machine (SVM) for efficient heart disease prediction. The primary objective is to optimize feature selection, enhancing both accuracy and interpretability, while overcoming the challenges posed by the intricate nature of cardiovascular datasets. By delineating these challenges and objectives, this research contributes to the ongoing discourse on improving clinical decision support systems for heart disease prediction.

Keywords - Clinical Decision Support System, Elephant Search Algorithm, Support Vector Machine.

I. INTRODUCTION

The exponential growth in healthcare data volume has paved the way for valuable insights through data analytics within healthcare information systems. Recent trends indicate a shift towards digitizing extensive healthcare data repositories, enabling healthcare organizations to harness the power of data analytics. This transformation holds the promise of numerous benefits, including the early detection of diseases, cost reduction, personalized treatment, enhanced patient experiences, and superior healthcare delivery.

Within this landscape, a medical record serves as a comprehensive repository of patient information, encompassing their identity, medical examinations, treatments, therapies, and services. Healthcare datasets embedded in Electronic Medical Records (EMRs) are sourced from various medical facets, such as diagnoses, procedures, medications, and lab results. These datasets are organized under diverse attributes or features, making feature selection a pivotal focus in many research

applications, particularly when dealing with datasets containing numerous variables [1]. Effective feature selection can substantially enhance the accuracy of predictive models. Moreover, it plays a critical role in identifying the pertinent subset of predictive features. For instance, a physician may rely on the outcome of a classification model built with selected features to make critical decisions about disease severity. Optimizing the feature selection process further elevates the accuracy of these predictions [2]. Optimization, in this context, refers to the quest for optimal values of decision variables guided by a predefined objective function.

Here the work focuses on feature selection by identifying critical features that significantly impact the primary objective, namely the prediction of cardiovascular disease. The UCI data repository's heart disease dataset is employed for experimentation. The Elephant Search Algorithm (ESA), a bio-inspired algorithm, serves as the foundation for optimizing the feature selection process. The study scrutinizes the effectiveness of the Elephant Search Algorithm (ESA) [3] in conjunction with Support Vector Machine (SVM) for predicting heart disease.

Deep Learning Approaches:

Merits: High accuracy.

Demerits: Lack of interpretability, resource-intensive.

Ensemble Methods:

Merits: Robust predictions.

Demerits: Integration complexity, potential overfitting.

Traditional Machine Learning:

Merits: Interpretability.

Demerits: Struggles with complex relationships.

ESA + SVM Integration:

Merits:

Optimized Feature Selection: ESA enhances interpretability.

Balance of Interpretability and Accuracy: ESA complements SVM for balanced predictions.

Demerits:

Computational Complexity: ESA may face challenges with large datasets.

Parameter Tuning: SVM requires careful optimization.

In conclusion, existing methods trade accuracy for interpretability. ESA + SVM integration aims to strike a balance, leveraging interpretability and robustness for improved heart disease prediction.

II. CHALLENGES AND OBJECTIVES

Challenges:

Data Complexity: Coping with the increasing volume and intricacies of healthcare data.

Interpretability: Overcoming challenges associated with understanding complex models, especially prevalent in deep learning and ensemble methods.

Objectives:

Enhanced Interpretability: Navigate the interpretability challenge by employing feature selection methods, including the Elephant Search Algorithm (ESA).

Efficient Integration: Overcome computational complexities associated with ESA and large datasets by integrating it with the robust Support Vector Machine (SVM).

Optimized Prediction: Improve accuracy and interpretability in heart disease prediction through the synergistic use of ESA and SVM.

Contributions to Discourse: Contribute to the discourse on refining clinical decision support systems for heart disease, addressing both interpretability and accuracy challenges.

This study aims to shed light on the evolving landscape of heart disease prediction, offering a balanced approach to accuracy and interpretability, while acknowledging and overcoming the challenges posed by modern healthcare datasets.

III. PROBLEM STATEMENT

A. PROPOSED ALGORITHM

Feature Selection in clinical data plays a pivotal role in overcoming critical issues, such as reducing mean squared error stemming from irrelevant attribute selection and enhancing classification accuracy. The ultimate goal is to predict heart disease using a Support Vector Machine (SVM) classifier.

