SCIENTIFIC OASIS

Decision Making: Applications in Management and Engineering

ISSN: 2560-6018, eISSN: 2620-0104

DECISION MAKING: APPLICATIONS IN MANAGEMENT AND ENGINEERING

FBGN: A FUSION BASED GENETIC PSO NETWORK FOR HEART ATTACK PREDICTION

Dr. D. Durga Prasad 1, Yamanuri Pratyusha 2, Akula Jaya Hari Sai Jaswitha 3, Vinjamuri Mounika 4,Bayana Geetha5

[1,2,3,4,5] Department of Computer Science and Engineering, PSCMR College of Engineering and Technology, Vijayawada, A.P.

ABSTRACT

The most prevalent disease is heart disease. The most challenging task is predicting cardiac disease. Although there exist tools to predict cardiac disease, they are either more pricey or unsuccessful at doing so. We may be able to reduce the death rate somewhat if we can immediately predict cardiac problems. We have access to many data in the modern world, allowing us to anticipate employing machine learning methods to study heart illness. The existing model used to forecast cardiac problems is based on a Customised ensemble classification technique that uses a decision tree, SVM and logistic regression technique. The following are some of the system's drawbacks: more classifiers should first be evaluated second because the suggested model is based on the wrapper feature selection technique, and its high data processing price and temporal complexity as the main limitations. One option is to use a genetic algorithm. Minor data restrictions are the second. So, to rectify drawbacks in this project, use a genetic algorithm with particle swarm optimisation. This paper chooses the optimal attributes for predicting heart illness with numerous equations. Using numerous equations to forecast cardiac disease is innovative in this work.

Keywords:

Heart disease; Feature selection; Customised ensemble classification; Particle Swarm Optimization; and Genetic Algorithm.

1. Introduction

Heart disease is growing daily worldwide, causing mortality in some cases. Heart disease can be identified by some of the symptoms, such as chest pain, blood pressure, cholesterol and age. This paper predicts cardiovascular disease using machine learning techniques. Machine learning can handle large amounts of data [11].

The dataset is taken as the Cleveland dataset since it has many samples [1]. To eliminate missing values, this paper used one hot encoding technique [2]. To improve the methodology specified in

[4], this paper used feature selection techniques to select the optimal features to reduce dimensionality. This paper has an Ensemble classifier to eliminate the slow classification process [3]. This paper uses a feature selection process to predict heart disease well for dimensionality reduction [5].

The feature selection methods in this research include the genetic algorithm and particle swarm optimisation. This paper merges these two to improve feature selection process capability and prevent local optima problems in PSO. PSO's local optima problem limits the search space in which features can be found. The search process can be significantly enhanced if the strategies listed above are combined with various aims. To classify data, employ an ensemble classification technique that combines SVM, logistic regression, and decision trees. Because we are integrating those three, this prediction accuracy will increase, and training time will be reduced compared to using only one classifier.

