International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I7P126

Volume 73 Issue 7, 328-338, July 2025

Original Article

Detection of Multiple Optimized Feature Subsets Using Genetic Algorithm for ECG-based Identification

Mamata Pandey[1], Anup Kumar Keshri[2]

[1,2]*Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India.*

[1]*Corresponding Author : m.pandey@bitmesra.ac.in*

Received: 05 March 2025 Revised: 22 May 2025 Accepted: 04 July 2025 Published: 30 July 2025

Abstract - *Selecting the most relevant features remains a crucial challenge in optimizing classification accuracy and computational efficiency, especially in high-dimensional datasets. Most of the approaches select a single subset, claiming it to be the most optimal. However, diverse combinations of features may compete on computational overhead and model selection. Different feature subsets may highlight distinct aspects of the dataset, helping domain experts gain better insights. This paper presents a novel approach for detecting multiple optimized feature subsets using a novel Genetic Algorithm (GA) variant for ECG-based identification. The proposed method employs GA to identify multiple optimal feature combinations, improving both accuracy and robustness. Experimental evaluations on the ECG dataset demonstrate the effectiveness of the approach in selecting optimal feature subsets, leading to enhanced classification performance. The results indicate that the proposed method can significantly improve identification accuracy while reducing feature dimensionality, making it a viable solution for real-world applications.*

Keywords - *Genetic Algorithm, ECG-based identification, Multiobjective optimization, Multi-optima optimization, Niching.*

1. Introduction

Biometric identification has become an essential component in security systems. ECG-based authentication is emerging as a promising alternative due to its inherent physiological uniqueness [1]. Unlike traditional biometric traits such as fingerprints or facial recognition, ECG signals offer the advantage of being difficult to replicate, enhancing security and reliability. However, ECG data's high dimensionality and variability pose challenges in developing efficient identification systems [2]. Feature selection is crucial for high-dimensional data, as redundant and irrelevant features can degrade classification performance and increase computational overhead [3]. Traditional feature selection methods, such as Principal Component Analysis (PCA)[4] and filter-based techniques [5], may not always yield optimal results, as they rely on predefined heuristics and correlation between features [6]. To overcome these limitations, researchers have increasingly explored Evolutionary Algorithms (EA) [7]. The evolutionary approach to feature selection dynamically explores the search space to identify optimal subsets that maximize classification performance [8]. EAs, including Differential Evolution (DE) [9]Genetic Algorithms (GA) [10, 11]. Ant Colony Optimization (ACO)[12], and Particle Swarm Optimization (PSO) [13], have been widely studied in feature selection [14]. They can also address multiple conflicting criteria simultaneously, such as minimizing feature count and maximizing model performance, leading to multiobjective optimization [15]. However, the standard EA typically generates a single optimal subset of features, whereas in real-world scenarios, multiple distinct subsets may satisfy the optimization criteria. Different subsets can reveal unique aspects of the data, enabling domain experts to discover hidden patterns or relationships between features and target variables [16]. In a dynamic environment, where data distribution or context may change over time, multiple feature subsets provide a range of options, allowing for seamless adaptation to new conditions. The identification of multiple subsets offers insights into the problem domain and clarifies the mapping between features and target labels.

Feature selection is often influenced by constraints such as computational efficiency, interpretability, and domain-specific priorities. Multiple subsets allow for the selection of features that align with these specific needs. Analyzing multiple subsets helps to balance trade-offs, such as performance versus computational cost, accuracy versus simplicity, and predictive power versus interpretability. Exploring multiple subsets can also reveal redundancies or interactions among features, potentially improving the model's generalization to unseen data. In situations where optimization involves competing objectives, such as maximizing accuracy while minimizing the number of features, having multiple subsets enables the selection of those that achieve the best balance. Lastly, in scenarios with missing

or corrupted data, alternative subsets serve as backup solutions, ensuring that the model remains reliable and effective. GA is a robust evolutionary algorithm suitable for high-dimensional and complex data [17]. This paper addresses the above-said requirements and overcomes limitations of the standard version by a novel variant of the genetic algorithm designed to detect multiple optimized feature subsets that enhance classification accuracy while minimizing subset size. In contrast with traditional approaches, the proposed algorithm generates multiple subsets that satisfy the optimization criteria. The proposed algorithm has been applied to an ECG dataset to identify individuals based on their ECG features. The contributions of this work are as follows.

1. A novel variant of GA for generating multiple optimized feature subsets from a high-dimensional dataset and its application to the ECG dataset.
2. Identification of critical features for ECG-based identification by analyzing multiple optimized feature subsets
3. Performance analysis evaluation across the ten best resultant optimal feature subsets.

The related research works, methodology, dataset, algorithm, and results are detailed in the subsequent sections of the paper.

2. Related Work

GA is a powerful optimization technique inspired by the process of natural evolution. The algorithm utilizes stochastic search methods to explore multiple solutions simultaneously [18]. GAs can be applied to a wide range of problems, as they work with a population of potential solutions, allowing them to search various areas of the solution space at the same time. It has been applied to several problem domains like heart disease diagnosis [19], heart rate variability [11], ear biometrics [20], ECG signal processing [21], solar radiation estimation [16], biodiesel synthesis [22], intrusion detection [23], phishing URL detection [24] GAs are also useful for parameter tuning of various machine learning models [25, 26]. The selection and crossover operators generate new solutions and update the population with fitter individuals. In mutation, random changes are introduced in offspring, such as flipping a bit in a binary-encoded solution to maintain diversity. This diversity prevents premature convergence to suboptimal solutions [27].

GAs can efficiently search vast, multi-dimensional spaces, making them useful for problems with thousands or even millions of potential solutions. However, GAs do not guarantee finding the absolute best solution. Instead, they are heuristic-based methods, which means they focus on finding satisfactory solutions rather than mathematically provable optimal ones [28]. GA has a high computation time due to slow convergence. To address these issues, various forms of genetic algorithms have been developed, utilizing different selection [29, 30], crossover [27, 31, 32], and mutation operators. Several hybrid approaches of GA with other techniques like particle swarm optimization [13] and fuzzy logic[19, 33] have been applied to various problem domains. GA has been enhanced using several techniques for obtaining multiple optima [34]. Several variants of niching[16, 35] and crowding [36] techniques have been studied for locating multiple optima. These techniques use multiple populations to explore diverse areas of the search space.

Crowding focuses on regulating how offspring replace individuals within the population. Rather than randomly replacing individuals, offspring are positioned to replace those that are most similar to them in terms of fitness or distance. This strategy prevents dominance by highly similar individuals, thereby ensuring a more diverse population. Variants of crowding include deterministic crowding, where offspring directly compete with their parents, and fitness-sharing-based crowding, which takes into account both fitness and similarity. Conversely, niching [37] promotes and preserves subpopulations, or "niches," within the search space, making it particularly effective for multimodal optimization problems. This approach encourages the algorithm to explore different regions of the solution space and maintain multiple optima.