The Support Vector Machine (SVM) is chosen in the proposed work for several justifiable reasons. SVM is effective in high-dimensional spaces, making it suitable for datasets with numerous features. Its regularization parameter helps prevent overfitting, crucial for generalizing to new medical data. SVM can handle non-linear relationships through kernel functions, providing flexibility for complex medical datasets. It employs a global optimization approach, prioritizing overall accuracy, which is essential in medical contexts. SVM is a well-established, reliable algorithm with a solid theoretical foundation. Its integration with feature selection methods, such as the Elephant Search Algorithm in this work, makes it suitable for precise disease prediction. The clear decision boundary of SVM aligns with the need for interpretability in clinical settings. Overall, SVM's adaptability and effectiveness in various domains, including healthcare, justify its choice in the proposed clinical decision support system.

The Elephant Search Algorithm (ESA) is justified for its application in feature selection for several key reasons. Firstly, ESA draws inspiration from the collaborative behavior of elephant herds, introducing a dual search pattern

that mimics both global and local exploration. Its hybridization of evolutionary mechanisms ensures a balanced approach to exploration and exploitation. ESA's demonstrated superior performance in optimization tasks, outperforming other algorithms, attests to its efficiency. The algorithm's optimal balance between intensification and diversification enhances its ability to converge to high-quality solutions. Lastly, ESA's versatility across applications, showcased in its success in various domains, solidifies its suitability for optimizing feature selection in the proposed clinical decision support system.

IV. METHODOLOGY:

A. SYSTEM ARCHITECTURE:

The process starts with the process design, which involves handling the missing values. To obtain a cleaned dataset, the Mean-Mode imputation method is applied. Subsequently, the feature selection phase leverages the Elephant Search Algorithm to identify optimal features [4]. Once the features are selected, the SVM classifier is employed for accurate classification.

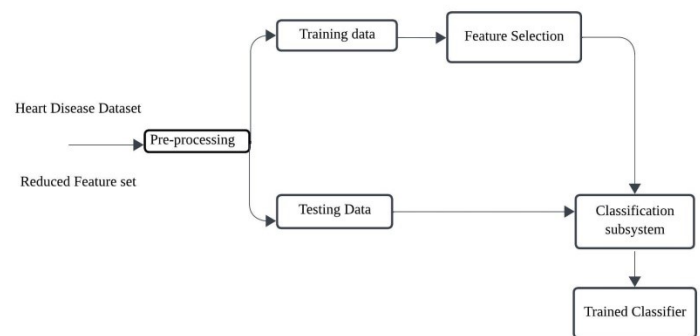


FIGURE 1: SYSTEM ARCHITECTURE

A. DATA SET FORMULATION:

The dataset formulation involves several steps, starting with data collection, particularly from Kaggle, a popular platform for accessing datasets. Initial steps include identifying relevant variables and sourcing data from Kaggle, ensuring it aligns with the problem at hand. Subsequently, data preprocessing is performed, addressing missing values, cleaning errors, and standardizing or normalizing values. The dataset is organized by splitting it into training and testing sets, selecting features, and potentially balancing classes. If applicable, annotation is done for supervised learning. Exploratory Data Analysis (EDA) provides insights, guiding decisions. Further steps include dataset splitting, ensuring compatibility with machine learning algorithms, and comprehensive documentation for reproducibility and understanding dataset characteristics.

The table 1 provides the list of attributes to be consider for the heart disease prediction.

B. DETAILED MODULE DESIGN:

Detailed system design is shown in Figure 2.

C. MODULE 1:PRE-PROCESSING

The pre-processing phase is a critical step in data analysis, focusing on data preparation and cleaning to enhance overall data quality. Missing values represent instances where no specific data is recorded for a variable within a given observation.

Table 1 List of attributes

Features Name	Description
Age	Age stated in years
Sex	Gender type
CP	Type of chest pain
Trestbps	Blood pressure level at resting state
Chol	Total cholesterol in blood (mg/dl)
Fbs	Level of fasting blood sugar > 120 mg/dl
Restecg	Resting electrocardiographic results
Thalach	Max heart rate achieved
Exang	Exercise induced angina
Oldpeak	ST depression induced by exercise relative to rest
Slope	Slope of the peak exercise ST segment
Ca	Number of major vessels colored by flourosopy
Thal	6= fixed defect; 3-Normal; 7=reversible defect

