



## COPY RIGHT



**ELSEVIER**  
**SSRN**

**2023 IJIEMR.** Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 18<sup>th</sup> Jul 2023. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 07](http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 07)

**10.48047/IJIEMR/V12/ISSUE 07/07**

Title **AI-based patient features analysis through selection learning**

Volume 12, ISSUE 07, Pages: 61-68

Paper Authors **Boda Mahesh, A.Raju, A.Praveen**



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

## AI-based patient features analysis through selection learning

**Boda Mahesh<sup>1</sup>, A.Raju<sup>2</sup>, A.Praveen<sup>3</sup>**

Assistant Professor, Department of CSE, Balaji Institute of Technology and Science, Warangal

[boda.mahesh6@gmail.com](mailto:boda.mahesh6@gmail.com)<sup>1</sup>, [rajucse531@gmail.com](mailto:rajucse531@gmail.com)<sup>2</sup>, [praveen28wgl@gmail.com](mailto:praveen28wgl@gmail.com)<sup>3</sup>

### Abstract

Artificial intelligence has significantly benefited the development of medical informatics and biomedicine by making some approaches available for application, ranging from rule-based expert systems and fuzzy logic to neural networks and evolutionary algorithms. Evolutionary techniques are well known for handling nonlinear restricted optimization problems. After a few generations, the population typically converges to an area around the global optimum, thanks to the exploratory power of evolution-based optimizers. Although the search space can be effectively reduced by this convergence, most of the current optimization techniques continue the search over the original space, losing a significant amount of time on searching unsuccessful regions. One of the most important unresolved problems in pattern recognition, feature selection is an NP-hard problem from the perspective of algorithm creation. We propose a novel evolutionary-incremental feature selection method in this study. The suggested method allows for the application of a standard evolutionary algorithm (EA), such as a genetic algorithm (GA). To make typical EAs compatible with solutions that can have a variable length, this framework suggests certain general adjustments. The solutions about the primary generations in this framework are concise. The length of solutions can then be steadily expanded through generations. The primary components of the suggested solutions are (a) regular recording and monitoring of the patient's vital sign measures taken at home and (b) management of the patient's electronic medical records (EMRs). It aids in the pursuit of superior patient monitoring at home and prompt emergency response.

**Key words**—Artificial intelligence; feature selection; medical conditions; learning method; genetic algorithms

### Introduction

Accurately predicting cardiac illness has the potential to save lives, but a wrong diagnosis can be fatal. Numerous risk features, such as excessive cholesterol, obesity, raised triglyceride levels, and hypertension, increase the likelihood of developing heart disease. Heart failure happens when the heart's muscles are unable to pump blood as efficiently as they should [1]. Breathing problems could be brought on by lung blood clots. The heart gradually weakens or stiffens over time as a result of some cardiac disorders, such as narrowing of the heart's arteries or high blood pressure. If they receive the appropriate care, people with heart disease may live longer. Reduced alcohol use, quitting smoking, a low-fat, low-sodium diet, 30 minutes of moderate

exercise five days a week (or more), and other risk features can all help. Your doctor may also prescribe medicines if lifestyle changes aren't enough to keep your cardiac disease under control. If you have heart disease, the medication you take will be connected to your specific needs. If medication is ineffective for you, your doctor may advise a certain therapy or surgery. Depending on the type of heart illness and the degree of cardiac damage, a certain surgery will be performed. Age and a family history of early coronary artery disease are three well-established risk features for the genesis of heart disease in men/women [1]. Conventional features are autonomous, unstructured features that reflect the fundamental data. First, these techniques provide equal weight to the features and the

feature analysis and then combine them to create the source feature aggregation. Each analysis of the feature is converted via encoding to a hidden state vector that acts as the context representation. Finally, using a decoder based on context representation, features are created from a predetermined feature [9]. However, there are still considerable disadvantages to these previous methods. They don't consider if the features are present in the natural data; instead, they just extract them from the normal feature data, which results in the creation of features that do not appear in the feature analysis.