Common niching methods include fitness sharing, where an individual's fitness is reduced based on the number of others within its niche; speciation, which groups individuals into species based on similarity; and clearing, which restricts the number of individuals in a niche by removing excess ones [38, 39]. These techniques are generally utilized on continuous search spaces and real values. They can be effectively used to discover multiple optimized subsets.

Electrocardiogram (ECG) signals are complex, noisy, and high-dimensional [40]. ECG-based biometric identification has gained significant attention due to its uniqueness and resilience against spoofing attacks. ECG data typically includes hundreds of features extracted from waveforms, such as the P wave, QRS complex, and T wave. However, not all features are useful; some may be redundant or noisy [41]. EA can help select the most important features, reducing computational requirements and enhancing classification accuracy [2, 8].

However, most previous works have utilized EAs to generate a single optimum subset, leaving other subsets that may compete significantly. Detection of multiple optimal subsets can facilitate a better understanding of relevance, dependencies, and redundancies in ECG features that are not interpretable in the single optimal solution. GA is a robust EA that has been successfully applied to various problem domains, making it dependable for ECG data. This study presents a novel variant of GA to identify multiple optimal

ECG feature subsets. The proposed GA uses neither a standard crowding nor a niching technique. It handles multiple populations that converge to different optima simultaneously. The solutions may have different fitness values. These solutions are further analyzed to detect critical features for ECG-based identification.

3. Methodology

A Genetic Algorithm (GA) is an optimization technique inspired by the principles of natural selection. It addresses complex problems by evolving solutions over several generations.

The process starts with a population of individuals, also known as candidate solutions or chromosomes, which are represented in a suitable format. Each individual is evaluated using a fitness function to determine its effectiveness. The best solutions are then selected and combined through a process called crossover to create new offspring.

Additionally, mutation introduces small changes to maintain diversity within the population. Over multiple generations, the population gradually evolves toward an optimal or near-optimal solution. The proposed GA uses binary encoding, random selection, uniform crossover and random mutation [18]. The specifics of these are defined as follows.

3.1. Encoding

GA requires a compatible representation of possible solutions into chromosomes. Each chromosome contains a set of genes representing a solution characteristic. In the proposed GA, each feature subset is encoded into a binary string. For n features, there are 2^n subsets. Each subset A is represented by a binary string B of n bits. For each ith bit, B[i]=1 if the ith feature belongs to A, otherwise, B[i]=0.

3.2. Fitness Function

The fitness function is a multiobjective composite function given by equation 1.

$$\text{Maximize } f(A) = \text{CLF}(A).\text{performance} + (n - |A|)/n \quad (1)$$

Where A is a feature subset. The first term of the fitness function is the performance of the classifier CLF when trained and tested with features only in subset A. CLF can be any state-of-the-art classifier, like a Neural Network or Support Vector Machine (SVM). Performance can be measured by metrics such as accuracy, precision, specificity, or F1-score or their combination. The second term represents the number of absent features in the subset. If the cardinality of A is 0 (minimum), then the value of this term is 1(maximum). If the cardinality of A is n(maximum), then its value is 0 (minimum). For all other subsets, its value is between 0 and 1. Maximizing fitness function maximizes classifier performance and the number of absent features in the subset, that is, minimizes feature count.

3.3. Initialization

A population of possible solutions (individuals) is generated randomly. Each solution is a subset encoded as a binary string as discussed above. The fitness value of each individual is calculated using Equation 1.

3.4. Selection

Individuals are randomly selected to form a mating pool, which may include duplicates. In this random selection process, every individual in the population has an equal chance of being chosen. Individuals are selected from the population at random and added to the mating pool.

3.5. Crossover (Recombination)

During crossover, two parent solutions are combined to create new offspring. Members of the mating pool are then paired randomly as parents P1 and P2, and a crossover operator is applied on each pair to create two offspring C1 and C2. The j^{th} ($0 \le j \le n$) bit of offspring is calculated by Equation 2. Where CR is the crossover rate defined between 0 and 1. A random number r is generated for each bit of offspring.

If r < CR, the corresponding bit is copied from P1 to C1 and from P2 to C2. If r > CR, then the corresponding bit is copied from P2 to C1 and from P1 to C2. It is like flipping a coin to choose a gene from any one of the parents.

$$\begin{pmatrix} C_{1j}=P_{1j}, C_{2j}=P_{2j} \; if \; r \le CR \\ C_{1j}=P_{2j}, C_{2j}=P_{1j} \; if \; r > CR \end{pmatrix} \quad (2)$$

3.6. Mutation

Mutation is performed on offspring randomly according to the mutation rate. The mutation rate is a real number between 0 and 1. Random mutation flips a random number of bits depending on the mutation rate, MR. MR is set between 0 and 1. A random number r is generated for each bit. If $r \le MR$, then the corresponding bit is flipped. The j^{th} bit of the i^{th} offspring is obtained by Equation 3.

$$C_{ij} = \begin{cases} 1 - C_{ij} & if \; r \le MR \\ C_{ij} & if \; r \ge MR \end{cases} \quad (3)$$

3.7. Replacement (Survivor Selection)

The Fitness of parents and offspring are compared, and those with better fitness value go to the next generation, and the cycle repeats. This continues until a stopping condition is met, that is, a maximum number of generations have evolved or a sufficiently good solution is found.

3.8. Multi-Optima Optimization

The above steps generate a single optimized subset. To achieve multiple optimal solutions, N populations evolve simultaneously.

The classifier's performance with all n features serves as the threshold for the minimum performance required from the subset. An optimized subset must perform equally well or better than this threshold. The proposed algorithm is designed

to generate multiple optimized subsets. The steps of the proposed GA are as follows. For simplification, the performance of CLF is being measured through accuracy.

1. Classifier CLF is trained and tested with all N features, and its accuracy, Acc, has been recorded as the threshold value.
2. N populations are initialized, each with M individuals, where each individual is a binary encoded subset.
3. Initial fitness values of all individuals in each population are calculated using Equation 1.
4. A mating pool of K individuals is created through random selection.
5. Individuals from the mating pool are paired, and crossover is performed using Equation 2.
6. The fitness value is calculated for each offspring, and individuals are replaced by their fitter child.
7. In each population, an individual with the maximum fitness value is obtained.
8. For each population i, if the accuracy of classifier CLF trained and tested with features in m_i is greater than or equal to Acc and if m_i is a unique subset derived so far, then it is saved as an optimal subset. The corresponding population is randomly reinitialized with new M individuals.
9. Steps 5 to 8 are repeated until a maximum number of epochs have expired or no new optimal solution is obtained over a set of iterations.