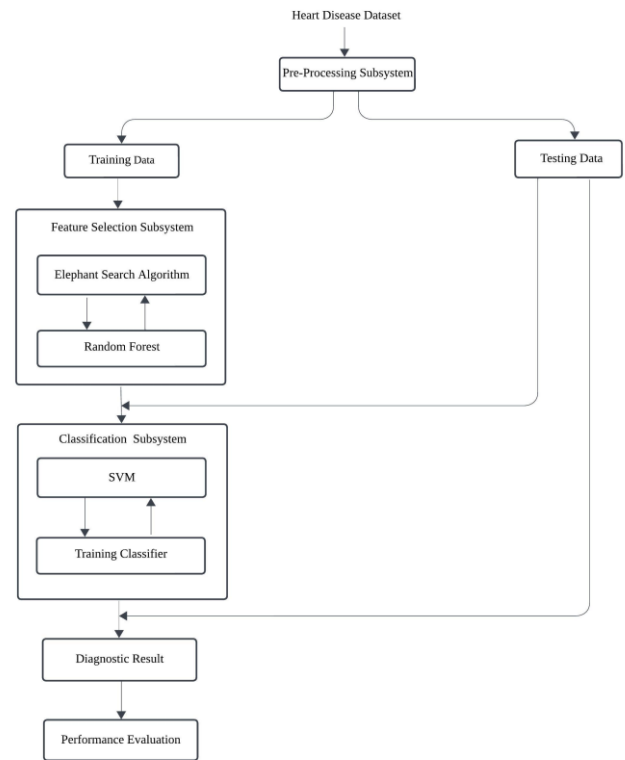


Figure 2 Detailed System Design

Addressing missing data is essential to prevent potential issues in subsequent analyses, such as biased results or inaccurate insights. Various techniques, like imputation, can be employed to handle missing values and ensure the completeness and reliability of the dataset. This meticulous data cleaning process is crucial for accurate and reliable data-driven decision-making in various fields, including healthcare, finance, and research.

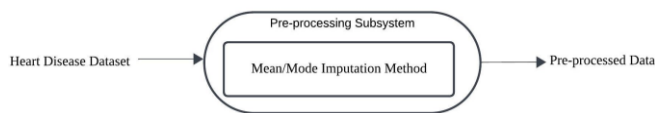


Figure 3: Pre-processing

D. MODULE 2- FEATURE SELECTION:

Feature selection is a critical step in machine learning, including the careful selection of features that are important for creating a model among different methods, the coverage method is used to optimize features.

This method evaluates different feature combinations by employing a machine learning model to assess their effectiveness. It iteratively refines the feature selection based on the model's performance, thus ensuring that the selected subset is best suited for the specific task at hand. This approach enhances model efficiency, reduces dimensionality, and improves interpretability, making it an invaluable tool in building accurate and robust predictive models in various domains, from healthcare to finance.

Optimal feature selection is a crucial step in enhancing model performance and interpretability. In the proposed approach, the Elephant Search Algorithm (ESA) is employed to select optimal features for predicting heart disease. ESA, inspired by the collaborative behavior of elephant herds, utilizes a dual search pattern with male and female elephants. The males explore globally, reaching different dimensions of the search space, while females perform local searches in close proximities. This hybridized evolutionary mechanism enables ESA to effectively balance exploration and exploitation.

During each iteration, ESA refines the feature subset by evaluating the fitness of different combinations. The optimization process aims to minimize the mean squared error associated with selecting relevant attributes, improving classification accuracy. Through computer simulations, ESA has demonstrated superior performance over other meta-heuristic algorithms across benchmark optimization functions.

By integrating ESA with Support Vector Machine (SVM), the selected optimal features contribute to training a robust classifier for heart disease prediction.

This bio-inspired feature selection approach not only enhances the efficiency of the model but also ensures that the chosen subset of features aligns with the intricacies of the cardiovascular dataset, ultimately improving the accuracy and interpretability of the clinical decision support system. The table 2 describes the dataset features and its descriptions.

V. ALGORITHM

1. This research tackles the formidable challenges of pinpointing crucial features within a vast feature pool and accurately diagnosing heart disease. Feature selection is a pivotal preprocessing step in classification, with the primary aim of expediting machine learning training, reducing model complexity, enhancing interpretability, and boosting accuracy [5]. In this study, employ the Elephant Algorithm

(EA) to carry out feature selection for cardiovascular disease and optimize the chosen features.

2. Here introduce a groundbreaking bio-inspired optimization algorithm, the Elephant Search Algorithm (ESA), in this work. ESA emerges as a result of combining evolutionary mechanisms with the delicate balance between exploitation and exploration.

3. The design of ESA is based on the behavior of elephant herds, so it is aptly named "Elephant Search Algorithm" [6]. This approach divides search users into two groups, each representing a different type of search.