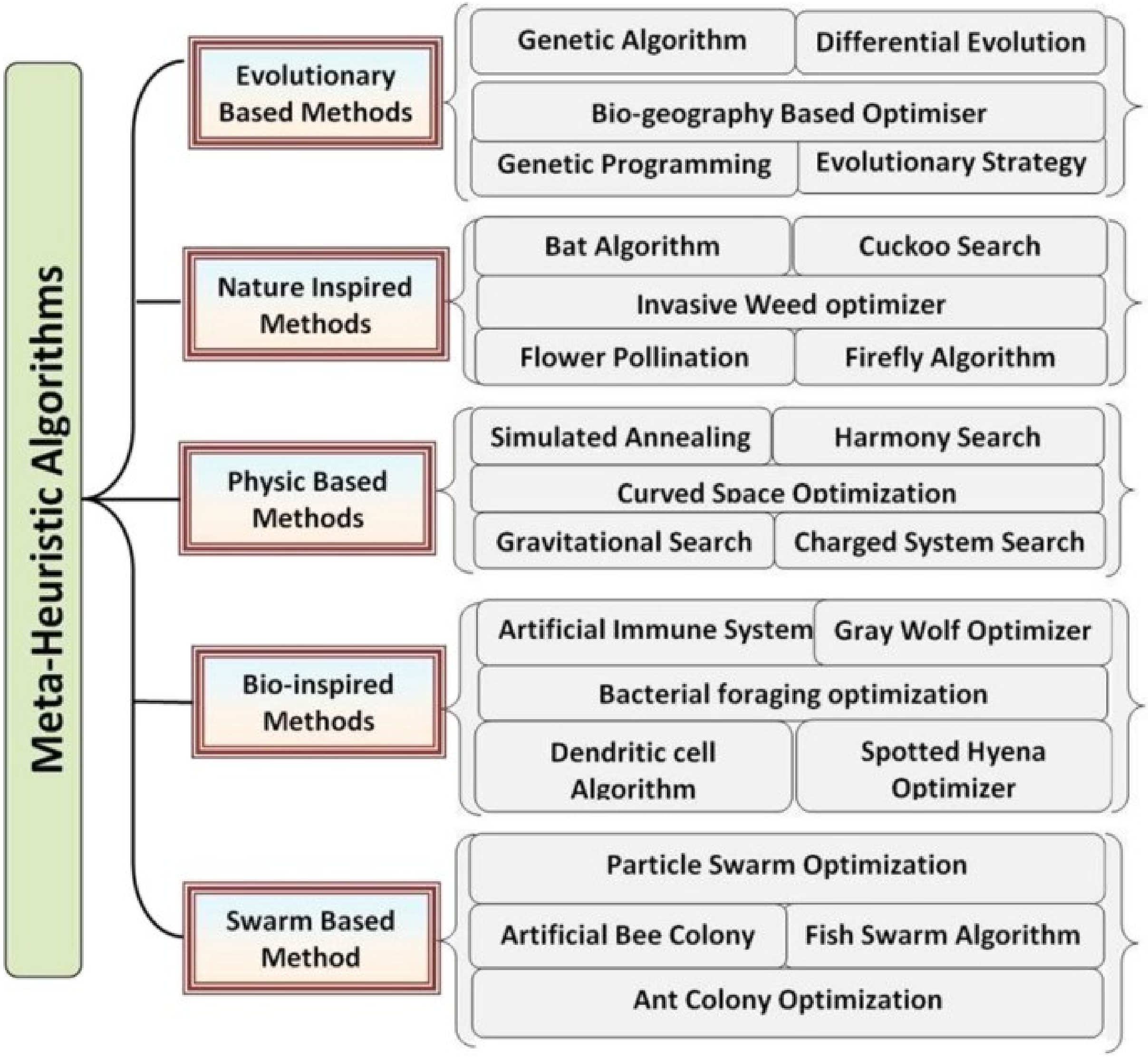


Figure 1: Meta-Heuristic Algorithms

A metaheuristic algorithm, a search method, can deal with an optimisation problem that is complicated and hard to solve optimally. In the real world of scarce resources (e.g., processing power and time), it is essential to identify a close to optimal solution based on inadequate or incomplete knowledge. One of the most notable developments in operations research over the past 20 years has been the development of metaheuristics for resolving these optimisation issues.

Some issues demand attention to develop superior solutions above the already used conventional methods. Authors who cover a wide range of applications for solving optimisation issues explain several metaheuristic techniques.

An evolutionary algorithm operates on a relatively simple principle called an EA. An EA's four main steps are initialisation, selection, genetic operators, and termination. Each of these phases corresponds to a distinct feature of natural selection and offers simple techniques for modularising applications of this algorithm category. In simple terms, more fit individuals will remain and multiply under an EA, but less fit individuals will perish rather than contribute to the gene pool of later generations, much like in natural selection.

Swarm Intelligence, sometimes known as artificial groups of simple agents. It is the collective intelligence behaviour of self-organised and distributed systems. Social insects create nests, work together to transfer objects and engage in communal sorting and clustering as examples of SI. Self-organisation and the division of work are two critical ideas regarded as essential elements of SI.

A group of algorithms known as nature-inspired optimisation algorithms (NIOA) are influenced by how things behave in the natural world. Animal behaviours, biology, chemical reactions, etc., have all served as inspiration for NIOAs. This has made engineering solutions, medical treatments, etc., possible. Multiple intricate sub-processes can be included in natural processes. Because of this, every algorithm is distinct and potent. The goal of the study of NIOAs is to improve the efficiency of nature-inspired algorithms by addressing concerns with algorithm selection, parameter tuning, and algorithm updates. Natural occurrences are constantly changing, and eventually, an algorithm's behaviour must also change.

Each bio-inspired optimisation algorithm has unique characteristics that make it a better alternative than other algorithms to handle a particular problem. Each bio-inspired optimisation algorithm has been built in a generic way to cope with any optimisation challenge. For instance, bee colony algorithms are a good choice for dealing with combinatorial optimisation problems, ant colony algorithms are typically a good choice for dealing with routing problems, and genetic algorithms are generally a good choice as long as complex solution reconstruction techniques do not need to be developed when obtaining infeasible solutions after applying evolutionary operators. Therefore, the selection of the appropriate algorithm to use for each problem and the algorithm's design depends on the expert's competence.

2. Literature Survey

In 2022, Md G. El- Shafiey et al. [1] proposed a cardiac disease prediction method employing PSO techniques. Cleveland and Statlog are the datasets used in this suggested methodology. The classification is contrasted with other cutting-edge techniques. As a classifier, Random Forest is employed. By choosing the best attributes, the suggested method improved accuracy. The Cleveland dataset's accuracy was 95.6%, while the Statlog dataset's accuracy was 91.4%. The key drawback of this methodology is that more classifiers must be examined to get better results. Therefore, intend to employ genetic algorithm to overcome this restriction.

In 2021, P. Ghosh et al. [2] proposed illness of heart prediction using LASSO as a feature selection technique. The categorisation of cardiac disorders in this Cleveland, Statlog and Hungarian datasets was employed. Using LASSO, the appropriate features are chosen. The classification results are presented differently so that they may be compared to those of the other classifiers. With the help of the Random Forest bagging approach, the suggested model produced excellent accuracy (RFBM). The drawback is that the model becomes increasingly accurate even for a massive dataset due to the high percentage of missing values and overreliance on a particular feature selection method. For a better categorisation, it is necessary to explore alternative feature choices to get around this.

In 2022, Wiharto et al. [3] proposed a feature selection method that depends on a genetic algorithm for detecting coronary heart disease. The Cleveland and statlog datasets are employed, and the suggested approach uses SVM-GA together through a fast correlation-based filter (FCBF) for feature selection. We have seen that in this study, the bagging method outperforms other algorithms like RF and C4.5, allowing the suggested diagnosis system model to function more effectively. Its weakness is that it needs more speed for real-time situations due to its slow categorisation process.

In 2019, Senthilkumar Mohan et al. [4] proposed a methodology using an effective cardiac disease prediction using hybrid machine learning techniques. The Cleveland dataset, accessible on the UCI repository, was utilised in this suggested technique. They developed a brand-new strategy for identifying essential information using machine learning techniques to enhance prediction. Various characteristics and classification approaches were used to make the prediction. By applying a mixed random forest with a linear model and various attribute combinations, they were able to build a prediction model for cardiac disease with a performance accuracy level of 88.7%. The 13 characteristics were used to evaluate each model's performance. Based on error optimisation on the dataset, the classifiers are assessed. Compared to other approaches, the HRFLM classification method offers the best accuracy. In this methodology, the limitation is that this model cannot control the mortality rate because of inefficient prediction. This can be further improved by using the best feature selection techniques.

In 2020, R. Spencer et al. [5] proposed a heart disease prediction using feature selection and classification methods. The dataset utilised in the suggested technique is a composite dataset. Different methodologies are used to examine these four datasets. The Bayes Net method and Chi-squared feature selection produced the best model, which had an accuracy of 85%. This paper's drawback is that just ML and a few feature selection techniques were employed. Different feature selection approaches can be employed to get around this restriction.

In 2022, Neil Gordon et al. [6] proposed an addressing optimisation challenge employing evolutionary algorithms to accomplish feature selection on datasets with plenty of variables. A dataset including actual heart failure is employed in the suggested technique. To identify the most important variables, this study used feature selection to a published dataset using an experimental technique. It has been reported that a feature selection approach is used as part of an optimisation technique that uses a genetic algorithm to improve accuracy. The primary drawback of this technology is the evolutionary algorithm used to wrap the selected characteristics. Since the chances of overcoming this restriction are slim, even a slight performance improvement is valuable.