Figure 1 depicts the flowchart of the algorithm. After deriving multiple subsets, further analysis is performed to find critical features for the given problem domain. In general, features in more than 90 percent of subsets can be considered highly critical and less than 30 percent as least critical [16].

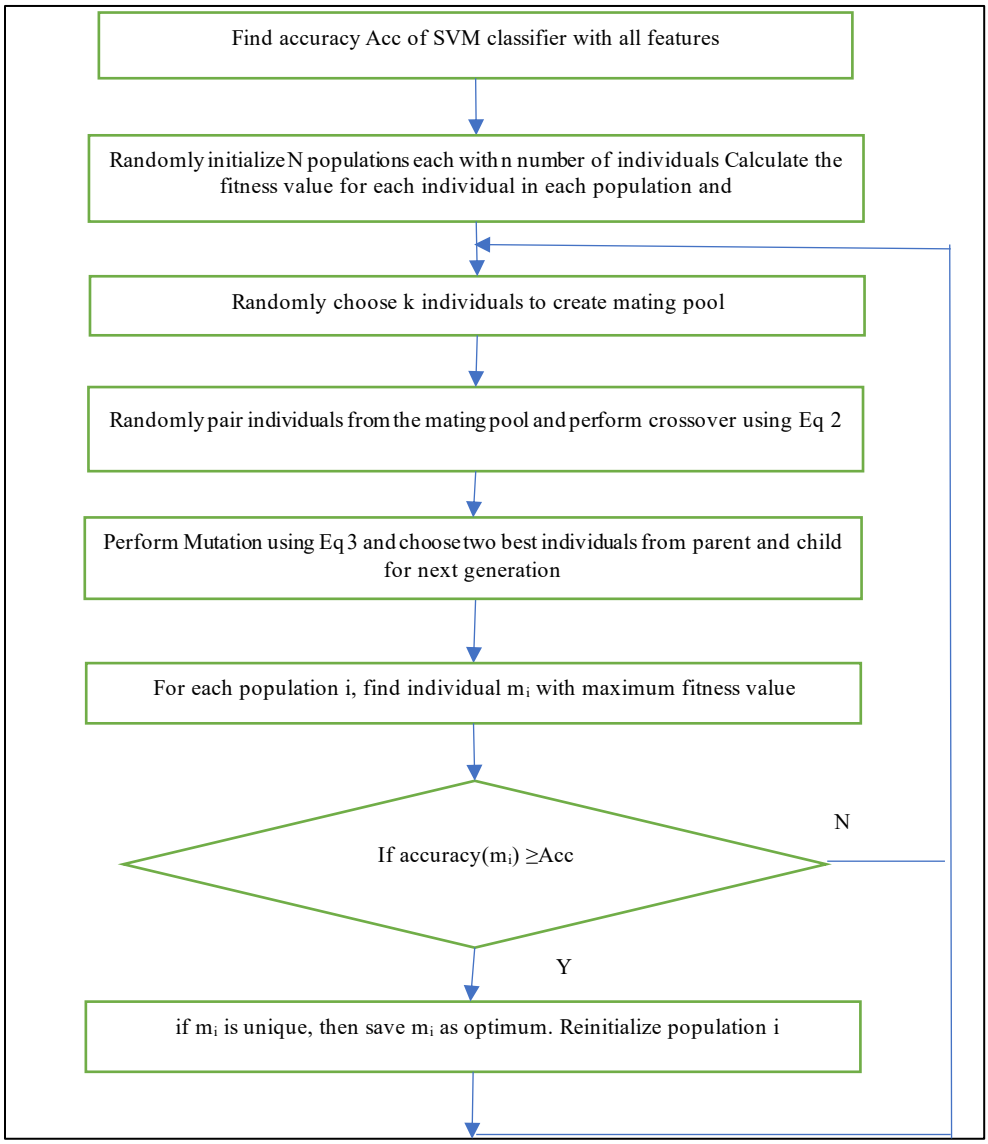

Fig. 1 Genetic algorithm with niching

4. Results and Discussion

4.1. Dataset

The dataset was obtained from digital ECG samples recorded in the Biotechnology and Bioengineering Lab, Birla Institute of Technology, Mesra, Ranchi, India. It comprises 71 fiducial features of 19 individuals (19-55 years male/female) listed in Table 1. P, Q, R, S, and T, known as fiducial points, represent the peaks and troughs in an ECG cardiac cycle, as illustrated in Figure 2. Each point in the digital ECG signal can be expressed as coordinate points (t, a), where a 'denotes the amplitude (y-axis) and 't' represents the time instance (x-axis). The features listed in Table 1 have been computed using the amplitude and time values of the fiducial points across all cardiac cycles for a specific ECG sample. The names of the features indicate which fiducial points were used to calculate their values. Temporal features represent the time intervals between the PQRST points, while amplitude features indicate the absolute differences in amplitudes. Distance, slope, angle, and miscellaneous features have been calculated using formulas from coordinate geometry. The suffix in each feature name denotes the type of feature. For example, QT, QTa, QTd, and QTs refer to the time interval, absolute amplitude difference, distance, and slope between the fiducial points Q and T, respectively.

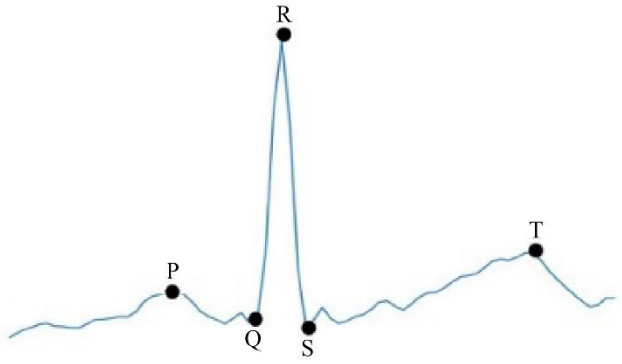

Fig. 2 ECG fiducial points PQRST

Table 1. ECG fiducial features

Temporal Features	PQ, QS, QT, PT, PS, ST, QT/QS, PT/QS	8
Amplitude Features	Py, Qy, Ry, Sy, Ty, ,PSa, QSa, RSa, STa, PQa, PTa, QRa, QTa, RS/QRa, RS/QSa, ST/QSa, PQ/RSa, PQ/QSa, RS/QTa, PQ/QTa, ST/PQa, PQ/PSa, ST/QTa, PQ/QRa	24
Distance Features	PQd, RSd, QSd, QRd, STd, ST/QSd, PRd, RS/QRd	9
Slope Features	PQs, STs, PSs, QRs, QSs, QTs, RSs, PTs, PRs, RTs	10
Angle Features	PQR, RST, RSQ, QRS, RQS, RTS	6
Miscellaneous Features	QRS area, S angle/PQ dis, S angle/QT time, QRS perimeter, QRS area/RS^2, (R/Q) angle, QRS in radius, (R/S)angle, (R/T) angle, QRS x-centroid, QRS y-centroid, R angle/QS time, (Q/T) angle, QRS area/ QR amp, RR	14
Total		**71**