The other sections of this paper are organized as follows: Section 2 introduces methodologies for evaluating feature analysis and its risk and Section 3 provides an analysis of proposed feature analysis. The experiment and analysis of the data are presented in Section 4, and the proposed work is concluded in Section 5.

### Previous Works:

Feature selection is the process of describing the data by picking just a few of its original features [5]. The selection of features plays a crucial role, especially in applications with a lot of unnecessary functionality. The selection of features is reduced to a search problem given a criterion function [4, 6]. When there are too many features, an exhaustive search is impractical; instead, a heuristic search might be used. Sequential forward and/or backward selection is one of the algorithms that have demonstrated success in real-world settings [7, 8]. But none of them can ensure optimality in any way. Utilizing feature weighting, which gives each feature a real-value number to signify its relevance, can solve this issue in medical analysis [6]. An incorrect diagnosis can be catastrophic, yet accurately forecasting cardiac sickness has the potential to save lives. Heart disease is more likely to occur when several risk features, including high cholesterol, obesity, elevated triglyceride levels, and hypertension, are present. Heart failure happens when the heart's muscles are unable to pump blood as efficiently as they should [1]. Currently,

many feature extra descriptor methods can be broadly divided into supervised and unsupervised methods [5]. The TextRank-based graph ranking algorithm was introduced [2] to find the relationships between probable features for unsupervised extra descriptor techniques. [11] used the Key Cluster method of clustering to find representative diagnoses from feature groupings. Using supervised machine learning methods, the feature extra descriptor task is transformed into a binary classification issue [7]. The Naive Bayes model (NB) is used to determine whether the patient diagnoses are features in the approach, which was created by [4]. It involves extracting patient features from diagnosis, calculating attributes like the feature value and location information for each patient phrase, and then using the NB to determine whether the patient diagnosis is a feature, through the C4.5 decision tree training classifier [12]. Support Vector Machine (SVM) was utilized by [13] to extract term usage features, such as diagnose frequency and diagnose location data. Conditional Random Field (CRF) was utilized by [14] to extract features from reports. The incremental method is one of the most popular and straightforward techniques for feature selection. There are two categories of incremental feature selection: forward selection and reverse elimination [3]. Online feature selection on a stream of data can benefit from incremental methods or modified versions of such methods [4]. Even though incremental feature selection approaches are straightforward and helpful in some unique situations. A probabilistic-incremental strategy for feature selection was given in [5] as a solution to this problem. This technique chose features using the Las Vegas algorithm probabilistically. This method can carry out feature selection more effectively, according to the results that were reported. EAs are appropriate approaches for handling such problems because the feature selection procedure is computationally quite difficult. In a short amount of time, EAs can find less-than-ideal solutions. GA [6] is one of the most well-liked and frequently utilized EAs. The most popular method of employing GA for



feature selection defines chromosomes as a bit string of length  $F$ . In this instance, a bit in the bit string represents each feature's presence. If a bit is set to 1, the associated feature is chosen; if not, it is not chosen. Numerous mobile applications for cardiology have been created to improve the research experiences of medical professionals and students [15–17]. Examples include (a) applications that show a 3D model of the human heart and let users view it from any angle, (b) calculators with frequently used formulas in cardiovascular medicine, (c) electrocardiography (ECG) guides with examples of various types of ECG and (d) tools for clinical practice and diagnosis. All of the aforementioned initiatives are targeted at medical professionals, primarily for instructional purposes and easy access to information for physicians. Furthermore, people who have heart issues shouldn't utilize them. In recent years, artificial intelligence (AI) has replaced both invasive and noninvasive diagnostic techniques to become the most promising technology for the identification of heart diseases. In [4], support vector machines (SVM) and multilayered perceptron neural networks (MLPNN) are evaluated for the identification of coronary artery disease (CAD) based on exercise stress testing (EST) data. The best strategy for predicting outcomes for postoperative cardiac patients is neural networks, according to [13]. To support non-specialist staff members' clinical judgment in the assessment of patients with heart failure, an AI-based Computer-Aided Diagnosis system is developed [16]. The system determines the patient's pathogenic condition and highlights the use of four AI-based techniques to discover potential irritants: neural networks, support vector machines, decision trees, and fuzzy expert systems with genetically derived rules. Neural networks produced the greatest results, with an 86% accuracy rate. It is a difficult and complex task to automatically analyze cardiology tests, and several expert systems, neural networks, and case-based reasoning techniques have been developed in this area [17].

