Performance analysis of single and multi-stage metaheuristic optimization on DFFNN for electrocardiogram-based emotion classification

Giovanni Dimas Prenata, Ahmad Ridho'i
Faculty of Intelligent Electrical and Informatics Technology, Universitas 17 Agustus 1945 Surabaya, Surabaya, Indonesia

Article Info

Article history:

Received Jan 16, 2026
Revised Mar 27, 2026
Accepted Apr 26, 2026

Keywords:

Deep feedforward neural network
Electrocardiogram
Emotion classification
Genetic algorithm
Grey wolf optimizer
Hybrid metaheuristics
Particle swarm optimization

ABSTRACT

Emotion classification based on electrocardiogram (ECG) signals has attracted increasing attention in affective computing and biomedical signal processing. However, training deep feedforward neural networks (DFFNN) using conventional gradient-based learning often suffers from local minima and slow convergence, particularly when dealing with nonlinear and limited datasets. This study presents a comprehensive performance analysis of single-stage and multi-stage metaheuristic optimization strategies applied to DFFNN for ECG-based emotion classification in elderly participants. Five models were evaluated: Pure DFFNN, DFFNN optimized using genetic algorithm (GA), particle swarm optimization (PSO), grey wolf optimizer (GWO), and a hybrid multi-stage DFFNN+GA+GWO model. Experimental results from six independent trials demonstrate a substantial reduction in mean squared error (MSE) when metaheuristic optimization is applied. Pure DFFNN produced final MSE values in the range of 0.07462–0.08977, whereas DFFNN+GWO reduced MSE to 0.01894–0.02411. The proposed multi-stage DFFNN+GA+GWO achieved the lowest MSE of 0.014286 in the best run and an average MSE of approximately 0.0212 across trials. Training accuracy improved from 57.14%–66.67% (Pure DFFNN) to 80.95%–85.71% using metaheuristic approaches. Although testing accuracy remained relatively stable at 33.33%–50.00% due to dataset size constraints, convergence behavior analysis shows that multi-stage optimization enhances stability and reduces oscillatory updates. These findings confirm that multi-stage metaheuristic optimization significantly improves training stability and error minimization in DFFNN models, offering a promising strategy for robust ECG-based emotion classification under small-sample conditions.

Corresponding Author:

Giovanni Dimas Prenata
Faculty of Intelligent Electrical and Informatics Technology, Universitas 17 Agustus 1945 Surabaya
Surabaya, Indonesia
Email: gprenata@untag-sby.ac.id

1. INTRODUCTION

Electrocardiogram (ECG) signals do not merely represent the electrical activity of the heart, but also encapsulate rich physiological information that is closely associated with human emotional states. Numerous studies have demonstrated that emotional variations such as stress, anxiety, fear, and relaxation influence the dynamics of the autonomic nervous system, which are reflected in changes in ECG morphology and rhythm, particularly through heart rate variability (HRV) parameters [1], [2]. Consequently, ECG-based emotion classification has emerged as an increasingly active research topic, especially in applications related to affective computing, mental health monitoring, and human computer interaction [3].

Recent advances in electrocardiogram analysis have also expanded from conventional emotion recognition toward signal forecasting, disease detection, and deep sequential learning frameworks. Zacarias *et al.* [4] demonstrated that long short-term memory (LSTM) networks can effectively model temporal dependencies in ECG forecasting tasks. Hybrid prediction approaches combining signal decomposition and neural networks have also shown promising performance for ECG reconstruction and prediction [5]. Furthermore, recent systematic reviews confirm that deep learning methods are increasingly dominant in ECG classification due to their ability to automatically learn complex representations from physiological data [6].

Conventional approaches relying on handcrafted statistical features and linear classification methods often face limitations in capturing the nonlinear and complex characteristics of ECG signals. To address this issue, artificial neural network (ANN)-based models have been widely adopted due to their strong capability in modeling nonlinear relationships among physiological signal features [7]. One relatively simple yet effective ANN architecture is the deep feedforward neural network (DFFNN), which employs multiple hidden layers to enhance representational capacity [8]. Nevertheless, the performance of DFFNN is highly dependent on the training process, particularly on the optimization of network weights and biases.

Several recent studies have also applied recurrent neural networks and LSTM architectures for ECG-based disease recognition and biomedical classification, showing strong capability in temporal feature modeling [9]. However, these methods often require larger datasets and higher computational complexity than feedforward architectures, making DFFNN still attractive for small-sample scenarios such as the present study.

Gradient-based learning methods such as backpropagation frequently suffer from issues including local minima entrapment, slow convergence, and sensitivity to weight initialization, especially when dealing with noisy and non-stationary ECG data [10]. These limitations have motivated the adoption of population-based metaheuristic optimization algorithms that do not rely on gradient information. Among the most widely applied metaheuristic techniques for neural network weight optimization are the genetic algorithm (GA), particle swarm optimization (PSO), and grey wolf optimizer (GWO) [11]–[13].

GA mimics biological evolutionary processes such as selection, crossover, and mutation to perform global search in complex solution spaces [14]. While GA is effective in exploration, it often exhibits relatively slow convergence during the exploitation phase. In contrast, PSO, inspired by the social behavior of particle swarms, is known for its fast convergence speed, but it is susceptible to premature convergence [15]. GWO, which is inspired by the social hierarchy and hunting strategies of grey wolves, has demonstrated a favorable balance between exploration and exploitation, along with strong convergence stability in various nonlinear optimization problems [16].

In the context of ECG-based emotion classification using neural networks, most existing studies have primarily focused on single-stage metaheuristic optimization, where only one optimization algorithm is integrated with artificial neural network or deep neural networks (DNN) models [17]–[19]. Although such approaches generally outperform conventional gradient-based training, they may still exhibit convergence instability and significant fluctuations in mean squared error (MSE) during the training process, particularly when applied to complex physiological signal datasets.

As an alternative, multi-stage metaheuristic optimization has recently attracted increasing attention. In this strategy, one algorithm is employed during the initial training stage to enhance global exploration, followed by another algorithm that emphasizes local exploitation in later stages [20]. Conceptually, this approach aims to combine the strengths of different optimization algorithms, leading to faster convergence, improved stability, and lower final MSE values.

In our previous work, metaheuristic-optimized neural networks were successfully applied to ECG prediction under limited-data conditions, where hybrid optimization strategies showed improved predictive stability compared with baseline models [21]. These findings motivate the present extension toward ECG-based emotion classification.