If an ECG recording lasts for 30 seconds and includes 35 complete cardiac cycles, there will be 35 sets of PQRST fiducial points within that sample. Consequently, this results in 35 sets of 71 fiducial features for the sample. The number of cardiac cycles in a given duration varies from person to person. The recordings vary from 8 seconds to 10 minutes in duration, leading to the creation of 10 to 680 feature sets. For the classification problem, a dataset was constructed by randomly selecting 300 feature sets from each sample and shuffling them. Each sample is labeled with a unique person ID (1, 2, 3, …, 19), resulting in 19 distinct class labels. Feature sets from shorter samples have been selected multiple times. Finally, the dataset consists of 5700 tuples with 71 columns and 19 class labels.

The study involves a set of 71 features with 271 possible subsets. The objective is to identify all subsets that optimize the fitness function defined by Equation 1. Each subset is encoded as a binary string of 71 bits as discussed in the methodology section. A feature subset A is represented by a binary string B, where B[i] = 1 indicates that the ith feature is included in subset A, while B[i] = 0 indicates its exclusion. The fitness value for each binary string is calculated using the objective function in Equation 1, where the classifier used is a Support Vector Machine (SVM) with a linear kernel and n = 71. The SVM classifier has been trained and tested to identify individuals based on these 71 fiducial features. The data is split into training and testing sets with an 80:20 ratio, achieving an accuracy of 0.98. This accuracy serves as the threshold A for the minimum performance of an optimized subset. The other parameters are set as follows: the Crossover Rate (CR) is 0.5, the Mutation Rate (MR) is 0.5, and the number of epochs is 500. When the simple genetic algorithm with a single population was applied, each run generated a different subset with varying numbers and types of features, achieving accuracy greater than or equal to 0.98, indicating more than one competing optimized subset. Figure 2 depicts the convergence of a single population. When the proposed version of the genetic algorithm with multiple populations was applied, 500 distinct optimal subsets were achieved of cardinality from 25 to 46, each with a classifier accuracy from 0.98 to 0.990476. The result clearly shows that there is a significant reduction in the cardinality of features without compromising the classifier's performance.