## 1. Proposed Work:

There is a vast amount of patient data in the medical industry. It is necessary to mine this data using various machine-learning techniques. To make accurate diagnostic conclusions, healthcare professionals analyze this data. Clinical aid can be provided by analyzing medical data with categorization algorithms. But nobody has developed a correlation equation that can be used to pinpoint the patients specifically. Previous studies have used medical information for classification purposes to help identify the patient's precise condition. To detect cardiac patients early on without angiography and CT angiography, this study used basic data as well as some relevant clinical data. The primary contributions of this research are as follows: Artificial neural network (ANN) and curve-fitting techniques were used to develop the correlation. The identification of cardiac patients can be done by experts using a model of an artificial neural network (ANN). Results from an ANN-based model are extremely accurate. The difficulties and parameters connected to cardiac disease have been identified for testing and identification purposes after a thorough discussion of the condition and a selective literature review.

The Kaggle database has been used to collect the data. The models' performances have been contrasted. The findings indicate that these linkages can make it simpler and more accurate to identify cardiac patients.

We suggest a preprocessing technique to enhance the Principal's performance in two parts to Principal Component Analysis (PCA) for classification issues; the first step involves calculating the weight of each feature using a feature weighting algorithm.

The features that have weights greater than a certain threshold are then chosen. The second phase is then applied to the chosen feature attributes.

The second phase involves changing feature variances until they match the significance of the features. We analyzed that the performance of PCA will improve

in classification issues by utilizing the benefit of step 2 to disclose the class structure.

One of the best techniques for evaluating the quality of features is shown in Algorithm 1.

**Algorithm 1:**

Step 1: Initialization: Feature Descriptor  $F_D$  i.e.,  $F_D = \{(x_i, y_i)\}_{i=1}^N$ ,  $y$  is the feature class labels described from 1 to  $f_c$ ,  $f_c$  is the number of feature classes;

Step 2: Set Feature Weights  $F_{Wj}$  i.e.,  $F_{Wj}=0$ , as the number of iterations takes place the  $F_W$  increases slowly with  $T$  which is the number of iterations varies as  $0 \leq j \leq t$ ;

Step 3:

loop for  $F_L = 1$  to  $T$ , where  $F_L$  is feature-length

select a feature pattern  $F_{Px}$  from  $F_D$  with feature class  $f_c$ ;

find  $k$  closest feature fits  $F_f$  from feature class  $f_c$  chosen;

for each feature class  $f_c$ , find identical feature class  $f_i$ , where  $f_i \neq f_c$ ;

from feature class  $f_i$ , find  $k$  closest non-feature patterns  $F_{NPx}(f_i)$ ;

for  $j = 1$  to  $t$

compute Feature Nearest Weights  $F_{NWj}$  i.e.,

$$F_{NW_j} = \sum_{i=1}^k |x_{cj} - F_{fj}| / T * k + \sum_{i=1}^k |x_{cj} - F_{NPx}(f_i)| / T * k$$

end

end

In Algorithm 1, the main purpose is to calculate the weights of features repeatedly based on how well values can differentiate between similar cases. The original work deals with the entire data and divides problems into two kinds [12].