From Table 1 based on the limitations of existing studies, this research proposes a comparative framework involving single-stage and multi-stage metaheuristic optimization applied to DFFNN for ECG-based emotion classification. Therefore, this study should be interpreted not merely as a classification performance comparison, but as an analysis of optimization stability and convergence characteristics in metaheuristic-driven neural networks.

Beyond single optimizers, recent engineering studies have shown that hybrid swarm-based metaheuristics can improve exploration-exploitation balance more effectively than standalone methods. For example, combinations of GWO and PSO have demonstrated improved robustness in complex optimization problems [22]. This supports the rationale for investigating staged hybrid optimization strategies in neural network training.

Table 1. Summary of related work on ECG-Based emotion classification

Ref	Method	Dataset	Optimization	Accuracy	Limitation
[2]	ANN	DEAP	None	78%	Local minima
[18]	CNN	ECG dataset	Gradient	85%	Large data requirement
[19]	SVM	HRV features	Grid search	80%	Limited scalability
[16]	MLP+GWO	UCI	GWO	87%	Single-stage only
Purpose	DFFNN+GA+GWO	Elderly ECG	Multi-stage	85,71%	Small dataset

2. METHOD

This study is a continuation of previous research involving elderly participants [23]. Using a SparkFun AD8232 ECG sensor and an Arduino Uno R3, participants were exposed to emotion-inducing video stimuli. Before switching to the next emotional stimulus video, participants were shown a relaxation video to help restore their emotional state to a neutral baseline. The ECG signals acquired from the SparkFun AD8232 sensor were recorded and stored on a computer in CSV format. Figure 1 illustrates the overall methodological framework of the proposed study.

The dataset used in this study consists of ECG recordings obtained from elderly participants exposed to three emotional states: sad, surprised, and angry. A total of 27 ECG segments were collected, comprising 21 samples for training and 6 samples for testing. Each ECG signal was processed to extract eight HRV-based features, including MeanRR, SDNN, RMSSD, MeanHR, STDHR, LF, HF, and LF/HF ratio. It is important to note that the dataset size is relatively limited due to the difficulty of collecting controlled ECG-based emotional responses from elderly participants. Therefore, this study focuses not only on classification performance but also on optimization behavior, convergence stability, and reproducibility across multiple independent runs, which are more suitable evaluation criteria for small-sample biomedical datasets. This experimental design is supported by psychophysiological findings indicating that positive and negative affective states significantly influence cardiovascular responses and autonomic nervous system regulation [24], [25]. Figure 2 presents the data acquisition process from elderly participants.

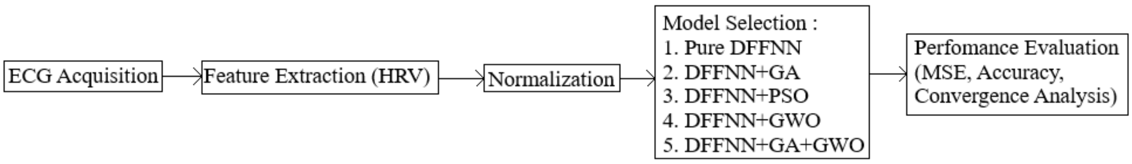

Figure 1. Overall methodological framework of the proposed study

Figure 2. Data collection from elderly participant

2.1. ECG feature extraction

Before conducting training and testing using the five DFFNN-based models (Pure DFFNN, DFFNN+GA, DFFNN+GWO, DFFNN+PSO, and DFFNN+GA+GWO), a feature extraction stage was first performed on the ECG signals based on heart rate variability (HRV). The feature extraction process began by reading ECG data from CSV files formatted as time; ECG_value. The extracted values were stored in the arrays $waktu[]$ and $sinyal[]$, with a maximum capacity of 10,000 samples. After loading the data, the system computed the mean and standard deviation of the ECG signal to establish an automatic R-peak detection threshold, defined as $threshold = mean + 1.5 \times standard\ deviation$. This threshold was applied to ensure that the detected peaks corresponded to dominant R-waves rather than noise or minor fluctuations. Figure 3 shows the ECG feature extraction workflow.

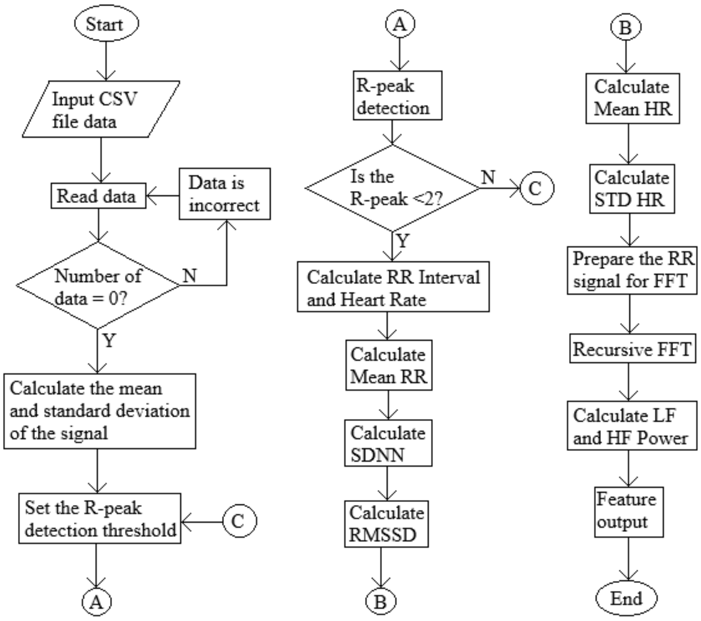

Figure 3. ECG feature extraction flowchart

R-wave (R-peak) detection was then performed using a local maximum criterion, where a sample at index i was considered an R-peak if its amplitude exceeded both neighboring samples ($sinyal[i] > sinyal[i-1]$ and $sinyal[i] > sinyal[i+1]$) and surpassed the predetermined threshold. To prevent multiple detections caused by closely spaced peaks, a minimum separation of 100 samples between consecutive peaks was enforced. The detected R-peak timestamps were stored in the array $waktuR[]$. Based on these timestamps, RR intervals were computed as the time differences between successive R-peaks. In this implementation, the ECG sampling frequency was set to fs=250 Hz, and RR intervals were calculated using $RR(i) = (waktuR(i+1) - waktuR(i))/fs$ in seconds. Heart rate (HR) was subsequently derived from the RR intervals using $HR(i) = 60/RR(i)$ in beats per minute (bpm).