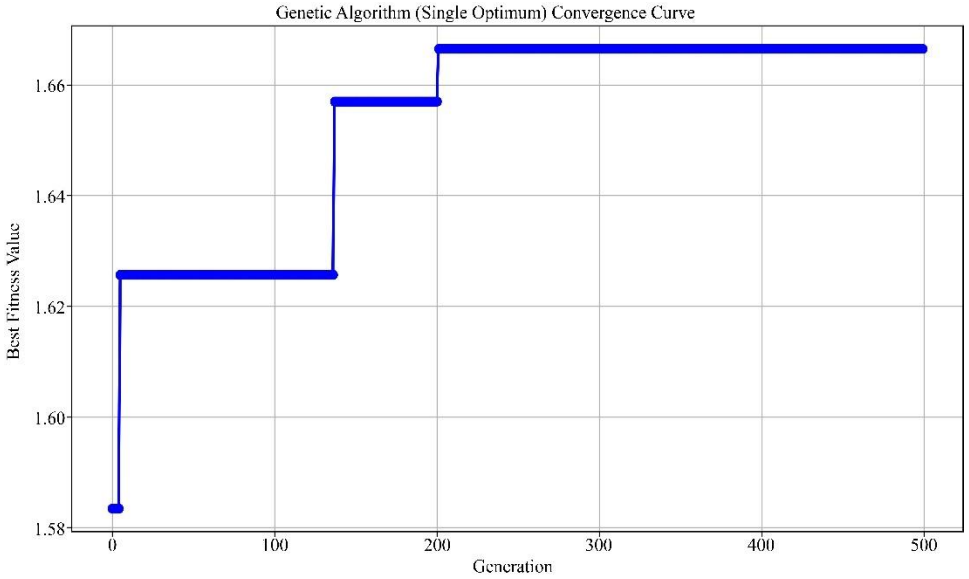

Fig. 3 Convergence of the simple genetic algorithm

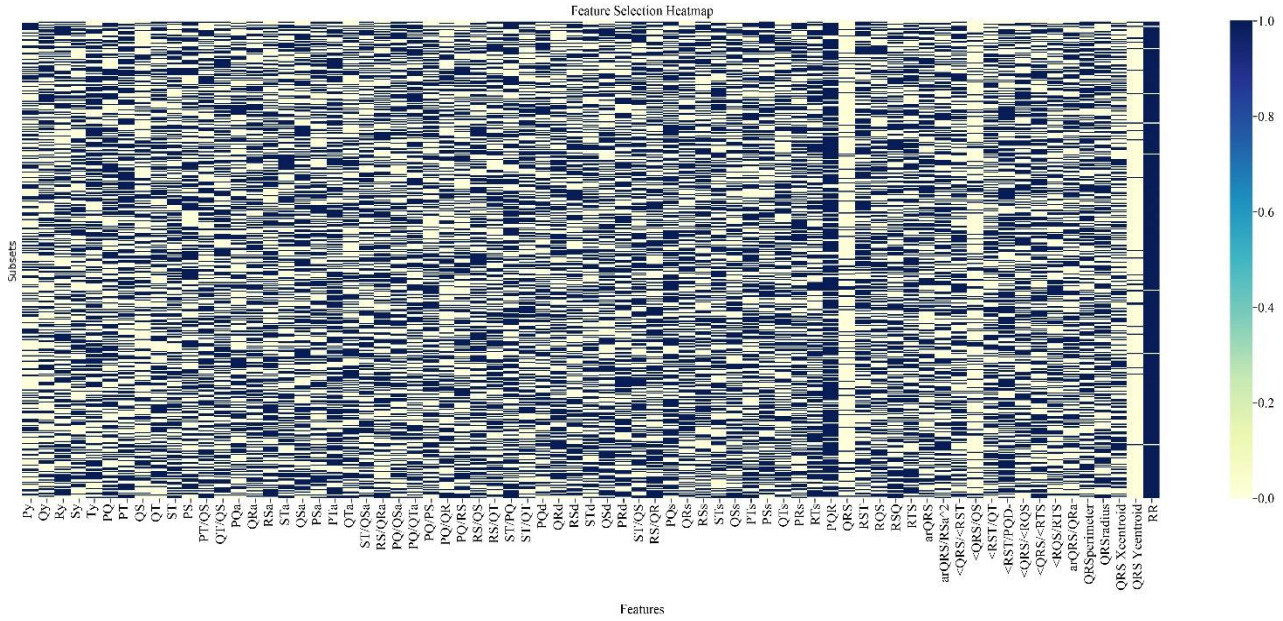

Fig. 4 Heat map of subsets

Table 2. Optimal feature subsets

SL No	Subset	Number of Features	Accuracy
1	'Qy', 'Ry', 'Ty', 'PQ', 'PQa', 'QRa', 'RSa', 'ST/QSa', 'PQ/QSa', 'ST/PQ', 'QRd', 'STd', 'STs', 'QSs', 'PTs', 'PRs', 'RTs', 'RST', 'RTS', 'arQRS', 'arQRS/RSa^2', '<QRS/<RST', '<RST/QT', 'QRS Xcentroid', 'RR'	25	0.990476
2	'Py', 'Qy', 'Ty', 'PT', 'PT/QS', 'QT/QS', 'PQa', 'QRa', 'PTa', 'PQ/PS', 'PQ/RS', 'RS/QT', 'ST/QT', 'PQd', 'ST/QS', 'QRs', 'RTs', 'PQR', 'RSQ', 'RTS', 'arQRS', '<RST/PQd', '<QRS/<RTS', 'arQRS/QRa', 'QRSradius', 'RR'	26	0.987302
3	'Qy', 'Ty', 'PQ', 'QT', 'PT/QS', 'RSa', 'ST/QSa', 'PQ/QSa', 'PQ/QTa', 'PQ/QR', 'RS/QS', 'ST/QT', 'STd', 'QSd', 'ST/QS', 'PQs', 'STs', 'QSs', 'PQR', 'RTS', 'arQRS', '<QRS/QS', '<QRS/<RQS', 'QRSperimeter', 'QRSradius', 'RR'	26	0.987302

4	'PT', 'ST', 'PS', 'QT/QS', 'QRa', 'RSa', 'PTa', 'QTa', 'RS/QRa', 'PQ/QR', 'PQ/RS', 'RS/QS', 'ST/PQ', 'ST/QS', 'QRs', 'STs', 'PSs', 'QTs', 'PRs', 'RTs', 'RST', 'RQS', 'arQRS/RSa^2', '<QRS/<RTS', 'QRSradius', 'RR'	26	0.987302
5	'Py', 'Sy', 'PT', 'ST', 'PS', 'PQa', 'QSa', 'PSa', 'ST/QSa', 'PQ/QTa', 'RS/QS', 'RS/QT', 'ST/PQ', 'ST/QT', 'RSd', 'STd', 'PRd', 'QRs', 'PSs', 'PRs', 'RSQ', 'arQRS', '<RST/QT', '<RQS/RTS', 'QRS Xcentroid', 'RR'	26	0.987302
6	'Qy', 'Ry', 'Ty', 'PQ', 'PT', 'ST', 'QRa', 'RSa', 'PQ/QTa', 'PQ/RS', 'RS/QT', 'QRd', 'STd', 'PRd', 'RSs', 'STs', 'QTs', 'RTs', 'PQR', 'RST', 'RQS', 'arQRS', '<RST/QT', '<RST/PQd', '<QRS/<RQS', '<RQS/RTS', 'RR'	27	0.987302
7	'Ry', 'Sy', 'PQ', 'ST', 'RSa', 'QSa', 'PSa', 'PTa', 'PQ/QSa', 'PQ/QTa', 'ST/PQ', 'STd', 'QSd', 'PQs', 'QRs', 'RSs', 'QSs', 'PRs', 'PQR', 'RSQ', 'arQRS/RSa^2', '<QRS/<RST', '<RST/QT', 'arQRS/QRa', 'QRSperimeter', 'QRSradius', 'RR'	27	0.987302
8	'Qy', 'Ry', 'Sy', 'PQ', 'PS', 'QT/QS', 'RSa', 'PTa', 'ST/QSa', 'PQ/QR', 'PQ/RS', 'RS/QS', 'RS/QT', 'ST/QT', 'RSd', 'PRd', 'ST/QS', 'RS/QR', 'QRs', 'STs', 'PTs', 'RTS', '<RST/PQd', 'arQRS/QRa', 'QRSradius', 'QRS Ycentroid', 'RR'	27	0.987302
9	'Ry', 'QT', 'PT/QS', 'QRa', 'PTa', 'QTa', 'ST/QSa', 'PQ/QTa', 'RS/QS', 'RS/QT', 'ST/QT', 'PQd', 'QRd', 'QSd', 'PRd', 'PQs', 'RSs', 'PTs', 'PSs', 'RTS', 'arQRS/RSa^2', '<QRS/<RST', '<RST/QT', '<QRS/<RQS', '<RQS/RTS', 'arQRS/QRa', 'RR'	27	0.987302
10	'Sy', 'Ty', 'PQ', 'QT/QS', 'QRa', 'STa', 'PSa', 'PTa', 'RS/QRa', 'RS/QS', 'RS/QT', 'ST/QT', 'QRd', 'STd', 'PRd', 'ST/QS', 'RS/QR', 'PSs', 'PQR', 'QRS', 'RSQ', 'arQRS/RSa^2', '<RST/QT', '<RST/PQd', 'arQRS/QRa', 'QRSradius', 'RR'	27	0.987302

Figure 6 depicts the simultaneous convergence of ten populations. The drift in a graph indicates reinitialization of the corresponding population. Figure 4 is a heatmap of all subsets. The rows correspond to subsets, and the columns to features. The dark cell indicates the presence, and the light cell indicates the absence of the corresponding feature. The darker column indicates that the corresponding feature is present in more subsets than the feature corresponding to the lighter column. Figure 5 illustrates the occurrences of various features within optimal subsets.

The presence of a feature in an optimal subset signifies its importance in identifying a person. Features that appear in 90% or more of the optimal subsets are deemed critical, while those present in less than 30% are considered least critical. It has been observed that the 'RR distance' is the most critical feature, appearing in 98% of the subsets. In contrast, the 'Y-centroid' and 'QRS angle' appear in less than 30% of the subsets, categorizing them as the least critical features.

Analyzing different subsets can reveal unique characteristics of ECG signals, aiding domain experts in uncovering subtle patterns or relationships between features and identity-related markers. In cases of missing or corrupted ECG data, alternative subsets provide reliable backup solutions, ensuring the identification system remains functional and effective. Table 2 lists the top ten optimal subsets of cardinalities 25, 26, and 27 and accuracies from 0.987302 to 0.990476. Figure 7 shows a comparison among the classifier performances of these ten subsets and the set of all 71 features. The identification precision is high with all 71 features, but accuracy, recall, and F1-score are low. The classifier performances with other optimal subsets are more stable.