In particular, it lacks the means to get rid of redundant features [13]. Algorithm 1 employs a strategy based on K-nearest-neighbor algorithms as proposed in Algorithm 1, which uses the 1-nearest-neighbor method. The pseudocode for this algorithm is shown in the pseudocode of Algorithm 1. The idea of distributed model learning, which enables the agent to learn and store the details of the final model, is a crucial component of this method design employing two modules as opposed to one. It is insufficient to try to encode the model of a complex learning problem in a single structure. The quantity of data needed to store the model must be split among multiple modules. The descriptor type feature and the parameter feature are added to the existing module used to store the final model in the proposed design.

A third module, the assembler, is utilized to combine one descriptor type and one vector of parameters into one executable descriptor during the learning phase.

The following are the roles that each of the three modules plays.

Descriptor-type features:

- (i) get knowledge from the scalar reward ( $r$ ), where  $r$  is the descriptor type should the agent carry out in the upcoming time stage,
- (ii) The temporal difference error TD instructs the parameter feature, here the parameter vector  $p$  is most appropriate for the descriptor type that the descriptor type feature has provided.
- (iii) The descriptor to be carried out by the agent is assembled by the assembler using the descriptor type at and the parameter vector  $p$ . Feature learning is implemented using descriptor type features [8].
- (iv) Finding the ideal descriptor type for each circumstance is the main goal of this module.
- (v) The learning process in this module is quick and reliable because there are a limited amount of activities, even when the states are continuous vectors. The fundamental standards

applied in each of the three features' decision mechanisms are listed in Table 1.

**Table 1: Patient Feature and Learning Decision**

Set	Patient feature	Fuzzy Class	Learning Decision
A	Heart Failure	<ul style="list-style-type: none"> <li>• Class variable</li> <li>• Age</li> <li>• Duration of Rate Control</li> </ul>	<ul style="list-style-type: none"> <li>• Activity is limited for the duration</li> </ul>
B	Heart disorders	<ul style="list-style-type: none"> <li>• Duration of Rate Control</li> <li>• Aortic root dilatation</li> </ul>	<ul style="list-style-type: none"> <li>• Limited Medication</li> </ul>
C	Hypertension	<ul style="list-style-type: none"> <li>• Age</li> <li>• Blood Pressure</li> </ul>	<ul style="list-style-type: none"> <li>• Limited Medication</li> </ul>

Other criteria are better used as distributed variables rather than being separated into fuzzy variables. The choice may involve beginning, stopping, or adjusting a pharmaceutical regimen, activity restriction, or surgical intervention. The features are intended for use by general practitioners, cardiologists, and other medical professionals. Although general practitioners may find features to be of great use when called upon to care for a cardiologic patient, cardiologists gain greatly from them. Our objective for feature selection in heart identification is to select a few Gabor wavelets to get the best level of accuracy. This paper's method is comparable to the [14]. The main differences between the suggested strategy and the [14] methods are as follows: We employ a set of fixed locations rather than the adjustable locations used

by the proposed work to extract Gabor features from distinct hearts. We choose the optimum locations to extract features using an EA as opposed to the proposed work, which chooses the areas to extract Gabor features through a matching approach for each heart feature. In contrast to the proposed work, which extracts a particular set of Gabor coefficients at each position, we simply extract one Gabor coefficient per feature location. The parameters of the applied Gabor wavelets at each location are selected from an EA. Four parameters make up a Gabor wavelet, as mentioned in algorithm 1. If a feature picture has a resolution of 512x512 pixels and has 256 potential values for and, then 10000 distinct Gabor wavelets can be used to extract 10000 different features from each feature image.