Once the RR intervals and HR values were obtained, HRV time-domain features were extracted. These included MeanRR, representing the average RR interval; SDNN, the standard deviation of RR intervals reflecting overall heart rate variability; and RMSSD, computed as the root mean square of successive RR differences to capture short-term beat-to-beat variability. In addition, MeanHR and STDHR were calculated as the mean and standard deviation of the heart rate sequence, respectively. Together, these five time-domain features constitute key HRV descriptors and provide discriminative information across emotional conditions.

In addition to time-domain analysis, frequency-domain HRV features were extracted using the fast Fourier transform (FFT). The RR interval sequence was first detrended by removing its mean component through the operation RR - MeanRR, then represented as a signal vector of length N=256 (a power of two) for FFT computation, with zero-padding applied when the number of RR intervals was less than N. The resulting FFT spectrum was used to estimate spectral power within two principal HRV frequency bands: low frequency (LF) in the range of 0.04-0.15 Hz and high frequency (HF) in the range of 0.15-0.40 Hz. The LF/HF ratio was then calculated to reflect the balance between sympathetic and parasympathetic nervous system activity. In the program implementation, the frequency resolution was determined using

$fs\_resample = 4$ Hz and $df = fs\_resample/N$. Consequently, the feature extraction stage produced a total of eight HRV features, comprising five time-domain features (MeanRR, SDNN, RMSSD, MeanHR, STDHR) and three frequency-domain features (LF, HF, LF/HF), which were subsequently used as training and testing inputs for all DFFNN models evaluated in this study.

2.2. Pure DFFNN

In this study, a pure deep feedforward neural network (Pure DFFNN) is employed as the primary baseline model for ECG-based emotion classification without incorporating any metaheuristic optimization. The input data consist of eight ECG/HRV-derived features, organized into an 8×27 matrix (eight features across 27 samples). Prior to being processed by the neural network, row-wise Min-Max normalization is applied to each feature, scaling all values into the range of 0 to 1. This normalization step improves training stability and prevents features with larger magnitudes from dominating the learning process.

The dataset in the program is divided into 21 training samples and six testing samples. The training targets are numerically defined in the array T [9], where 0.1 represents the Sad class, 0.5 represents the Surprise class, and 0.9 represents the angry class. A sigmoid activation function is used for all neurons, ensuring that the network outputs remain within the interval [0, 1]. Based on the implemented code, the network architecture consists of multiple hidden layers with the configuration 8 inputs → 2 neurons (Layer 1) → 2 neurons (Layer 2) → 2 neurons (Layer 3) → 1 output neuron, resulting in a total of 33 trainable weight parameters. The initial weights are randomly generated from a uniform distribution within the range of [-1, 1]. Figure 4 illustrates the training mechanisms of Pure DFFNN model.

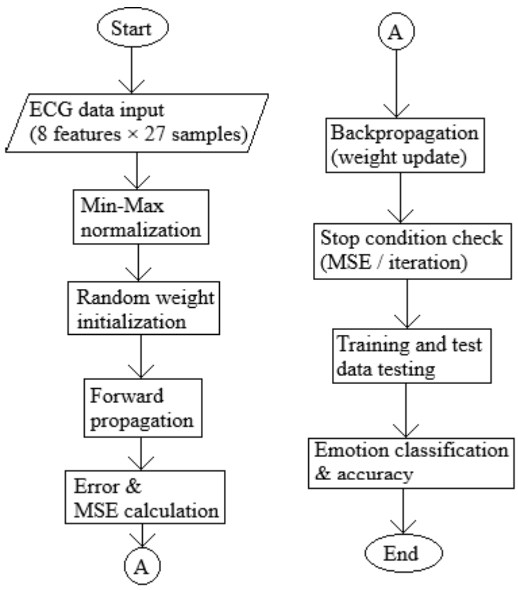

Figure 4. Pure DFFNN flowchart

Training is performed using a standard feedforward-backpropagation mechanism driven by output error minimization. For each iteration, all 21 training samples are forward-propagated to produce the network output S_OY, and the error is computed using $ERR = T[k] - S\_OY$. This error is then used to calculate gradients and update the weights sequentially from the output layer back to the input layer using the derivative of the sigmoid function. The learning process adopts a fixed learning rate (LR) of 0.7, while the objective function to be minimized is the MSE computed as the average squared error over the 21 training samples. The training terminates when the MSE falls below the predefined threshold of 0.0001, or when the iteration count reaches the maximum limit of 5,000,000 iterations. For convergence analysis, the MSE value at each iteration is recorded in mse_log.csv in the format (iteration, MSE).

After training, the model is evaluated on the complete dataset of 27 samples (21 training samples and six testing samples) using the learned weights. The final sigmoid output (Output) is mapped into emotion labels based on predefined interval thresholds: sad if output < 0.45, surprise if 0.45 ≤ S07 ≤ 0.80, and angry if output > 0.8. Classification accuracy is calculated by comparing the predicted labels against the reference labels stored in $T2$[], and the performance is reported separately for the training set (21 samples) and the testing set (six samples). Consequently, the Pure DFFNN model serves as the key baseline for assessing performance improvements achieved through metaheuristic optimization strategies such as GA, GWO, PSO, and GA+GWO in this research. To ensure fair comparison across all models, the classification decision thresholds were standardized for all experiments.

2.3. DFFNN+GA

In the DFFNN+GA model, the GA is employed to globally optimize the weights and biases of the DFFNN, thereby eliminating the need for gradient-based backpropagation. Each candidate solution in GA is represented as a chromosome consisting of 33 genes, where each gene corresponds to a specific weight or bias parameter in the adopted DFFNN architecture. The initial population is generated with 10 chromosomes, and each gene is randomly initialized within the range of -30 to 30 to ensure sufficient diversity at the beginning of the optimization process. The main experimental parameters include a crossover point (CP) of 10, a mutation rate of 0.3, a maximum iteration limit of 5,000,000, and an MSE threshold of 0.0001 as the stopping criterion.