The subset with the fewest features is often considered the most optimal. However, a subset with low cardinality may not always be the most computationally efficient. Different types of features require varying amounts of computation time. For example, calculating straight distance, slope, amplitude, and temporal difference involves evaluating a single expression, making them relatively simple. In contrast, computing ratios of these features requires a three-step process (calculating both the numerator and the denominator). Calculating angles, areas, and centroids is more complex.

By detecting multiple subsets, we can facilitate comparisons and choose from various solutions based on computation time. For instance, in Table 2, subsets 2 and 3 both have 26 features. However, set 2 has five ratio features, whereas set 3 has seven; hence, set 2 has less computation time.

Additionally, subsets can be compared across various models, situations, or configurations of ensemble models. In dynamic environments where the distribution of ECG signals or context may change, maintaining multiple feature subsets ensures adaptability. This approach allows models to switch to subsets that are better suited for new conditions.

In the future, identifying multiple optimal subsets and critical features will be highly beneficial and insightful for studying Heart Rate Variability (HRV) and developing robust systems for authentication, disease and stress prediction using various ECG features. The authors also plan to experiment with combining other feature optimization techniques alongside the proposed method. Although the proposed algorithm is applied and tested on the ECG dataset, it is a generalized approach for feature optimization.

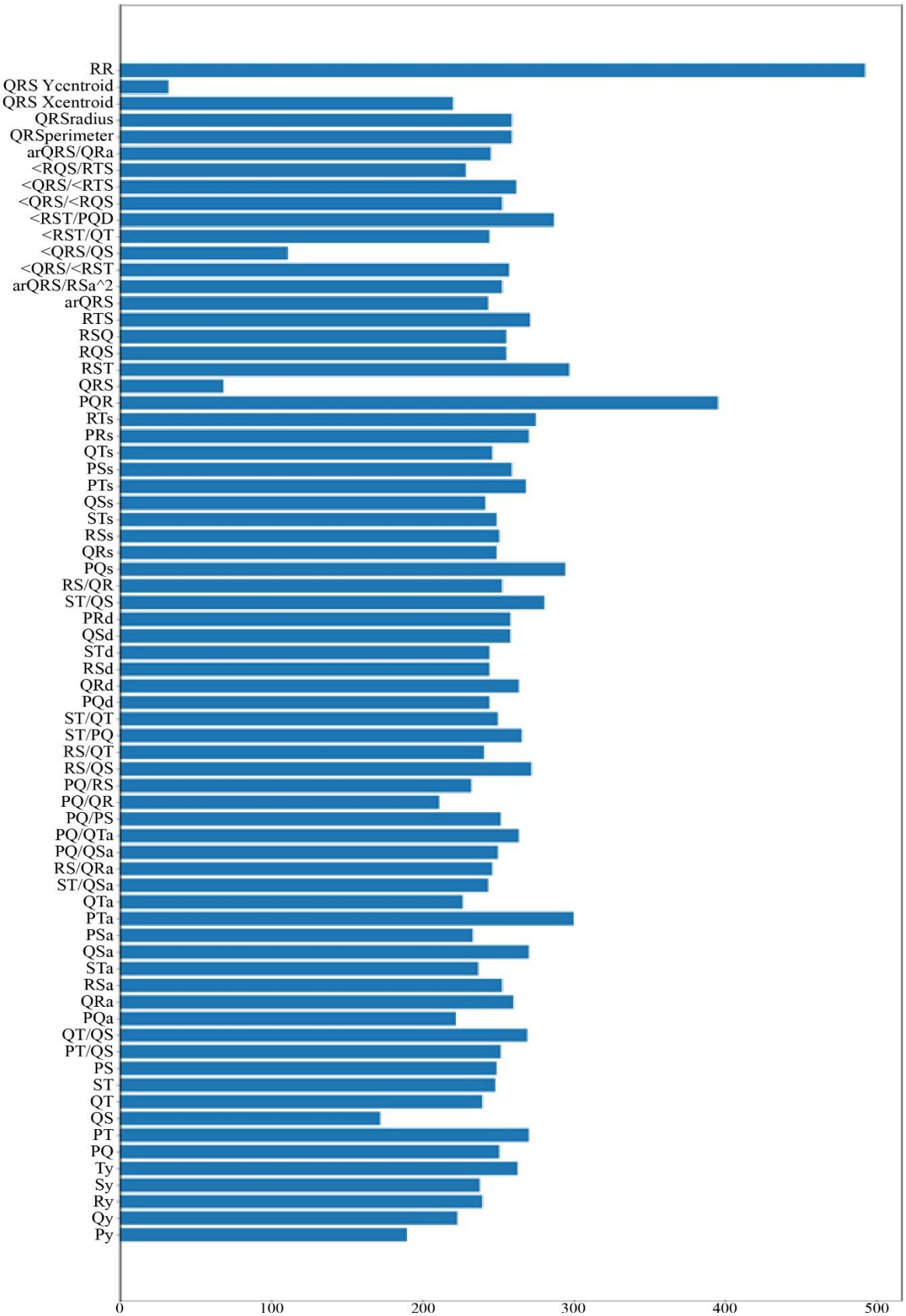

Fig. 5 Occurrences of features in optimal subsets

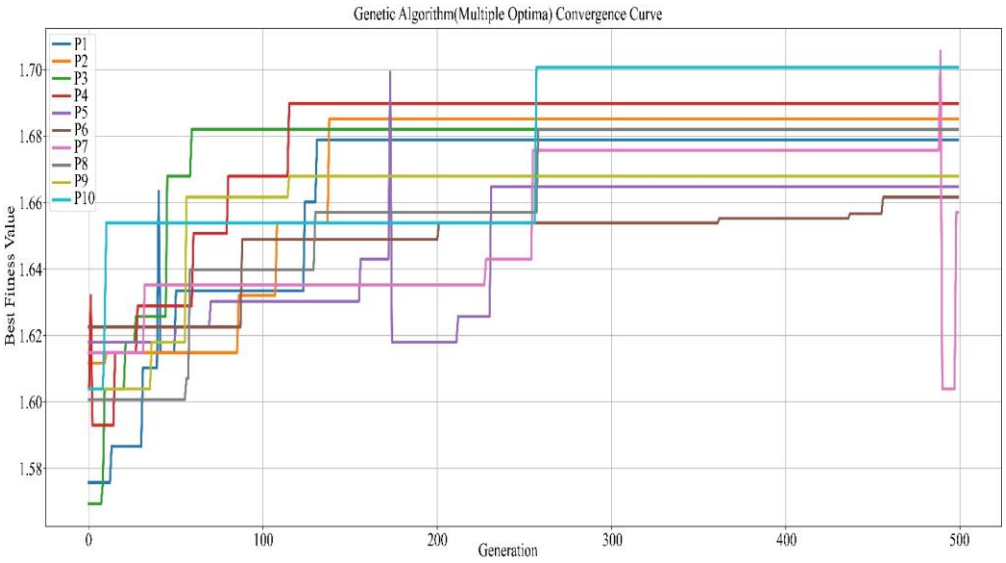

Fig. 6 Simultaneous convergence of ten populations

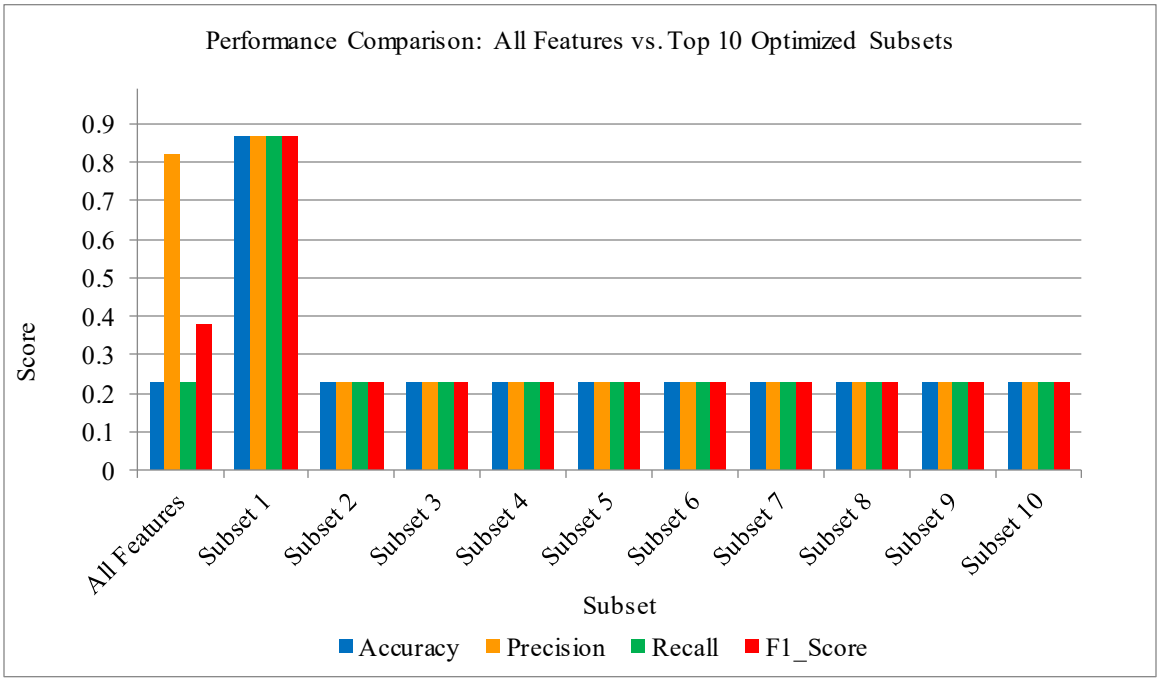

Fig. 7 Performance comparison of all features vs Top 10 subsets

5. Conclusion

This paper presents a novel variant of the Multiobjective Multi-Optima Genetic Algorithm (GA) that incorporates niching techniques for identifying multiple optimal feature subsets in ECG-based identification. The approach explores multiple populations simultaneously to generate diverse feature subsets, each designed to maximize classifier performance while minimizing the cardinality. The algorithm has been applied to an ECG dataset containing 71 fiducial features, resulting in 500 distinct feature subsets with cardinalities ranging from 25 to 46. This approach improved SVM classifier accuracy from 0.98 to 0.9904. Further analysis of these subsets revealed critical features for identification. The RR-distance and QRS angle features were found to be the most critical, appearing in over 90% of the subsets, while the Y-centroid feature was the least critical, with a presence in less than 30% of the subsets.