In 2016, Kamilia Menghour et al. [7] put forth an ACO-PSO based approach for feature selection using I ant colony optimisation (ACO), particle swarm optimisation (PSO), and hybrid bio-inspired approach for feature selection in this proposed methodology. We pointed out that these approaches' efficacy is measured concerning basic bio-inspired feature selection methods. The effectiveness of these methods is evaluated against more established bio-inspired feature selection techniques.

In 2019, Youness Khourdifi et al. [8] proposed a cardiovascular disease prediction using particle swarm and ant colony optimisation. The Fast Correlation-Based Feature Selection (FCBF) approach was employed in the article's proposed strategy to avoid redundant data and enhance classification accuracy. Results show the effectiveness and consistency of the suggested hybrid technique for managing various kinds of data for the division of cardiac disease, as it is discovered. The suggested mixed strategy is applied to a heart disease patient dataset. The knowledge base of the author, the equipment used for the inquiry, such as the processing power of the computer, and the time limits of the study are only a few of the limitations of this study. Modern technologies and subject-matter knowledge are thus required for this sort of study to overcome this constraint.

In 2022, Jan Carlo T. Arroyo et al. [9] proposed a modified neural network using a genetic algorithm for cardiac disorder prediction. Artificial neural networks were trained and tested on 70,000 instances of the cardiovascular disease dataset, which contains 12 variables (ANN). It has been noted that this effort tries to improve ANN performance to boost prediction accuracy. To attain this, it was observed that the Genetic Algorithm (GA) was applied. The GA-ANN prediction model fared better than other ML techniques for predicting cardiovascular disease. It is suggested to include GA in prediction models other than ANN.

In 2021, Laith Abualigah et al. [10] proposedcombining a hybrid cosine method and evolutionary algorithm to discover unique features for data mining tasks. KNN classifier is also used in this suggested methodology in addition to the evolutionary algorithm and the binary version of the sine-cosine technique. The suggested SCAGA performed better when the exploitation and exploration approaches for the search space were balanced. PSO, ALO, SCA, and SCAGA are analysed in classification using KrvskpEW and waveformEW. The proposed approach may be applied in further research to generalise the approach across other domains and additional datasets; using innovative optimisers, such as the arithmetic optimisation method, the feature selection problem can be handled (AOA).

S No	Author	Findings	Technique	Limitations
1	Md G. El-Shafiey	The proposed approach increased the accuracy by selecting the optimal features. The accuracy obtained on the Cleveland dataset was 95.6%, and on the Statlog dataset was 91.4%.	PSO based feature selection is made. Random Forest is used as a classifier.	The main limitation is that more classifiers must be evaluated to improve the results.
2	P. GHOSH	When the classification is done, the results are shown differently than the other classifiers. The proposed model obtained high accuracy using the Random forest bagging method (RFBM).	In this Cleveland, Hungarian and Statlog datasets are used to classify heart disease. The convenient features are selected using the Least Absolute Shrinkage and Selection Operator (LASSO).	The limitation of this is that there are high values of missing values.
3	Wiharto	The proposed	SVM-GA	Its shortcoming

		diagnosis system model can perform better when the bagging-LMT algorithm is used than other algorithms, such as RF and C4.5.	merged with a fast correlation-based filter (FCBF) was applied for Feature selection, and The Datasets Cleveland and statlog were used.	is its slow classification process, which makes it unsuitable for real-time instances.
4	Senthilkumar Mohan	The classifiers are evaluated based on error optimisation on the dataset. HRFLM classification method	They proposed a novel method for finding significant features by applying ML techniques to enhance accuracy in diagnosing cardiac disease.	The limitation is that this model cannot control the mortality rate because of inefficient prediction; this can be further improved using the best feature selection techniques.
5	R. Spencer	The best model created using Chi-squared feature selection merged with the BayesNet algorithm.	A combined dataset is used as a dataset. These four datasets are evaluated using different techniques.	The limitation of this paper is that they have used only ML and tried some feature selection techniques.

3. Proposed Methodology

Figure 2 discusses the methodology of the current paper. It contains multiple phases. The foremost step is dataset collection. The dataset here is considered the Cleveland dataset taken from the UCI repository since it has a large number of samples, 1024, and the dataset initially has 14 features. The second step is preprocessing. The samples contain duplicate data, which is why they preprocessed the dataset. Therefore, we employed one hot encoding strategy to clean up the dataset.

In the next step, the preprocessed dataset undergoes feature selection technique i.e., genetic algorithm to get the optimal features as feature subset. The word Objective indicates an equation. Here equation is used to enhance the feature selection process, which is also this paper's novelty. After the genetic feature selection technique, only 10 features are encountered as reduced feature subsets. This reduced feature subset is named FS1 for understanding.