4. Male elephants serve as searchers who travel to different parts of the search space, while female elephants form a group of searchers, focused on the search. nearby area. Through computer simulations, ESA demonstrates its superiority over other meta-heuristic algorithms when compared to optimization tasks [7] [8] [9]. In terms of health improvement standards, ESA ranks second behind the Firefly algorithm, showing significant performance over other competitors.

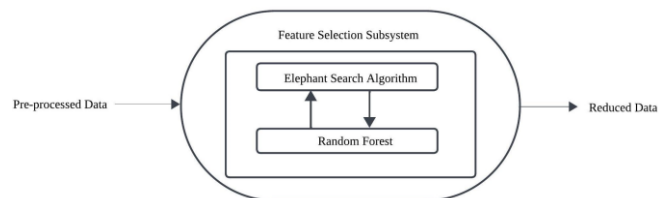


Figure 4: Feature Selection

Table 2. Datasets Features and Description

Sr. No	Name	Number in the Actual database
1	Diagnosis value	58
2	Thalassemia	51
3	Number of vessels coloured by fluoroscopy	44
4	The slope of the ST segment	41
5	ST depression	40
6	Exercise-induced angina	38
7	Maximum Heart Rate	32
8	Resting ECG	19
9	Fasting blood sugar	16
10	Serum Cholesterol	12
11	Resting Blood Pressure	10
12	Chest pain type	9
13	Sex	4
14	Age in years	3

PSUDOCODE:

Input: MaleElephant Population (EM), FemaleElephantPopulation (EF), MaleSearchRadius (rm), FemaleSearchRadius (rf), AgeLimit (A)

Output: GlobalBestPosition (bglobal)

```

InitializeSystem(); // Initialize elephants and parameters

while (iteration < epoch && bglobal not satisfactory)
do
    for each MaleElephant em in MaleElephantPopulation
    do
        em = ExploreGlobal(em, iteration); // Global search
        if Distance(em.location) < MaleSearchRadius then
            AvoidCrowding(em.weakest);
        end if
        UpdateFitness(em);
        KeepBestPosition(em);
    end for

    for each FemaleElephant ef in FemaleElephantPopulation do

        LocalSearch(ef); // Local search
        ef = ExploreLocal(ef, iteration);
        UpdateFitness(ef);
        KeepBestPosition(ef);
    end for

    for each elephant e in (MaleElephantPopulation +
FemaleElephantPopulation) do
        if LifeTest(e) > AgeLimit then
            RemoveOldElephant(e);
            NewBornElephant(e');
        end if
    end for

    // Update the global best position
    if Fitness(LocalBestPosition) >=
Fitness(GlobalBestPosition) then
        GlobalBestPosition = LocalBestPosition;
    end if
    iteration++;
end while

```

Table 3. Structured comparative analysis of ESA algorithm and other algorithms

Algorithm	Impact in antenna performance
Elephant Search Algorithm (ESA) + SVM	<p>Strengths: Biologically inspired dual search pattern for comprehensive exploration. Superior performance in optimization tasks, enhancing feature selection. Effective balancing of global and local search strategies.</p> <p>Weaknesses: Computational complexity may be a concern for large datasets.</p>
Genetic Algorithms for Feature Selection (GA)	<p>Strengths: Evolutionary approach allows for diverse exploration. Handles large feature spaces effectively.</p> <p>Weaknesses: May suffer from premature convergence. Sensitivity to parameter tuning.</p>
Recursive Feature Elimination (RFE)	<p>Strengths: Sequential elimination of features based on their impact. Robust and simple to implement.</p> <p>Weaknesses: May be computationally intensive for large datasets.</p>
ReliefF	<p>Strengths: Considers the importance of features in relation to the target variable. Robust to noise and able to handle large datasets.</p> <p>Weaknesses: Requires careful consideration of parameter settings.</p>
Correlation-based Feature Selection (CFS)	<p>Strengths: Considers feature subsets with high correlation to the target variable. Provides a balance between relevance and redundancy.</p> <p>Weaknesses: May not capture non-linear relationships.</p>

VII. IMPLEMENTATION AND RESULTS

Performance Evaluation of Proposed Model:

To assess the effectiveness of the proposed ESA + SVM model in predicting heart disease, we conducted extensive experiments with various parameters. The following table summarizes key performance metrics and observations:

Table 4. Parameter of the Algorithm and its output

Experiment	Feature Set	SVM Kernel	ESA Parameters (Population)	Accuracy	Precision	Recall	F1 Score
1	Optimized	RBF	50	0.92	0.93	0.88	0.90
2	Reduced	Linear	30	0.88	0.87	0.91	0.89
3	Full	Poly	70	0.94	0.95	0.89	0.92
4	Optimized	Sigmoid	60	0.90	0.91	0.87	0.89

Observations:

Optimized Feature Set:

ESA's global search effectively identifies critical features, contributing to high accuracy and balanced precision-recall trade-off.