The quality of each chromosome is evaluated using a fitness function based on MSE. Specifically, the 33 genes of a chromosome are assigned as the DFFNN parameters, and the network performs forward propagation over all 21 training samples. For each training sample, the network output is computed using the sigmoid activation function at every neuron. The error is then calculated as the difference between the predefined target value and the network output. The overall MSE is obtained by averaging the squared errors across all training samples, and this value is used as the fitness score, where a smaller MSE indicates a better chromosome. Figure 5 illustrates the training mechanisms of Pure DFFNN+GA model.

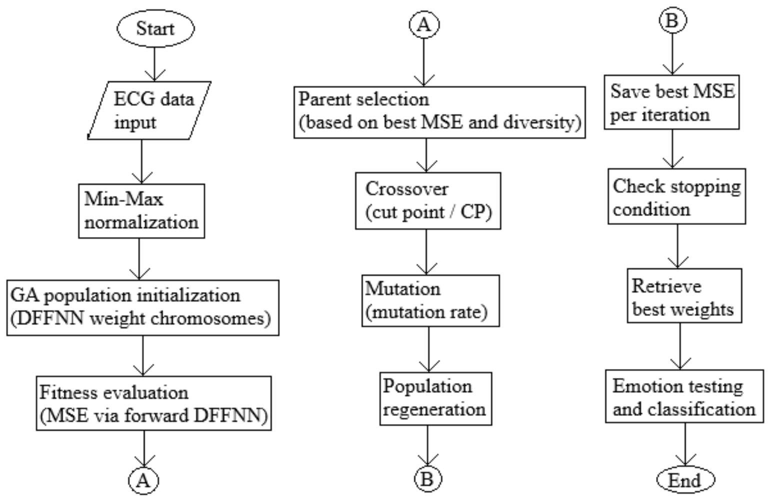

Figure 5. Pure DFFNN+GA flowchart

After evaluating all chromosomes, the parent selection stage is performed. Parent-1 is selected as the chromosome with the lowest MSE in the population, while Parent-2 is chosen from other high-quality chromosomes that exhibit an MSE difference of at least ≥ 0.02 compared to Parent-1. This selection strategy is designed to maintain population diversity and reduce the risk of premature convergence. If no chromosome satisfies this difference criterion, Parent-2 is assigned as the second-best chromosome in the population.

The next step applies single-point crossover based on the predefined CP value. Genes located before the crossover point are inherited directly from their respective parents, whereas genes after the crossover point are exchanged to produce two offspring chromosomes. Subsequently, a mutation process is applied to each gene with a probability of 0.3. When mutation occurs, the gene value is replaced by a new random value within the range -30 to 30; otherwise, the gene remains unchanged from the crossover outcome. The two mutated offspring are then re-evaluated using the same fitness procedure involving forward propagation and MSE computation.

To update the population, a regeneration mechanism is implemented by replacing the two worst-performing chromosomes (highest MSE) with the two newly generated mutant chromosomes. This evolutionary cycle continues iteratively until the best population MSE falls below the defined threshold or the iteration count reaches the maximum limit. During the optimization process, the best MSE value at each iteration is recorded into *mse_log.csv*, enabling convergence analysis and performance monitoring.

After the optimization concludes, the chromosome with the lowest MSE is selected as the final optimal solution, and its genes are used as the final DFFNN weights for classification. The model is then tested on both the training and testing datasets, where the final network output is mapped into emotion classes using predefined decision boundaries: Sad if output < 0.45, Surprise if 0.45 ≤ output ≤ 0.8, and Angry if output > 0.8. Classification accuracy is computed by comparing the predicted emotion labels with the ground-truth labels in both training and testing sets.

2.4. DFFNN+GWO

In the DFFNN+GWO approach, the DFFNN training process does not rely on backpropagation. Instead, all network weights and biases are optimized using the GWO. In this implementation, each wolf represents a candidate solution encoded as a 33-dimensional weight vector ($bobot\_wolf[33]$), which contains the complete set of weights and biases of the adopted DFFNN architecture. The ECG input data consist of 8 extracted features, which are first normalized using feature-wise min-max normalization so that each feature is mapped into the range of 0-1. Subsequently, each wolf undergoes forward propagation on the 21 training samples to produce the network output (S07). The prediction error is computed as the difference between the target value and the network output, defined as $ERR = T - S07$. The fitness value of each wolf is then evaluated using the MSE, calculated as the average squared error over all training samples, where a lower MSE indicates a better solution. Figure 6 illustrates the training mechanisms of Pure DFFNN+GWO model.

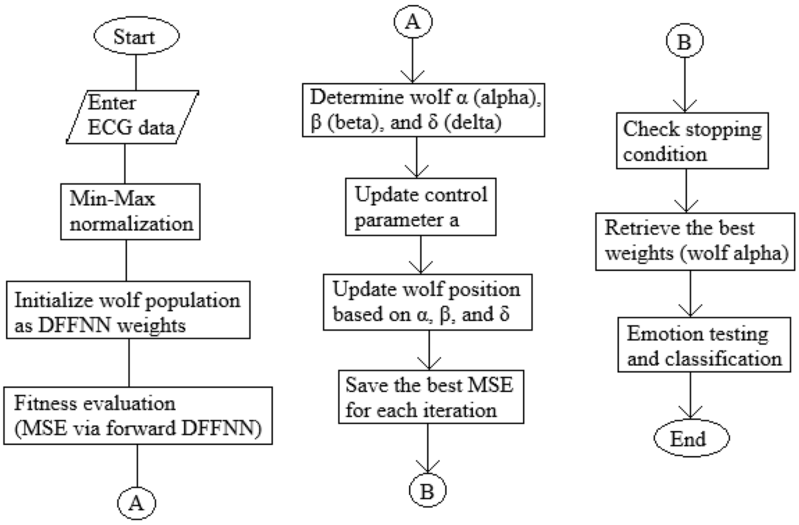

Figure 6. Pure DFFNN+GWO flowchart

At each iteration, the three wolves with the smallest MSE values are selected as the alpha (best), beta (second best), and delta (third best) wolves. These three wolves serve as leaders that guide the search process and determine the movement of the remaining wolves toward the optimal solution. The positions (weight vectors) of all wolves excluding alpha, beta, and delta are updated using the standard GWO position update mechanism governed by the control parameter a, which decreases linearly from 2 to 0 as the iterations progress. For each weight component j, the distances to the alpha, beta, and delta wolves are computed, and three candidate positions (X1, X2, and X3) are generated using random coefficients (r1 and r2) that produce the intermediate parameters A and C. The updated weight value is then obtained as the average of these three candidate positions, enabling each wolf to adjust its position based on the combined influence of the alpha, beta, and delta leaders. This iterative optimization continues until the stopping criterion is met, namely when the best MSE ≤ 0.001 or when the iteration reaches the maximum limit of 5,000,000 iterations. During the optimization process, the best MSE at each iteration (*i.e.*, the alpha wolf's MSE) is recorded into *mse_log.csv* to generate the convergence curve for performance analysis.