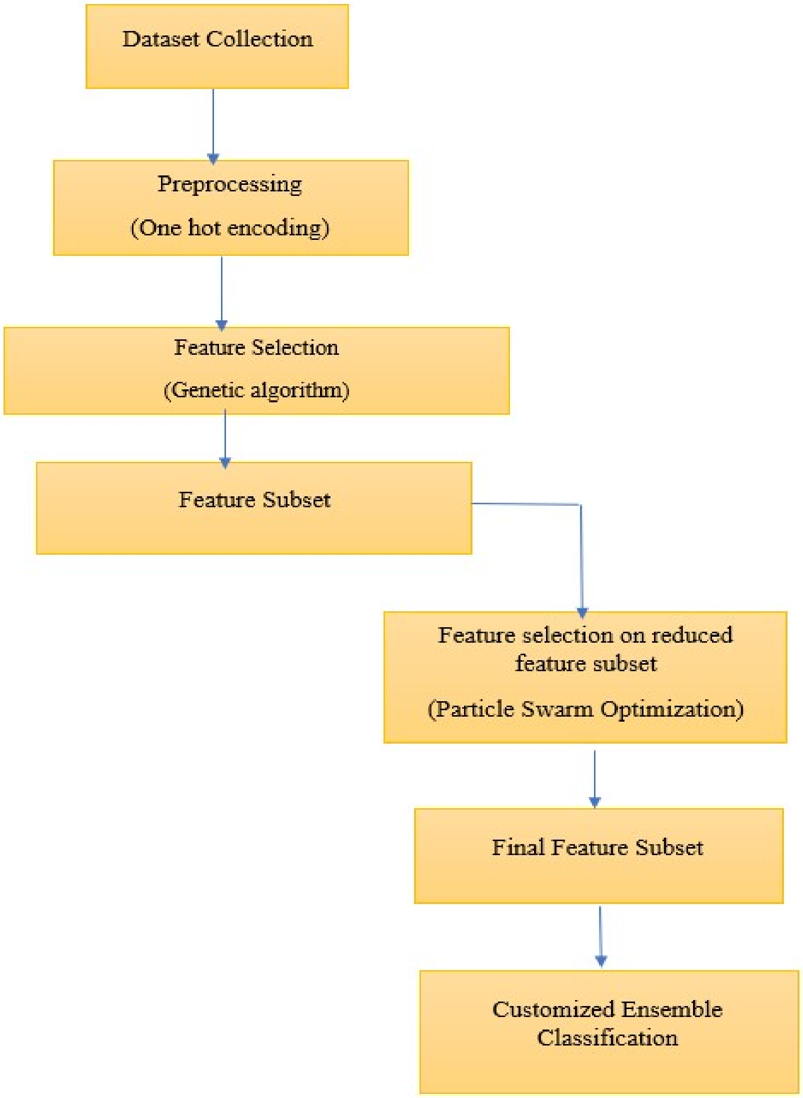


Figure 2: Proposed methodology

On the reduced feature subset FS1, the particle swarm optimisation feature selection technique is used. This feature subset is named FS2, the final feature subset generated after two feature selection techniques. The final feature subset FS2 has 9 features in total.

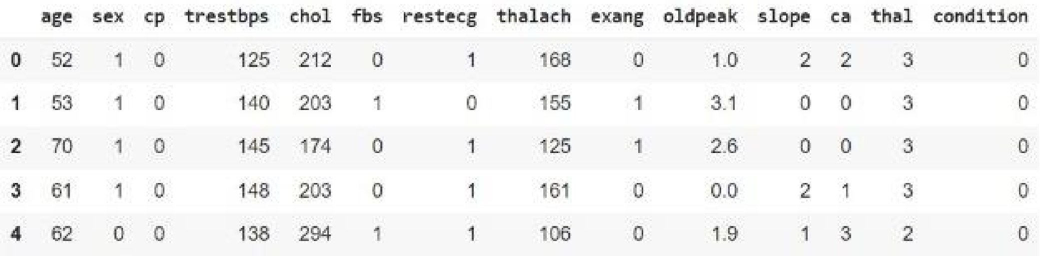
This paper involves the mixture of GA and PSO to avoid local optima problem in particle swarm optimisation technique. Local optima problem restricts the search space, so to increase the capability of the searching process, this paper combines PSO with a genetic algorithm.

After the final feature subset FS2 is generated, the classification process starts, which is the final step in the proposed methodology. Here Customised ensemble classification process is used. Generally, the word ensemble indicates the use of more than one classifier. Here multiple classifiers are used to perform a customised ensemble classification process: support vector machine (SVM), logistic regression and decision tree. After a customised classification process, this paper encountered an accuracy of 98%.

3.1 Dataset description

The Cleveland dataset is the dataset that initially came from the UCI repository. The dataset aims to make a prediction based on specific dataset properties. The dataset consists of 1024 entries with 13 distinct attributes and the class label target. The binary value of the class label in the dataset only comprises 0 and 1. Age, sex, gender, and other factors are among the 13 different qualities.

The top five records in the dataset:



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	condition
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Figure 3: Records of the dataset

3.2 Preprocessing using One hot encoding technique

The current study uses one hot encoding approach to clean the dataset. Because some machine learning models cannot work with categorical data, categorical data needs to be converted to numerical data before it can be put into a machine learning model. Consider a dataset having a Gender column that includes categorical items like Male and Female. Since the input includes string labels, machine learning models need a more precise preference order for these labels for a hierarchical structure. One strategy for resolving this issue is label encoding, in which we give these labels a number value, such as Male and Female being mapped to 0 and 1. To deal with such types of issues, one hot encoding is used.

3.3 Genetic Algorithm as Feature Selection Technique

Figure 4 describes the genetic algorithm. A genetic algorithm is an adaptable heuristic search. "Using a method based on evolution, genetic algorithms choose the best set. Generating a population based on subsets of the potential features is the first step in the feature selection process. A predictive model for the intended task is used to assess the subgroups from this population.

Like a genetic algorithm, a meta-heuristic would improve the search's objectivity and can choose features independently. However, as several potential combinations are inside the solution space, whether an ideal solution might be found is still being determined when the number of attributes is enormous.

The five steps are as follows, and they are primarily used to tackle complex optimisation problems:

- Initialisation
- Fitness Assignment
- Selection
- Reproduction
- Termination

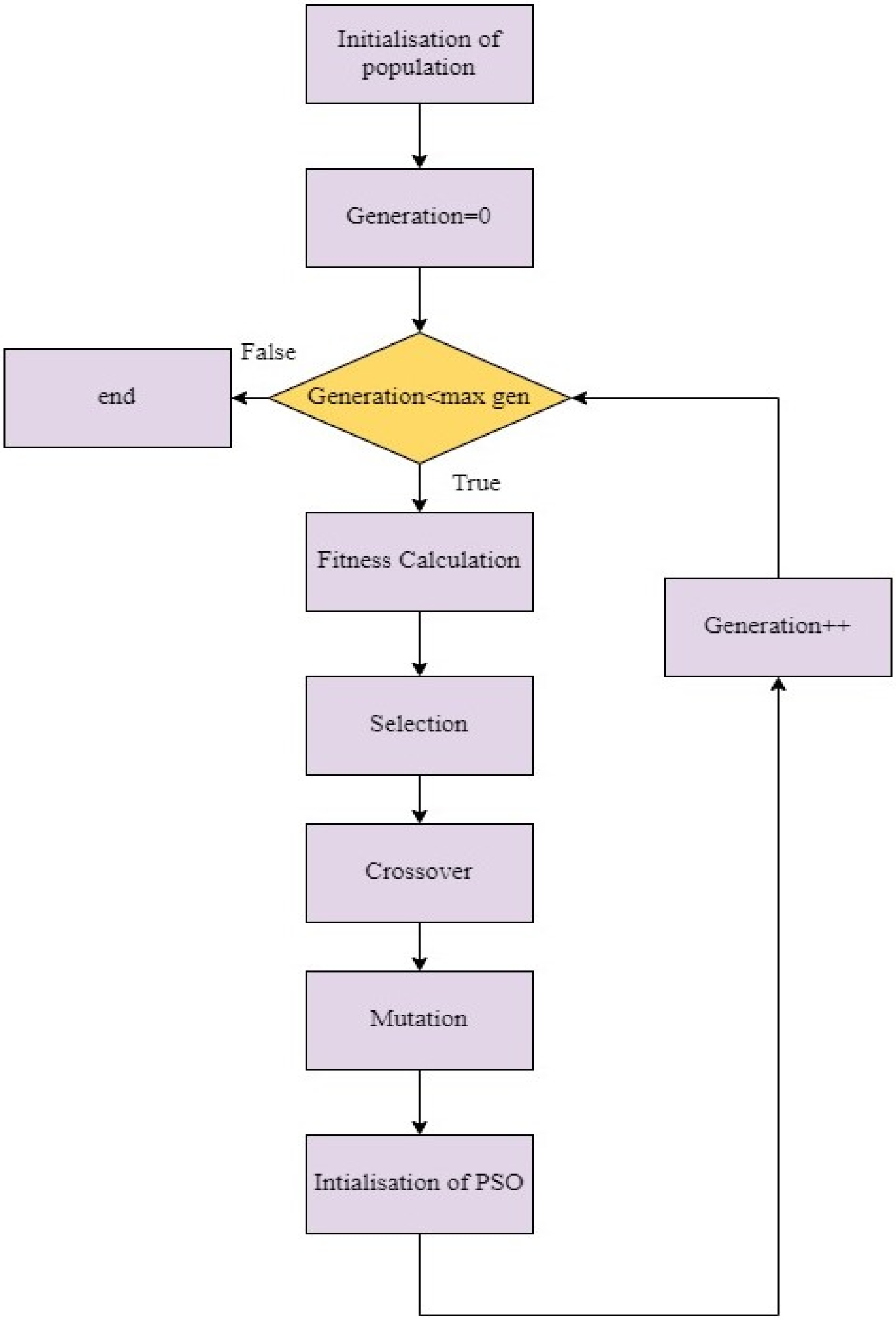


Figure 4: Genetic algorithm

3.4 Particle Swarm Optimisation as Feature Selection Technique on Reduced Feature Subset

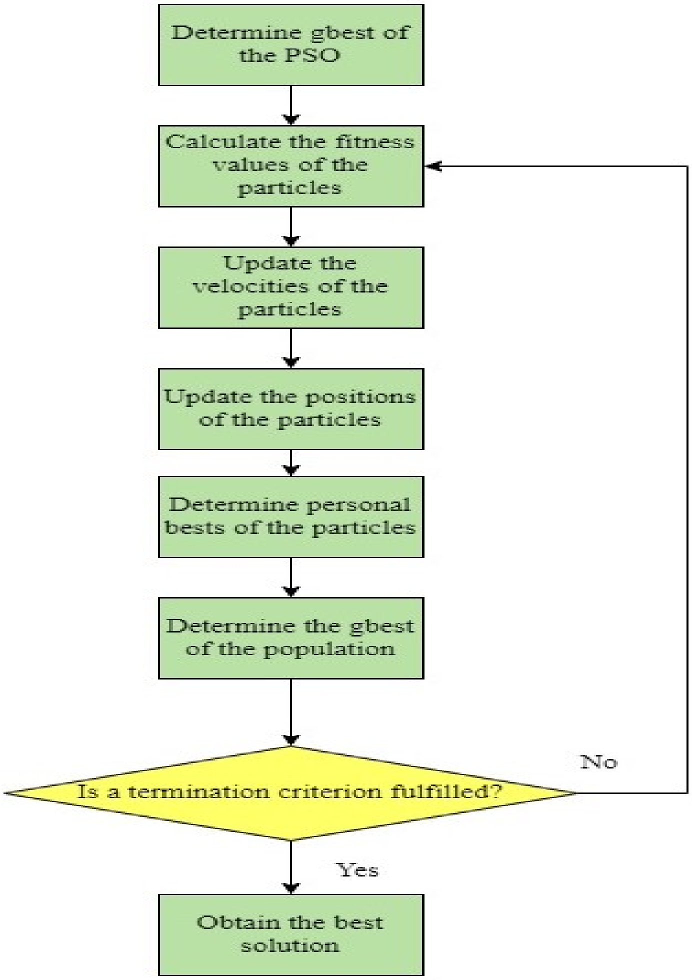


Figure 5: PSO

Figure 5 discusses another optimisation method based on the behavioural analysis of animals and birds, particle swarm optimisation (PSO). Consequently, a feature subset is used in this paper to suggest PSO-based selection since it performs better than Genetic Algorithm.

A metaheuristic PSO may search huge spaces of candidate solutions and make little to no assumptions about the problem being optimised. The PSO parameter selection can significantly influence the performance of the optimisation. Therefore, there has been much research on choosing PSO parameters that produce good performance. After applying PSO on the reduced feature subset, only 9 features are in the final feature subset.

3.5 Customised Ensemble Classification

Ensemble learning raises the efficacy of machine learning by integrating many models. With this procedure, forecasting accuracy can be boosted beyond that of a single model. The basic concept is to learn a group of classifiers (experts) and give them voting rights.

In customized ensemble classification, this paper mixes the following techniques:

3.5.1 Support Vector Machine

A supervised classification method that uses a line to differentiate between two groups. Another term for the SVM is the support vector network.An SVM creates a learning model that divides up brand-new examples into different groups. SVMs are referred to as non-probabilistic, binary linear classifiers based on these functions. An SVM needs labelled training data, just like other supervised learning machines. Materials are labelled in groups for classification. SVM training materials are organised into distinct groups and categorised individually at various points in space. SVMs may do unsupervised learning after understanding a large number of training instances. The algorithms will work to optimise the boundary around the hyperplane and maintain equality between the two sides in order to obtain the optimal data separation.