References

[1] Fahim Sufi, Ibrahim Khalil, and Jiankun Hu, *ECG-Based Authentication*, Handbook of Information and Communication Security, Springer, Berlin, Heidelberg, pp. 309-331, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[2] Kiran Kumar Patro et al., "An Efficient Optimized Feature Selection with Machine Learning Approach for ECG Biometric Recognition," *IETE Journal of Research*, vol. 68, no. 4, pp. 2743-2754, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Noor Jameel, and Hasanen S. Abdullah., "Intelligent Feature Selection Methods: A Survey," *Engineering and Technology Journal*, vol. 39, no. 1B, pp. 175-183, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Matheus Cezimbra Barbieri, Bruno Iochins Grisci, and Márcio Dorn, "Analysis and Comparison of Feature Selection Methods towards Performance and Stability," *Expert Systems with Applications*, vol. 249, pp. 2-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Huanhuan Gong et al., "A New Filter Feature Selection Algorithm for Classification Task by Ensembling Pearson Correlation Coefficient and Mutual Information," *Engineering Applications of Artificial Intelligence*, vol. 131, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Varun Gupta et al., "Detection of R-peaks using Fractional Fourier Transform and Principal Component Analysis," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 2, pp. 961-972, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Zohre Sadeghian et al., "A Review of Feature Selection Methods based on Meta-Heuristic Algorithms," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 37, no. 1, pp. 1-51, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] Hongqiang Li et al., "Genetic Algorithm for the Optimization of Features and Neural Networks in ECG Signals Classification," *Scientific Reports*, vol. 7, no. 1, pp. 1-12, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[9] Rami N. Khushaba, Ahmed Al-Ani, and Adel Al-Jumaily, "Differential Evolution based Feature Subset Selection," *2008 19th International Conference on Pattern Recognition*, Tampa, FL, USA, pp. 1-4, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[10] Gilseung Ahn, and Sun Hur, "Efficient Genetic Algorithm for Feature Selection for Early Time Series Classification," *Computers & Industrial Engineering*, vol. 142, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Nazneen Akhter et al., "Feature Selection for Heart Rate Variability based Biometric Recognition using Genetic Algorithm," *Intelligent Systems Technologies and Applications*, vol. 1, pp. 91-101, 2016 [CrossRef] [Google Scholar] [Publisher Link]