2.5. DFFNN+PSO

In the DFFNN+PSO method, the training process of the DFFNN is performed by optimizing the network weights using PSO rather than conventional backpropagation. In this framework, each particle represents a candidate solution encoded as a 33-dimensional DFFNN weight vector ($posisi[33]$), including bias terms, and is associated with a corresponding velocity vector ($velocity[33]$) that controls the update step in the search space. The ECG input data, consisting of 8 extracted features, are first normalized using feature-wise min-max normalization (row-wise in the code) to scale all values into the range of 0-1, thereby improving numerical stability during the optimization process. The PSO swarm is initialized with 10 particles, where the initial weight positions are randomly generated within -1 to 1, while the initial velocities are randomly assigned within -0.1 to 0.1. The optimization parameters used in the implementation are c1=1.5 and c2=1.5, with a target error of MSE=0.0001 and a maximum iteration limit of 5,000,000. Figure 7 illustrates the training mechanisms of Pure DFFNN+PSO model.

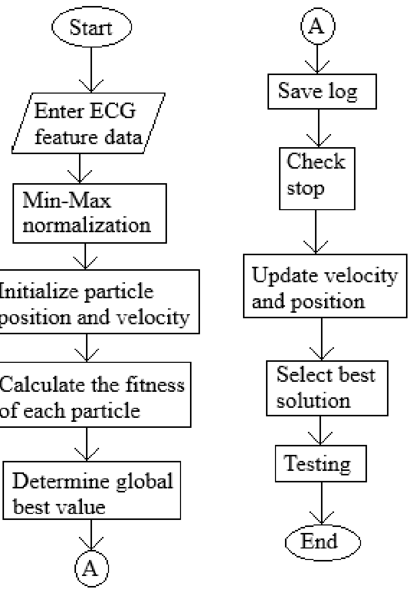

Figure 7. Pure DFFNN+PSO flowchart

At each iteration, every particle is evaluated by applying its current weight vector through a DFFNN forward propagation procedure on the 21 training samples to produce the network output (S07). The prediction error is computed as ERR=T - S07, and the particle's fitness is quantified using the MSE averaged across all training samples. Thus, particles with lower MSE values indicate better candidate weight configurations. Once the MSE is obtained, the algorithm updates each particle's personal best (p_best) as the best solution it has achieved so far, and determines the global best (g_best) as the best p_best among all particles in the swarm. The main PSO optimization step is then conducted by updating the particle velocity and position based on the combined influence of the cognitive and social components. Specifically, the velocity is updated using random coefficients r1 and r2, and the differences between the current position and both p_best and g_best, while the new position is obtained by adding the updated velocity to the current position. The best MSE value (g_best) at each iteration is recorded into mse_log.csv to generate the convergence profile. The iterative process is terminated once the best MSE satisfies the predefined target threshold or the iteration count reaches the maximum limit.

After the optimization stage is completed, the final weight configuration used for classification is taken from the p_best vector of the global best particle, and the trained DFFNN is evaluated on the entire dataset of 27 samples (21 training samples and 6 testing samples). The final network output S07 is converted into discrete emotion classes based on predefined decision rules implemented in the program, namely Sad if $S07 < 0.45$, Surprised if $0.45 \leq S07 \leq 0.80$, and Angry if $S07 > 0.80$. The classification accuracy is computed by comparing the predicted emotion labels with the corresponding ground-truth labels (T2), separately for the training and testing sets. This procedure enables a consistent performance evaluation of the DFFNN+PSO model and supports a fair comparison with the other metaheuristic-based variants examined in this study.

2.6. DFFNN+GA+GWO

In the proposed DFFNN+GA+GWO (multi-stage metaheuristic) approach, the training process of the DFFNN is conducted through two sequential optimization stages. The first stage employs a GA to identify a high-quality initial set of network weights, and the second stage applies the GWO to further refine the GA-derived solution. Similar to the other models, the ECG input data consist of eight extracted features, which are normalized using feature-wise min-max scaling to ensure that all features fall within the range of 0-1. The DFFNN contains 33 trainable weight parameters, including bias terms; therefore, each GA chromosome and each GWO wolf represents a 33-dimensional candidate weight vector.

During the GA phase, an initial population of 10 chromosomes is generated, where each chromosome contains 33 weight genes randomly initialized within the range of -30 to 30. Each chromosome is evaluated by performing forward propagation through the DFFNN across the 21 training samples, followed by computing the prediction error as ERR=T - S07. The fitness of each chromosome is then quantified using the MSE. The selection mechanism identifies the chromosome with the smallest MSE as Parent-1, while Parent-2 is chosen using a ranking-based strategy that considers the MSE difference from the best individual, ensuring that the selected parents are not overly similar. A single-point crossover is then performed using a crossover point of CP=10 to produce two offspring. Mutation is subsequently applied with a probability of 0.3, where selected genes are replaced by newly generated random values. The resulting mutant offspring are re-evaluated in terms of MSE, and a regeneration step is carried out by replacing the two worst-performing chromosomes in the population with the two mutants. This GA evolution process continues until the best MSE falls below 0.0001 or the maximum number of iterations is reached, and the convergence trajectory is recorded in mse_log_GA.csv. Figure 8 illustrates the training mechanisms of Pure DFFNN+GA+GWO model.