Following are the types of SVM:

- **Linear or simple SVM**: When categorising data that can be linearly separated, a linear SVM is utilised. This indicates that a dataset is considered to be linearly distinct or separable when it can be divided into categories or classes with the use of a single straight line. Additionally, the classification algorithm that uses such data is known as a linear SVM classifier. Problems involving classification and regression analysis are frequently solved with a simple SVM.
- **Non-Linear or Kernel SVM**: A kernel or non-linear SVM is used to classify non-linear data that cannot be divided into multiple groups with the aid of a straight line. Instead of relying just on 2D space, classification can be carried out with a non-linear data type by inserting features into higher dimensions. Here, freshly introduced features are in line with a hyperplane that makes it simple to divide classes or groups. Multiple variable optimisation problems are often handled by kernel SVMs.

3.5.2 Decision tree

It is one of the popular tools used in prediction and classification. It is a branch structure with an internal node that denotes a test on an attribute; every branch resembles the output of the test, and the leaf node denotes the target label.By using a greedy search to find the ideal split points inside a tree, decision tree learning uses a divide and conquer technique. When most or all of the records have been classified under distinct class labels, this splitting procedure is then repeated in a top-down, recursive fashion. The intricacy of the decision tree plays a significant role in determining whether or not all data points are categorised as homogenous sets. Pure leaf nodes, or data points belonging to a single class, are easier to obtain in smaller trees. It gets harder to preserve this purity as a tree gets bigger, which typically leads to too little data falling under a particular subtree.This is known as data fragmentation, and it frequently results in overfitting. Because of this, decision trees like small trees.

Following are the types of Decision trees:

- **ID3**: Iterative Dichotomiser 3. Entropy and information gain are used by this technique as metrics to assess possible splits.
- **C4.5**: This method is regarded as a subsequent ID3 iteration. To assess split points inside decision trees, it might employ information gain or gain ratios.
- **CART**: Classification and regression trees, or CART. Gini impurity is often used by this algorithm to determine the best characteristic to split on. Gini impurity quantifies the frequency of misclassification for a randomly selected attribute. A lower value is preferable when using Gini impurity to evaluate.

3.5.3 Logistic Regression

Within supervised learning, logistic regression has become one of its most well-known machine learning algorithms. This method uses independent factors to predict the category-dependent variable.When predicting a categorical variable as opposed to a continuous one, logistic regression is used to evaluate the relationship between a dependent variable and one or more independent variables. A categorical variable may have the values true, false, yes, no, 1, 0, etc.

There are three different kinds of categorical response-based logistic regression models:

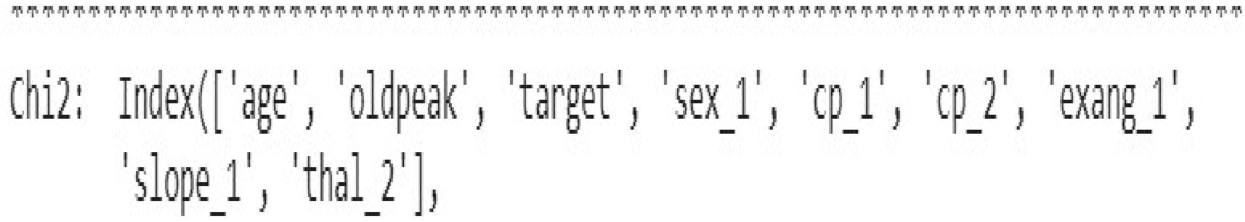
- **Binary Logistic Regression**: In this method, the response or dependent variable is dichotomous, meaning that there are only two possible outcomes (for example, 0 or 1). It is frequently used to determine if an email is spam or not, as well as whether a tumour is malignant or not. This method is most frequently used in logistic regression, and it is also one of the most widely used classifiers for binary classification in general.
- **Multinomial Logistic Regression**: In this kind of logistic regression model, the dependent variable contains three or more potential values, but there is no predetermined order in which they should be ranked. For instance, in order to more successfully sell their films, movie companies aim to forecast the type of film a viewer will likely watch. The studio may find out how much of an impact a person's age, gender, and dating status may have on the genre of movie they enjoy by using a multinomial logistic regression model. The studio can therefore target a particular movie's advertising campaign at an audience that is most likely to go watch it.
- **Ordinal Logistic Regression**: When the response variable includes three or more possible outcomes, however in this situation, these values do have a fixed order, the

ordinal logistic regression model is used. Grading scales from A to F or rating scales from 1 to 5 are two examples of ordinal responses.

4. Results & Discussion:

Prior studies had some restrictions, but this project overcame them. For example, the small dataset makes applying these findings to heart disease easier. Therefore, the dataset here using comprises 1024 samples and 13 characteristics. An increase in pointless features Therefore, this paper uses a genetic algorithm along with PSO. There is a need to use additional classifiers to improve accuracy. To increase accuracy, this paper uses the ensemble method as a classifier. The current model, which combines a Genetic algorithm with a Particle swarm optimisation algorithm used to forecast heart disease more precisely, has an accuracy of 98% compared to earlier publications that are evaluated.

4.1 Generation of Final Feature Subset



```
********************************************************************************

Chi2:  Index(['age', 'oldpeak', 'target', 'sex_1', 'cp_1', 'cp_2', 'exang_1',
       'slope_1', 'thal_2'],
```

Figure 6: Final Feature subset

Figure 6 describes the final feature subset. These features are present in the final feature subset after applying particle swarm optimisation on the reduced feature subset. The reduced feature subset is obtained by applying a genetic algorithm as a feature selection technique. The total features obtained are 9 out of 14 features.