## 2. Comparative Results and Discussions:

A feature of about 200 heart classes from the clinical database was utilized to generate the experimental results. Set A of 100 heart classes was used to choose the features, while Set B of 100 heart classes was utilized to assess the selected features, and Set C of 100 heart classes was utilized to analyze the selected features. Two 72 x 58-pixel frontal heart photos are included in each heart class. The features on these feature hearts have been standardized in terms of size, rotation, and location. In the tests, the fitness function of each prospective solution is compared against the effectiveness of heart recognition using the selected features. Therefore, employing the selected features is necessary to train and test the feature recognition system. Set A: For this, one training image and one test image from each class were used. The fitness value for a subset of features on set A is therefore equal to the accuracy determined using that subset of features. Upon completion of the EA, the best features are selected by the most successful feedback from the most recent feature. To determine the quality of this feature set, it is evaluated on set B. Depending on how well set B's heart

recognition performs, we can determine whether the final feature is over-fitted. In other diagnoses, if the final features yield more accuracy on set C, they are closer to the ideal features. Table 1 shows the specification of the final optimal attributes using the method we recommend. The number of features that were chosen as well as the effectiveness of heart recognition on sets A, B, and C are displayed in Table 1.

**Table 2: Comparative analysis of feature selection learning model with previous works**

Model / Parameters	Incremental GA	Variable Length GA	Incremental IWO	Proposed Work
No. of selected features	40	90	25	55
Accuracy of features on Set A (%)	98	98	98	98
Accuracy of features on Set B (%)	92	89	93	92
Accuracy of features on Set C (%)	93	91	94	94

Table 2 displays the specification of the ultimate best features utilizing variable length GA. The amount of features that were chosen as well as the effectiveness of heart recognition on sets A and B are displayed in the table below. Table 2 displays the specification of the top choice in the most recent iteration of the suggested evolutionary-incremental IWO-based feature selection method. Table 2 displays the accuracy values for each set. The table shows that all of the generation models outperform the conventional baseline approach. The proposed work uses a decoder with feature-aware capability and an attention mechanism to aggregate the feature information for each diagnosis in the diagnosis content. With the proposed work, the ability to produce learning-sensitive features has been improved.

**Conclusions and Future Scope:**

Datasets from public clinical are compared. Table 2 shows that all of the generation models outperform the conventional baseline approach. Furthermore, it should be noted that the model proposed in this research greatly outperforms the baseline model in both datasets. The feature's descriptor has a guiding influence on features; the feature's semantic content is not taken into account. We offer a feature descriptor methodology that combines feature analysis and patient health diagnosis. It uses feature semantics as a guide for encoding feature terms. A feature-aware decoder is employed during the decoding process. As a result, the proposed model generates features of higher feature descriptors. The experimental results show that the better feature information found in the feature's latent feature may be more efficiently used by the proposed model in this study. The proposed model is improved compared to other methods and can generate features with excellent accuracy in terms of feature prediction. The relative effectiveness of the suggested evolutionary-incremental framework for feature selection was assessed in the tests. The suggested method can be used for other optimization issues, such as clustering, which has solutions of varying lengths.