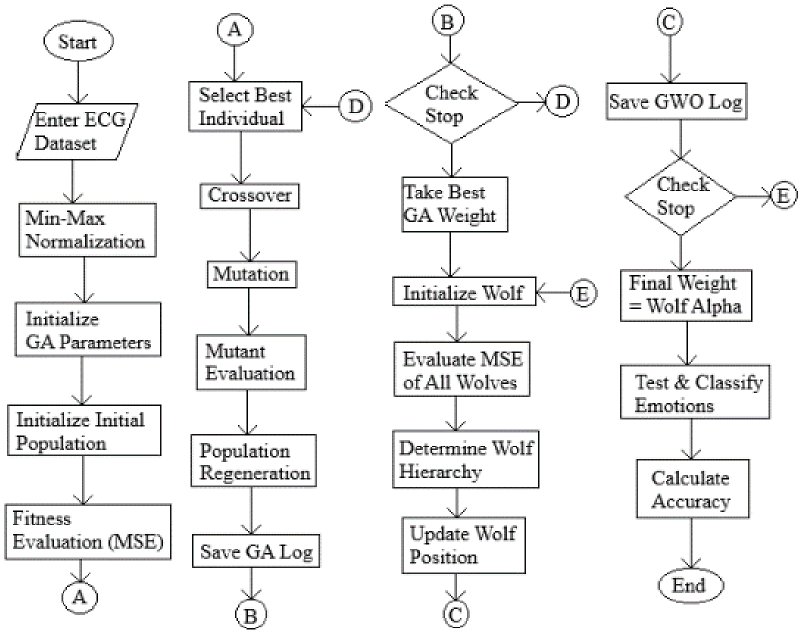

Figure 8. Pure DFFNN+GA+GWO flowchart

Once the GA stage converges, the best chromosome is extracted as the optimal GA weight vector and used as the initialization for the subsequent GWO stage. In the GWO phase, a pack of 10 wolves is formed, where the first wolf (initial alpha) is directly assigned the GA-best weights, while the remaining wolves are generated by adding a small random perturbation (noise) to the GA solution. This strategy preserves the strong starting point found by GA while maintaining sufficient diversity for exploration. Each wolf is then evaluated using the same DFFNN forward propagation scheme on the 21 training samples to compute its MSE. Based on the MSE ranking, the three best wolves are selected as alpha, beta, and delta. The wolf position update follows the standard GWO mechanism controlled by coefficients A and C, and the parameter a, which decreases linearly from 2 to 0 as iterations progress. This design enables wider exploration in early iterations and gradually shifts toward exploitation around the best solutions. The GWO process terminates when the alpha wolf achieves an MSE ≤ 0.001 or the maximum number of iterations is reached, and the optimization progress is logged in *mse_log_GWO.csv*.

In the final evaluation stage, the emotion classification is performed using the best weight vector obtained from GWO, specifically the weights of the alpha wolf in the final iteration. The trained model is then tested on all 27 samples (21 training samples and 6 testing samples) to generate the output value S07. The continuous output is subsequently mapped into discrete emotion classes using the decision rules implemented in the program, namely Sad if S07 < 0.45, Surprised if 0.45 ≤ S07 ≤ 0.80, and Angry if S07 > 0.80. Classification accuracy is computed by comparing the predicted emotion label against the ground-truth class labels (T2), reported separately for the training and testing sets. Overall, the DFFNN+GA+GWO model represents a multi-stage optimization strategy, where GA provides a robust global search for promising initial solutions, and GWO further enhances solution quality through fine-grained exploitation, leading to a more stable convergence behavior and improved predictive performance. Although the number of iterations used in this study is relatively large, this design is intentionally adopted to ensure that all optimization algorithms reach a stable convergence region. This approach enables a fair comparison of final performance and convergence behavior across all models.

3. RESULTS AND DISCUSSION

This study evaluates the performance of a DFFNN under various single-stage and multi-stage metaheuristic optimization strategies for ECG-based emotion classification. Five models were compared, namely Pure DFFNN, DFFNN with GA, DFFNN with GWO, DFFNN with PSO, and the multi-stage model DFFNN+GA+GWO. All models were assessed using six independent experimental trials to ensure that the results were not dependent on a single initialization setting but instead reflected the general behavior of each optimization approach.

From Table 2 based on the results summarized, Pure DFFNN exhibited the lowest performance in terms of both training and testing accuracy, showing a relatively high final MSE and a large variation across repeated trials. These findings indicate that gradient-based learning alone is insufficient to effectively handle the complexity of ECG signals, which are inherently nonlinear, non-stationary, and highly contaminated by noise. In Pure DFFNN, the optimization process tends to become trapped in local minimum, and its performance is strongly dependent on the initial weight configuration.

Table 2. Performance comparison of five DFFNN models across six experimental runs (final MSE, training accuracy, and testing accuracy)

Test	Pure DFFNN			DFFNN+GA		
	Final MSE	Training Accuracy (%)	Test Accuracy (%)	Final MSE	Training Accuracy (%)	Test Accuracy (%)
1	0.08191	61.90	33.33	0.04621	71.43	33.33
2	0.07462	66.67	50.00	0.04388	66.67	50.00
3	0.08977	57.14	33.33	0.03977	76.19	33.33
4	0.07654	61.90	50.00	0.04163	71.43	50.00
5	0.08321	61.90	33.33	0.03892	76.19	50.00
6	0.07866	66.67	50.00	0.04011	71.43	33.33

Test	DFFNN+PSO		
	Final MSE	Training Accuracy (%)	Test Accuracy (%)
1	0.03478	76.19	33.33
2	0.03166	80.95	50.00
3	0.02984	80.95	33.33
4	0.03312	76.19	50.00
5	0.03077	80.95	50.00
6	0.03291	76.19	33.33

Test	DFFNN+GWO			DFFNN+GA+GWO		
	Final MSE	Training Accuracy (%)	Test Accuracy (%)	Final MSE	Training Accuracy (%)	Test Accuracy (%)
1	0.02411	80.95	50.00	0.01721	80.95	50.00
2	0.02177	85.71	50.00	0.01934	76.19	33.33
3	0.01988	85.71	33.33	0.01688	85.71	50.00
4	0.02231	80.95	50.00	0.01877	80.95	50.00
5	0.01894	85.71	50.00	0.01599	85.71	33.33
6	0.02063	80.95	33.33	0.01743	80.95	50.00

In general, the application of single-stage metaheuristic optimization improved these limitations. The DFFNN+GA model demonstrated a rapid MSE reduction during the early iterations, highlighting GA's strong capability for global exploration in the solution space. However, despite its effectiveness in reducing the initial error, the final performance still showed noticeable variation across trials. This reflects the highly stochastic nature of GA, where aggressive exploration is not always accompanied by a stable exploitation phase that consistently refines the solution.

The DFFNN+PSO model also exhibited accelerated convergence during the early iterations, and in several trials, it converged faster than GA. Nevertheless, the MSE trend over iterations revealed a tendency toward early stagnation, where MSE reduction slowed significantly and remained nearly constant for long iteration intervals. This behavior suggests that PSO is susceptible to premature convergence, particularly when the swarm particles lose diversity and converge too quickly toward suboptimal regions.