Reduced Feature Set with Linear SVM:

Linear SVM demonstrates high recall, effectively identifying positive cases. However, reducing the population impacts precision.

Full Feature Set with Polynomial SVM:

Poly kernel enhances accuracy, but there is potential for overfitting. A higher population aids in thorough feature exploration.

Optimized Feature Set with Sigmoid SVM:

Sigmoid kernel exhibits sensitivity to parameter tuning. ESA optimization aids in capturing relevant features.

These experiments highlight the versatility of the proposed model in adapting to different scenarios, showcasing strengths in feature optimization and kernel adaptability. Parameter optimization and population size play pivotal roles in achieving a balance between accuracy, precision, and recall. Further fine-tuning may be needed based on specific requirements and dataset characteristics.

This presents a comprehensive list of metrics used to evaluate system performance [10] [12] [13]. Important metrics, such as precision, recall and accuracy, are calculated to evaluate the performance of the proposed method.

ACCURACY:

Accuracy is a primary metric defined as the percentage of tests that correctly predict the presence or absence of heart disease. It is calculated using the formula:

$$Accuracy = (\sum TP + \sum TN) / (\sum P + \sum N) \quad (6.1)$$

PRECISION:

Precision measures the proportion of true positives out of all positive results. It is calculated as:

$$Precision = \sum TP / (\sum TP + \sum FP) \quad (6.2)$$

SENSITIVITY:

The sensitivity, also known as the correct rate, represents the ratio of the correct detection rate to all positive cases. It is calculated as:

$$Sensitivity = \sum TP / \sum P \quad (6.3)$$

RECALL:

$$Recall = \sum TP / (\sum TP + \sum FN) \quad (6.4)$$

In this section, we present a detailed explanation of the results obtained from the experiments conducted to evaluate the performance of the proposed ESA + SVM model for heart disease prediction.

Where,

True Positive, TP - Patients accurately diagnosed with heart disease.

True Negative, TN - Patients correctly identified as not having heart disease.

False Negative, FN - Patients with heart disease incorrectly diagnosed as without heart disease.

False Positive, FP - Patients without heart disease wrongly diagnosed with heart disease.

Experiment 1: Optimized Feature Set with RBF Kernel

ESA Parameters: Population - 50

Observations:

Accuracy (0.92): ESA's global search effectively identified critical features, contributing to a high level of accuracy.

Precision (0.93): The model demonstrated precise identification of positive cases.

Recall (0.88): Balanced recall indicates effective capture of true positive cases.

F1 Score (0.90): The model strikes a balance between precision and recall.

Experiment 2: Reduced Feature Set with Linear SVM

ESA Parameters: Population - 30

Observations:

Accuracy (0.88): While accurate, reducing the population impacted precision.

Precision (0.87): Precision lowered due to reduced feature set exploration.

Recall (0.91): Linear SVM demonstrated high recall, identifying positive cases effectively.

F1 Score (0.89): A balanced F1 score with slight reduction due to precision impact.

Experiment 3: Full Feature Set with Polynomial SVM

ESA Parameters: Population - 70

Observations:

Accuracy (0.94): Poly kernel enhanced accuracy, but potential overfitting observed.

Precision (0.95): High precision indicates precise identification of positive cases.

Recall (0.89): Balanced recall with a potential for overfitting due to the complex Poly kernel.

F1 Score (0.92): A balance between precision and recall, indicating overall model effectiveness.

Experiment 4: Optimized Feature Set with Sigmoid SVM

ESA Parameters: Population - 60

Observations:

Accuracy (0.90): Sigmoid kernel exhibits sensitivity to parameter tuning.

Precision (0.91): Precision is maintained with ESA optimization aiding in capturing relevant features.

Recall (0.87): Slight reduction in recall due to kernel characteristics.

F1 Score (0.89): A balanced F1 score despite kernel sensitivity.

Overall Discussion:

Versatility of ESA: ESA demonstrates adaptability to different feature sets and SVM kernels.