## References:

1. N. Tasnim and S. A. Mamun, "Comparative Performance Analysis of Feature Selection for Mortality Prediction in ICU with Explainable Artificial Intelligence," 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ECCE57851.2023.10101553.
2. R. Chandra, M. Kapil and A. Sharma, "Comparative Analysis of Machine Learning Techniques with Principal Component Analysis on Kidney and Heart Disease," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021, pp. 1965-1973, doi: 10.1109/ICESC51422.2021.9533011.
3. M. H. Bin Abdul Ghafar, N. Aleena Binti Abdullah, A. H. Abdul Razak, M. Syahirul Amin Bin Megat Ali and S. A. Mutalib Al-Junid, "Chronic Kidney Disease Prediction based on Data Mining Method and Support Vector Machine," 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia, 2022, pp. 262-267, doi: 10.1109/ICSPC55597.2022.10001806.
4. F. Tasnim and S. U. Habiba, "A Comparative Study on Heart Disease Prediction Using Data Mining Techniques and Feature Selection," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), DHAKA, Bangladesh, 2021, pp. 338-341, doi: 10.1109/ICREST51555.2021.9331158.
5. T. Amarbayasgalan, V. -H. Pham, N. Theera-Umpon, Y. Piao and K. H. Ryu, "An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well-Ordered Training Datasets," in IEEE Access, vol. 9, pp. 135210-135223, 2021, doi: 10.1109/ACCESS.2021.3116974.
6. G. Tallapureddy and D. Radha, "Analysis of Ensemble of Machine Learning Algorithms for Detection of Parkinson's Disease," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 354-361, doi: 10.1109/ICAAIC53929.2022.9793048.
7. A. Deb, M. S. Akter Koli, S. B. Akter and A. A. Chowdhury, "An Outcome Based Analysis on Heart Disease Prediction using Machine Learning Algorithms and Data Mining Approaches," 2022 IEEE World AI IoT Congress (AIoT), Seattle, WA, USA, 2022, pp. 01-07, doi: 10.1109/AIoT54504.2022.9817194.
8. K. V. V. Reddy, I. Elamvazuthi, A. A. Aziz, S. Paramasivam and H. N. Chua, "Heart Disease Risk Prediction using Machine Learning with Principal Component Analysis," 2020 8th International Conference on Intelligent and Advanced Systems (ICIAS), Kuching, Malaysia, 2021, pp. 1-6, doi: 10.1109/ICIAS49414.2021.9642676.
9. K. M. Abiodun, J. B. Awotunde, D. R. Aremu, and E. A. Adeniyi, "Explainable AI for fighting COVID-19 pandemic: Opportunities, challenges, and future prospects," in Computational Intelligence for COVID-19 and Future Pandemics. Singapore: Springer, 2022, pp. 315-332.
10. A. Anand, T. Kadian, M. K. Shetty, and A. Gupta, "Explainable AI decision model for ECG data of cardiac disorders," Biomed. Signal Process. Control, vol. 75, May 2022, Art. no. 103584.
11. P. M. Kumar, L. U. Khan, and C. S. Hong, "Affirmative fusion process for improving wearable sensor data availability in artificial intelligence of medical things," IEEE Sensors J., early access, Feb. 22, 2022, doi: 10.1109/JSEN.2022.3153410
12. W. Gong, Y. Wang, M. Zhang, E. Mihankhah, H. Chen, and D. Wang, "A fast anomaly diagnosis approach based on modified CNN and multisensor data fusion," IEEE Trans. Ind. Electron., vol. 69, no. 12, pp. 13636-13646, Dec. 2022, doi: 10.1109/TIE.2021.3135520
13. A. Haleem, M. Javaid, R. P. Singh, and R. Suman, "Applications of artificial intelligence (AI) for cardiology during COVID-19 pandemic," Sustain. Oper. Comput., vol. 2, pp. 71-78, Nov. 2021.
14. A. P. Singh, N. R. Pradhan, A. K. Luhach, S. Agnihotri, N. Z. Jhanjhi, S. Verma, U. Ghosh, and D. S. Roy, "A novel patient-centric architectural framework for blockchain-enabled healthcare applications," IEEE Trans. Ind. Informat., vol. 17, no. 8, pp. 5779-5789, Aug. 2021



15. F. Shamout, T. Zhu, and D. A. Clifton, "Machine learning for clinical outcome prediction," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 116–126, 2021
16. T. L. Rodziewicz, B. Houseman, and J. E. Hipskind, "Medical error reduction and prevention," in *StatPearls [Internet]*. Treasure Island, FL, USA: StatPearls Publishing, Jan. 2021.
17. T. Makino et al., "Differences between human and machine perception in medical diagnosis," *Sci. Rep.*, vol. 12, no. 1, pp. 1–13, 2022.
18. Kumar, V. & Ramana, T.. (2022). Fully scheduled decomposition channel estimation based MIMO-POMA structured LTE. *International Journal of Communication Systems*. 35. 10.1002/dac.4263.
19. V. M. Kumar and T. V. Ramana, "Position-based Fully-Scheduled Precoder Channel Strategy for POMA Structured LTE Network," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-8, doi: 10.1109/ICECCT.2019.8869133.