[12] Nishath Ansari, "A Survey on Feature Selection Techniques Using Evolutionary Algorithms," *Iraqi Journal of Science*, vol. 62, no. 8, pp. 2796-2812, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Manosij Ghosh et al., "Binary Genetic Swarm Optimization: A Combination of GA and PSO for Feature Selection," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1598-1610, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Bing Xue et al., "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606-626, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[15] Amal Bouraoui, Salma Jamoussi, and Yassine BenAyed, "A Multi-Objective Genetic Algorithm for Simultaneous Model and Feature Selection for Support Vector Machines," *Artificial Intelligence Review*, vol. 50, no. 2, pp. 261-281, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[16] Jorge Bustos, Victor A. Jimenez, and Adrian Will, "A Comparison of Different Types of Niching Genetic Algorithms for Variable Selection in Solar Radiation Estimation," *arXiv Preprint*, pp. 1-10, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Jonas da S. Bohrer, and Márcio Dorn, "Enhancing Classification with Hybrid Feature Selection: A Multi-Objective Genetic Algorithm for High-Dimensional Data," *Expert Systems with Applications*, vol. 255, pp. 1-91, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Sourabh Katoch, and Sumit Singh Chauhan, and Vijay Kumar, "A Review on Genetic Algorithm: Past, Present, and Future," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091-8126, 2021 [CrossRef] [Google Scholar] [Publisher Link]

[19] G. Thippa Reddy et al., "Hybrid Genetic Algorithm and a Fuzzy Logic Classifier for Heart Disease Diagnosis," *Evolutionary Intelligence*, vol. 13, no. 2, pp. 185-196, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Lamis Ghoualmi, Amer Draa, and Salim Chikhi, "An Efficient Feature Selection Scheme based on Genetic Algorithm for Ear Biometrics Authentication," *2015 12th International Symposium on Programming and Systems (ISPS)*, Algiers, Algeria, pp. 1-5, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[21] Lei Lu, Jihong Yan, and Clarence W. de Silva, "Feature Selection for ECG Signal Processing using Improved Genetic Algorithm and Empirical Mode Decomposition," *Measurement*, vol. 94, pp. 372-381, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[22] Syaiful Anam, "Parameters Estimation of Enzymatic Reaction Model for Biodiesel Synthesis by using Real Coded Genetic Algorithm with Some Crossover Operations," *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 5, pp. 1-9, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] Elijah M. Maseno, and Zenghui Wang, "Hybrid Wrapper Feature Selection Method based on Genetic Algorithm and Extreme Learning Machine for Intrusion Detection," *Journal of Big Data*, vol. 11, no. 1, pp. 1-25, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Emre Kocyigit et al., "Enhanced Feature Selection using Genetic Algorithm for Machine-Learning-based Phishing URL Detection," *Applied Sciences*, vol. 14, no. 14, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Halit Bakır, and Özlem Ceviz, "Empirical Enhancement of Intrusion Detection Systems: A Comprehensive Approach with Genetic Algorithm-based Hyperparameter Tuning and Hybrid Feature Selection," *Arabian Journal for Science and Engineering*, vol. 49, no. 9, pp. 13025-13043, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Rongtao Zhang et al., "GA-FCFNN: A New Forecasting Method Combining Feature Selection Methods and Feedforward Neural Networks using Genetic Algorithms," *Information Sciences*, vol. 669, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] A.J. Umbarkar, and P.D. Sheth, "Crossover Operators in Genetic Algorithms: A Review," *ICTACT Journal on Soft Computing*, vol. 6, no. 1, pp. 1083-1092, 2016. [Google Scholar] [Publisher Link]

[28] Tea Tušar, and Bogdan Filipič, "Differential Evolution versus Genetic Algorithms in Multiobjective Optimization," *International Conference on Evolutionary Multi-Criterion Optimization*, Matsushima, Japan, pp. 257-271, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[29] Alex Rogers, and Adam Prugel-Bennett, "Genetic Drift in Genetic Algorithm Selection Schemes," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 298-303, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[30] Anupriya Shukla, Hari Mohan Pandey, and Deepti Mehrotra, "Comparative Review of Selection Techniques in Genetic Algorithm," *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, Greater Noida, India, pp. 515-519, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[31] Rakesh Kumar, Girdhar Gopal, and Rajesh Kumar "Novel Crossover Operator for Genetic Algorithm for Permutation Problems," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, no. 2, pp. 252-258, 2013. [Google Scholar] [Publisher Link]

[32] Varun Kumar SG, and R. Panneerselvam, "A Study of Crossover Operators for Genetic Algorithms to Solve VRP and its Variants and New Sinusoidal Motion Crossover Operator," *International Journal of Computational Intelligence Research (IJCIR)*, vol. 13, no. 7, pp. 1717-1733, 2017. [Google Scholar] [Publisher Link]

[33] N. Malarvizhi, P. Selvaran, and Pethuru Raj, "Adaptive Fuzzy Genetic Algorithm for Multi Biometric Authentication," *Multimedia Tools and Applications*, vol. 79, no. 13, pp. 9131-9144, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[34] Evgenii A. Sopov, "Multiple Optima Identification using Multi-Strategy Multimodal Genetic Algorithm," *Journal of Siberian Federal University - Mathematics and Physics*, vol. 9, no. 2, pp. 246-257, 2016. [Google Scholar] [Publisher Link]

[35] E. Pérez, M. Posada, and F. Herrera, "Analysis of New Niching Genetic Algorithms for Finding Multiple Solutions in the Job Shop Scheduling," *Journal of Intelligent Manufacturing*, vol. 23, no. 3, pp. 341-356, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[36] Ole J. Mengshoel, and David E. Goldberg, "The Crowding Approach to Niching in Genetic Algorithms," *Evolutionary Computation*, vol. 16, no. 3, pp. 315-354, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[37] Michael G. Epitropakis, Vassilis Plagianakos, and Michael N. Vrahatis, "Finding Multiple Global Optima Exploiting Differential Evolution's Niching Capability," *2011 IEEE Symposium on Differential Evolution (SDE)*, Paris, France, pp. 1-8, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[38] D. Parrott, and Xiaodong Li, "Locating and Tracking Multiple Dynamic Optima by a Particle Swarm Model using Speciation," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 4, pp. 440-458, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[39] Changhe Li et al., "An Adaptive Multipopulation Framework for Locating and Tracking Multiple Optima," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 590-605, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[40] Hong Liang, "ECG Feature Elements Identification for Cardiologist Expert Diagnosis," *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, China, pp. 3845-3848, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[41] Tae Wuk Bae, and Kee Koo Kwon, "ECG PQRST Complex Detector and Heart Rate Variability Analysis using Temporal Characteristics of Fiducial Points," *Biomedical Signal Processing and Control*, vol. 66, pp. 1-21, 2021. [CrossRef] [Google Scholar] [Publisher Link]