4.2 Iterations vs Accuracy

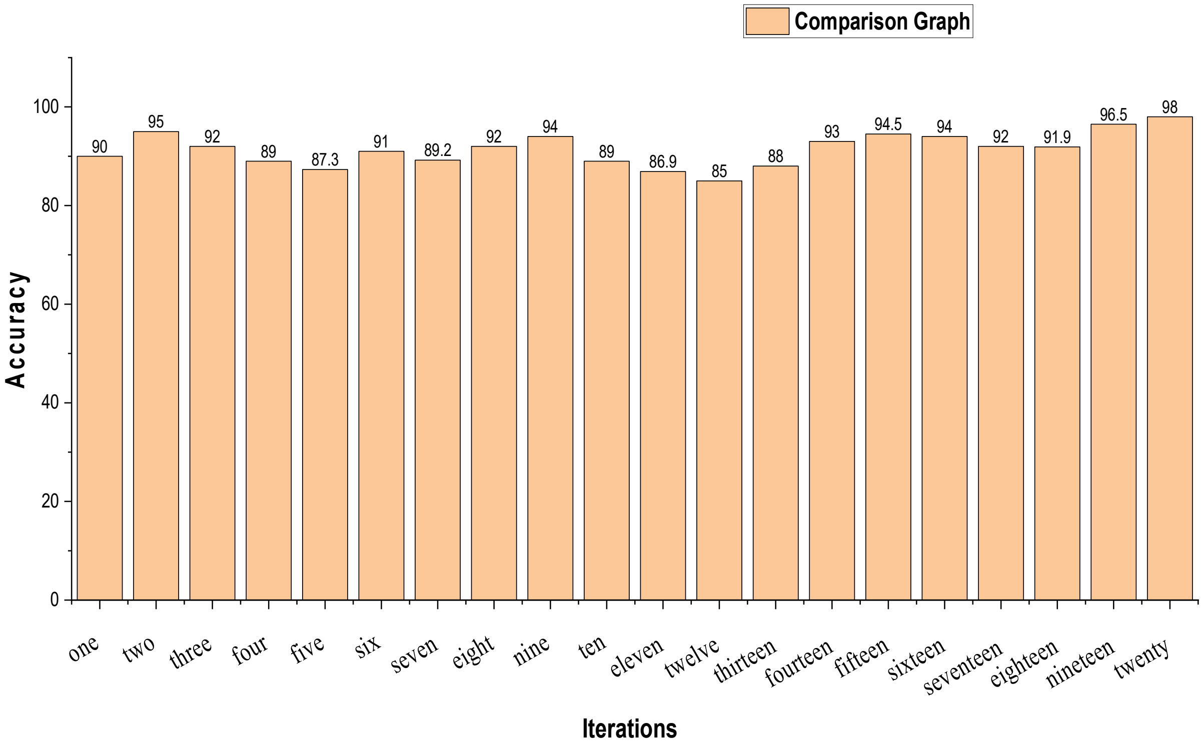


Figure 7: Iterations vs Accuracy

Figure 7 is a graph that discusses the iterations and their accuracies. Iterations are taken on the X-axis of the graph, and on the Y-axis, accuracies are taken. For the sample, this paper presents a graph of only 20 iterations. Out of all the accuracies, 85% occurred on the 12th iteration, which is the least, and 98% occurred on the 20th iteration, which is the highest. So, considering the highest accuracy is the accuracy, 98% is the accuracy of this paper.

4.3 Classification Report

Table 1: Classification Report

	Precision	recall	F1- score	Support
0	0.59	0.99	0.74	196
1	0.41	0.01	0.26	139
accuracy			0.98	335
Macro average	0.29	0.50	0.37	335
Weighted average	0.34	0.59	0.43	335

Table 1 discusses the classification report of the paper. The classification report generally includes precision, recall, f1- score, support, accuracy, macro average and weighted average. These are the performance metrics to measure the performance of a project. All the metrics rely on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values. Since the dataset considered is the binary class label of 0 and 1. 0 indicates no prediction of cardiovascular disease, and 1 indicates the prediction of cardiovascular disease. The accuracy obtained is 98%.

The formulae to calculate performance metrics are as follows

Equation-1: Precision = $\frac{TP}{TP+F}$

Equation-2: F1- score =2* $\frac{Precision*Recall}{Precision+Recal}$

Equation-3: Recall = $\frac{TP}{TP+}$

Equation-4: Accuracy= $\frac{TP+T}{TP+FP+TN+FN}$

4.4 Model vs Accuracy

Table 2: Model vs Accuracy

Authors	**Model**	**Accuracy**
Md G. El- Shafiey	PSO feature selection with Random forest on Cleveland and Statlog dataset	95.6% on Cleveland dataset & 91.4% on Statlog dataset
Wiharto	SVM-GA	94.60%
Senthilkumar	Hybrid Random Forest with Linear Model	88.7%
R. Spencer	Bayes-net Method & chi-square feature selection produced the best model.	85%
Neil Gordan	Genetic algorithm with feature selection	87.8%
Jan Carlo T. Arroyo	GA-ANN	87.3%
Lalit Abualigah	SCAGA	96%

Table 2 discusses the researched papers. Table 2 includes different authors and their models. Each author described a different model, and each model has different accuracy. Among all the models, SCAGA has obtained the highest accuracy of 96%, proposed by Lalit Abualigah. This paper obtained an accuracy of 98% by applying a genetic algorithm with PSO as a feature selection technique which is 2% greater than the accuracy obtained by the SCAGA model.