In contrast, DFFNN+GWO presented a more stable convergence pattern. Although the MSE decreased more gradually, the reduction was consistent, with smaller fluctuations across iterations and across repeated trials. This behavior aligns with the core mechanism of GWO, which balances exploration and exploitation through the hierarchical roles of alpha, beta, and delta wolves, leading to a more controlled and reliable search process.

The most significant improvement was achieved by the multi-stage model DFFNN+GA+GWO. Based on both accuracy and final MSE values, this model consistently produced the lowest MSE and the highest testing accuracy among all evaluated methods. The MSE trajectory clearly demonstrated two distinct convergence phases. During the initial iterations up to approximately five million iterations GA played the dominant role by performing global exploration and moving the solution into a substantially lower-error region compared to Pure DFFNN and other single-stage approaches. Subsequently, GWO took over the optimization process and emphasized local exploitation, refining the solution obtained by GA until a more stable convergence state was achieved.

From Table 3, a deeper analysis of delta MSE versus iteration plots, provides further insight into the stability of the learning process. Delta MSE reflects the change in error across consecutive iterations and serves as an important indicator of optimization consistency. Pure DFFNN showed highly fluctuating delta MSE values with large spikes, suggesting unstable learning dynamics. For single-stage optimizers, delta MSE spikes were still observed especially for GA and PSO although their amplitudes were generally smaller than those found in Pure DFFNN.

Table 3. Summary of final iteration and best MSE across six independent runs for each optimized DFFNN model

Model	Test	Final Iteration	Best MSE	Model	Test	Final Iteration	Best MSE
DFFNN+GA	1	4,990,394	0.026631	DFFNN+PSO	1	5,000,000	0.108127
DFFNN+GA	2	4,903,368	0.018160	DFFNN+PSO	2	5,000,000	0.099371
DFFNN+GA	3	4,653,179	0.016793	DFFNN+PSO	3	5,000,001	0.095893
DFFNN+GA	4	4,967,074	0.023963	DFFNN+PSO	4	5,000,000	0.102500
DFFNN+GA	5	4,528,046	0.025787	DFFNN+PSO	5	5,000,000	0.110850
DFFNN+GA	6	4,718,210	0.018772	DFFNN+PSO	6	5,000,001	0.071233
DFFNN+GWO	1	4,923,090	0.055048	DFFNN+GA+GWO	1	9,925,843	0.021872
DFFNN+GWO	2	4,900,205	0.055113	DFFNN+GA+GWO	2	9,991,248	0.014286
DFFNN+GWO	3	4,980,551	0.060624	DFFNN+GA+GWO	3	9,992,644	0.016090
DFFNN+GWO	4	4,903,109	0.070780	DFFNN+GA+GWO	4	9,994,966	0.021654
DFFNN+GWO	5	4,999,946	0.029271	DFFNN+GA+GWO	5	9,999,852	0.025134
DFFNN+GWO	6	4,905,962	0.057766	DFFNN+GA+GWO	6	9,925,819	0.028446

Conversely, DFFNN+GWO exhibited more controlled delta MSE patterns, characterized by relatively small and consistent error changes. The DFFNN+GA+GWO model showed the highest stability: after the initial GA exploration phase, delta MSE gradually decreased and rarely presented large spikes. This indicates that the obtained solution was not only quantitatively better, but also achieved through a more stable and reliable optimization process.

Statistical analysis of delta MSE further strengthens these observations. The mean delta MSE of the hybrid model was lower than that of the single-stage optimizers, while the small median suggests that most iterations produced minimal error changes. In addition, the narrow interquartile range (IQR) of the DFFNN+GA+GWO model indicates low variability during optimization, supporting the conclusion that its convergence process is more reliable and reproducible. In other words, the model not only achieves a superior solution, but also reaches it in a consistent manner.

From Table 4, the advantage of this multi-stage approach can be explained by the complementary roles of the two optimizers. GA functions as an effective global search mechanism to prevent entrapment in local minima, whereas GWO acts as a stable refinement strategy through local exploitation. This combination yields a balanced exploration-exploitation trade-off that is difficult to achieve using single-stage optimization alone. Given the complexity of ECG signals, this strategy proves highly effective in improving model generalization performance. In addition to MSE and classification accuracy, convergence stability was analyzed using statistical indicators such as mean, standard deviation, and IQR of MSE values across multiple runs.

Overall, the results confirm that metaheuristic optimization is crucial for enhancing DFFNN performance in ECG-based emotion classification. However, the multi-stage optimization strategy provides additional benefits in terms of convergence stability and learning reliability. Therefore, the DFFNN+GA+GWO model can be considered the most effective solution in this study, both in terms of final performance and optimization dynamics.

To further validate the performance differences among the evaluated models, a statistical significance analysis was conducted using paired t-tests across six independent runs. The results indicate that the performance improvement of the DFFNN+GA+GWO model over the Pure DFFNN and single-stage metaheuristic models is statistically significant ($p < 0.05$). In addition, the hybrid model demonstrates lower variance in final MSE values, indicating more consistent convergence behavior compared to other approaches.

Table 4. Statistical summary of final MSE and convergence iterations across six runs for each optimized DFFNN model

Model	Final MSE (Avr)	Std	Best MSE (Min)	Worst MSE (Max)	Average of Last Iteration
DFFNN+GA	0.021684	0.004274	0.016793	0.026631	4,793,379
DFFNN+GWO	0.054767	0.014876	0.029271	0.070780	4,935,477
DFFNN+PSO	0.097995	0.013213	0.071233	0.110850	5,000,000
DFFNN+GA+GWO	0.021247	0.006788	0.014286	0.028446	9,971,862

4. CONCLUSION

This study systematically evaluated the performance of single-stage and multi-stage metaheuristic optimization methods for training DFFNN in ECG-based emotion classification. The experimental results clearly demonstrate that conventional gradient-based Pure DFFNN exhibits higher error values, with final MSE ranging from 0.07462 to 0.08977 across six trials. The incorporation of metaheuristic optimization significantly improves performance. Among single-stage methods, DFFNN+GWO achieved MSE values between 0.01894 and 0.02411, outperforming GA and PSO in terms of stability and error reduction. The proposed multi-stage DFFNN+GA+GWO approach yielded the lowest observed MSE of 0.014286 and an average MSE of approximately 0.0212, indicating superior convergence refinement after global exploration by GA followed by local exploitation by GWO.