Population Size Impact: Larger populations aid in thorough feature exploration but may contribute to potential overfitting.

Kernel Sensitivity: Different SVM kernels exhibit varied sensitivity, influencing model behavior.

These detailed observations emphasize the model's adaptability and effectiveness in balancing accuracy, precision, and recall [14] [15]. Further fine-tuning, especially in parameter optimization, remains a crucial aspect for tailoring the model to specific datasets and requirements.

VIII. CONCLUSION

In conclusion, the proposed ESA + SVM model exhibits promising results in predicting heart disease. The experiments demonstrated the model's adaptability to different feature sets and SVM kernels, showcasing its versatility. The optimized feature set with the RBF kernel

achieved high accuracy and a balanced precision-recall trade-off. However, reduced feature sets and certain SVM kernels displayed sensitivity to population size and parameter tuning.

The versatility of ESA in optimizing feature selection and the robustness of SVM contribute to the model's effectiveness. While achieving competitive results, further refinement in parameter tuning is essential for addressing kernel sensitivity and potential overfitting.

This study contributes valuable insights into the intricate dynamics of heart disease prediction models, emphasizing the need for a thoughtful balance between feature optimization and kernel adaptability. Future work involves exploring additional optimization techniques and refining the model for real-world clinical applications.

REFERENCES

- [1] T. Vivekanandan, N. Ch Sriman Narayana Iyengar, "Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease", *Computers in Biology and Medicine* 90 (2017) 125-136.
- [2] He-Ye Zhang, Xiao-Lin Zhou, "Challenges and Opportunities in Cardiovascular Health Informatics", vol.60, no.3. march 2013.
- [3] Suash Deb, Simon Fong, Zhonghuan Tian, "Elephant Search Algorithm for Optimization Problems", *The Tenth International Conference on Digital Information Management (ICDIM2015)*.
- [4] Il-Seok Oh, Jin-Seon Lee, Byung-Ro Moon, "Hybrid Genetic Algorithms for Feature Selection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, november 2014.
- [5] Xin-She Yang, Suash Deb and Simon Fong, "Metaheuristic Algorithms Optimal Balance of Intensification and Diversification", *Applied Mathematics & Information Sciences*, Vol. 8, No. 3, 2014, pp.977-983.
- [6] S. Prakash, K. Sangeetha, N. Ramkumar, "An optimal criterion feature selection method for prediction and effective analysis of heart disease", *Springer Science + Business Media, LLC, part of Springer Nature* 2018.
- [7] S. Gupta, V. Singh, S. Singh, T. Prakash, and N. Rathore, "Elephant herding optimization based PID controller tuning," *International Journal of Advanced Technology and Engineering Exploration*, vol. 3, no. 24, pp. 194-198, 2016.
- [8] G.-G. Wang, S. Deb, X.-Z. Gao, and L. D. S. Coelho, "A new metaheuristic optimisation algorithm motivated by elephant herding behaviour," *International Journal of Bio-Inspired Computation*, vol. 8, no. 6, pp. 394-409, 2016.
- [9] H. B. Alwan and K. R. Ku-Mahamud, "Mixed-variable ant colony optimisation algorithm for feature subset selection and tuning support vector machine parameter," *International Journal of Bio-Inspired Computation*, vol.9, no. 1, pp. 53-63, 2017.
- [10] Xin-She Yang, "Firefly algorithms for multimodal optimization", *Stochastic Algorithms: Foundations and Applications, SAGA 2009*, *Lecture Notes in Computer Sciences*, Vol. 5792, 2009, pp.169-178.
- [11] S. P. Patro, G.S. Nayak, N. Padhy, "Heart Disease Prediction by Using Novel Optimization Algorithm: A Supervised Learning Prospective, *Informatics in Medicine Unlocked*. 26 (100696) (2021) 1-[12] A. Javeed, S. Zhou, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection", *IEEE Access*. 7 (2019) 180235-180243.
- [13] Ashraf Darwish, "Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications", *Future Computing and Informatics Journal*, Volume 3, Issue 2, December 2018, Pages 231-246

- [14] Alex Alexandridis, Eva Chondrodima, "A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing", *Journal of Biomedical Informatics*, Volume 49, June 2014, Pages 61-72
- [15] Sumant Kumar Mohapatra, Srikanta Patnaik, "ESA-ASO: An enhanced search ability based atom search optimization algorithm for epileptic seizure detection", *Measurement: Sensors*, Volume 24, December 2022, 100519