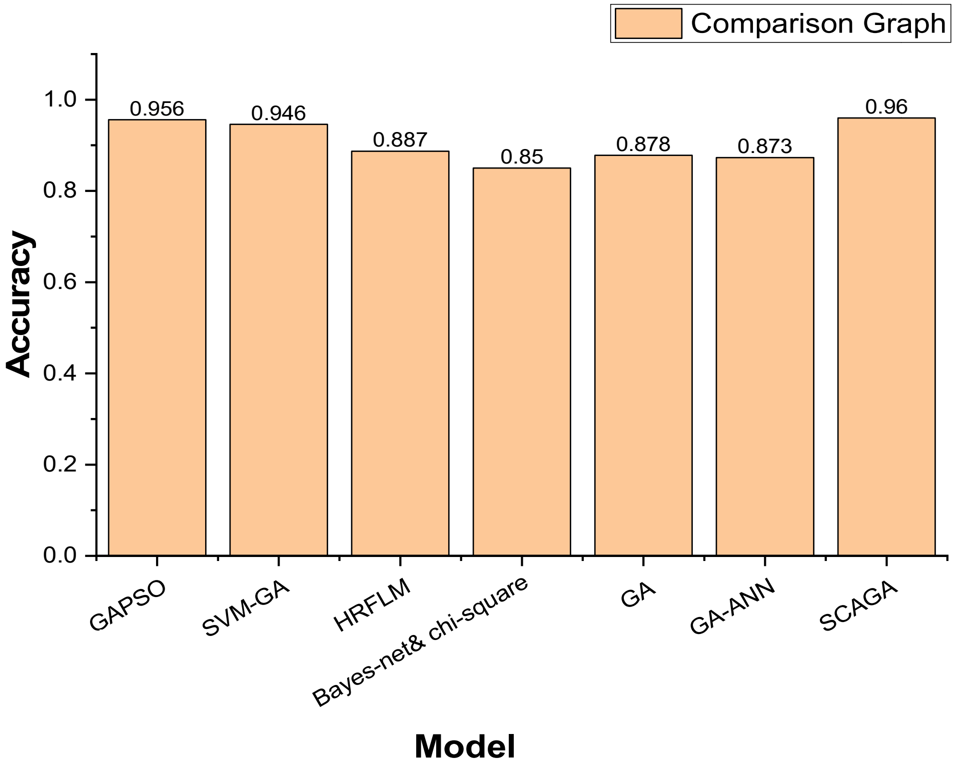


Figure 8: Comparison of accuracies

Figure 8 is a graph that compares the accuracies of previous methodologies. On the X axis, the methodologies of previous papers have been taken. On the Y axis, the accuracies of previous papers have been taken. Each methodology has different accuracy. Among all the researched papers, 96% is the highest accuracy, and its model is SCAGA. This paper obtained an accuracy of 98% which is 2% greater than that of SCAGA, one of the researched papers.

5. Conclusion:

Since heart failure is now a more common cause of death, the model described in this paper primarily focuses on predicting heart disease. Although there are currently methods for utilising machine learning to predict cardiac disease, our work combines feature selection with machine learning algorithms to identify the disease early and reduce patient risks using the genetic algorithm and particle swarm optimisation. To get around the PSO-caused local optima problem, shall employ feature selection. This feature selection aids in obtaining more ideal attributes that help in improved optimisation to more precisely predict cardiac disease. Using this forecast, medical professionals can diagnose the patients early, utilising the optimal features of heart disease prediction information.

Funding

Please add: "This research received no external funding"

Data Availability Statement

In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. You might choose to exclude this statement if the study did not report any data

Conflicts of Interest

Declare conflicts of interest or state "The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results"

References

[1]. Mohamed G. El-Shafiey, Ahmed Hagag, El-Sayed A. El-Dahshan, Manal A.Ismail, A hybrid GAand PSO optimised approach for heart-disease prediction based on random forest, Springer, 2022, https://doi.org/10.1007/s11042-022-12425-x

[2]. PRONAB GHOSH, SAMI AZAM, MIRJAM JONKMAN, (Member, IEEE), ASIF KARIM, F. M. JAVED MEHEDI SHAMRAT, EVA IGNATIOUS, SHAHANA SHULTANA, ABHIJITH REDDYBEERAVOLU, FRISO DE BOER, Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques, IEEE, vol.9, 2021

[3]. Wiharto, Esti Suryani, Sigit Setyawan, Hybrid Feature Selection Method Based on Genetic Algorithm for the Diagnosis of Coronary Heart Disease, JICCE, 2022, https://doi.org/10.6109/jicce.2022.20.1.31

[4]. Senthilkumar Mohan, chandrasegar thirumalai, gautam Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE

Access,volume7,2019,pgno-81542, Digital Object Identifier 10.1109/ACCESS.2019.2923707

[5]. Robinson Spencer, Fadi Thabtah, Neda Abdelhamid, Michael Thompson, exploring feature selection and classification methods for predicting heart disease, Sage, 2020, 10.1177/2055207620914777

[6]. Neil Gordon, Chandrasekhar Kambhapati and Asmaa Alabed, Addressing Optimisation Challenge For Datasets With Many Variables, Using Genetic Algorithms To Implement Feature Selection, IntechOpen, 2022, 1-21, https://doi.org/10.5772/acrt.01

[7]. Kamilia Menghour, Labiba Souici-Meslati, Hybrid ACO-PSO Based Approaches for Feature Selection, INASS, Vol.9, Issue No.3, 2016, DOI:10.22266/ijies2016.0930.07

[8]. Youness Khourdifi, Mohammad Bahaj, Heart disease prediction and classification using machine learning algorithms optimised by PSO and Ant colony Optimisation, ResearchGate, Vol.12, Issue No.1, 2019, https://doi.org/10.22266/IJIES2019.0228.24

[9]. Jan Carlo T. Arroyo, Allemar Jhone P. Delima, An Optimised Neural Network Using Genetic Algorithm for Cardiovascular Disease Prediction, Journal of Advances in Information Technology, Vol. 13, Issue No.1,2022, http://dx.doi.org/10.12720/jait.13.1.95-99

[10]. Laith Abualigah, Akram Jamal Dulaimi, A novel feature selection for data mining tasks using hybrid cosine algorithm and genetic algorithm, Springer Nature, 2021, https://doi.org/10.1007/s10586-021- 03254-y(0123456789().,-volV)

[11]. P. S. Silpa et al., "Designing of Augmented Breast Cancer Data using Enhanced Firefly Algorithm," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), 2022, pp. 759-767, doi: 10.1109/ICOSEC54921.2022.9951883.

[12]. Famila, S., Arulmurugan, A., Mahendar, A., Kalaiyarasan, R., Supriya, N. and Maridu, B., 2024. Transfer learning based feature extraction with metaheuristic optimization algorithm for detecting gastric cancer using optoelectronic sensor in endoscope. Optical and Quantum Electronics, 56(4), p.509.

[13]. Mahendar, A., and Dr K. Shahu Chatrapati. "Detection and Prevention of Cyber Attacks on Cloud-Based Data Centers using Machine Learning." International Journal of Computing and Digital Systems 12.1 (2022): 1063-1070.