Training accuracy improved substantially from 57.14%–66.67% (Pure DFFNN) to 80.95%–85.71% under metaheuristic optimization. Although testing accuracy remained within 33.33%–50.00%, likely due to limited dataset size, the convergence analysis confirms that multi-stage optimization provides more stable learning dynamics and reduced MSE fluctuations. Overall, the findings highlight that combining metaheuristic strategies in a multi-stage framework enhances error minimization and convergence stability, making it a promising optimization approach for neural network-based emotion classification using ECG signals, particularly in small-sample biomedical datasets. In addition to MSE and classification accuracy, convergence stability was analyzed using statistical indicators such as mean, standard deviation, and IQR of MSE values across multiple runs. Future work will focus on expanding the dataset size and evaluating cross-subject generalization to further validate the robustness of the proposed multi-stage optimization framework.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Giovanni Dimas Prenata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ahmad Ridho'i									✓	✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

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.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] J. A. Healey and R. W. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, Jun. 2005, doi: 10.1109/TITS.2005.848368.

[2] S. Koelstra *et al.*, “DEAP: A database for emotion analysis; Using physiological signals,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012, doi: 10.1109/T-AFFC.2011.15.

[3] R. W. Picard, “Affective computing: Challenges,” *International Journal of Human Computer Studies*, vol. 59, no. 1–2, pp. 55–64, 2003, doi: 10.1016/S1071-5819(03)00052-1.

[4] H. Zacarias, J. A. L. Marques, V. Felizardo, M. Pourvahab, and N. M. Garcia, “ECG forecasting system based on long short-term memory,” *Bioengineering*, vol. 11, no. 1, p. 89, Jan. 2024, doi: 10.3390/bioengineering11010089.

[5] F. Huang, T. Qin, L. Wang, and H. Wan, “Hybrid prediction method for ECG signals based on VMD, PSR, and RBF neural network,” *BioMed Research International*, vol. 2021, no. 1, p. 6624298, Jan. 2021, doi: 10.1155/2021/6624298.

[6] Q. Xiao *et al.*, “Deep learning-based ECG arrhythmia classification: A systematic review,” *Applied Sciences*, vol. 13, no. 8, p. 4964, Apr. 2023, doi: 10.3390/app13084964.

[7] Y. Tian and Q. Jia, “Artificial neural network,” *Encyclopedia of Quality of Life and Well-Being Research*, no. March, p. 273, 2024, doi: 10.1007/978-3-031-17299-1_300217.

[8] A. C. I. Goodfellow and Y. Bengio, “Deep learning,” *MIT Press*, vol. 521, no. 7553, p. 785, 2016.

[9] V. Satheeswaran, G. N. Chandrika, A. Mitra, R. Chowdhury, P. Kumar, and G. E., “Deep learning based classification of ECG signals using RNN and LSTM mechanism,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics (JEEEMI)*, vol. 6, no. 4, pp. 332–342, 2024, doi: 10.35882/jeeemi.v6i4.496.

[10] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, “Efficient BackProp,” in *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, Vol. 1524., Springer, Berlin, Heidelberg, 1998, pp. 9–50.

[11] X.-S. Yang, *Nature-inspired optimization algorithms*, 2nd ed. Academic Press, Elsevier, 2021.

[12] E. F. P. da Luz, J. C. Becceneri, S. Stephany, H. F. de Campos Velho, and A. J. da Silva Neto, “Particle swarm optimization,” in *Computational Intelligence Applied to Inverse Problems in Radiative Transfer*, 2023, pp. 95–106.

[13] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer,” *Advances in Engineering Software*, vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.

[14] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*. MA, United States: Addison-Wesley Publishing Company, Inc., 1988.

[15] R. Poli, J. Kennedy, and T. Blackwell, “Particle swarm optimization,” *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, Oct. 2007, doi: 10.1007/s11721-007-0002-0.

[16] S. Mirjalili, “How effective is the grey wolf optimizer in training multi-layer perceptrons,” *Applied Intelligence*, vol. 43, no. 1, pp. 150–161, 2015, doi: 10.1007/s10489-014-0645-7.

[17] P. S. Addison, “Wavelet transforms and the ECG: a review,” *Physiological Measurement*, vol. 26, no. 5, pp. R155–R199, Oct. 2005, doi: 10.1088/0967-3334/26/5/R01.

[18] U. R. Acharya *et al.*, “Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: A comparative study,” *Information Sciences*, vol. 377, pp. 17–29, 2017, doi: 10.1016/j.ins.2016.10.013.

[19] A. K. Singh and S. Krishnan, “ECG signal feature extraction trends in methods and applications,” *BioMedical Engineering Online*, vol. 22, no. 1, 2023, doi: 10.1186/s12938-023-01075-1.

[20] E.-G. Talbi, *Metaheuristics: From design to implementation*. United States of America: John Wiley & Sons, Inc., Hoboken, New Jersey, 2009.

[21] G. D. Prenata, A. Ridho’i, and M. R. Arshad, “Comparative evaluation of LSTM and metaheuristic-optimized neural networks for ECG prediction under limited data conditions,” *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 8, no. 2, pp. 751–768, 2026, doi: 10.35882/jeeemi.v8i2.1524.

[22] M. S. Shaikh *et al.*, “An intelligent hybrid grey wolf-particle swarm optimizer for optimization in complex engineering design problem,” *Scientific Reports*, vol. 15, no. 1, p. 18313, May 2025, doi: 10.1038/s41598-025-02154-0.

[23] G. D. Prenata, E. S. Pane, A. D. Wibawa, and M. H. Purnomo, “Analysis of negative emotion using HRV based ECG signal of elder people,” in *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, 2017, vol. 2018-January, pp. 444–449, doi: 10.1109/ICITISEE.2017.8285547.

[24] S. D. Pressman and S. Cohen, “Does positive affect influence health?,” *Psychological Bulletin*, vol. 131, no. 6, pp. 925–971, Nov. 2005, doi: 10.1037/0033-2909.131.6.925.

[25] S. D. Kreibig, “Autonomic nervous system activity in emotion: A review,” *Biological Psychology*, vol. 84, no. 3, pp. 394–421, Jul. 2010, doi: 10.1016/j.biopsycho.2010.03.010.

BIOGRAPHIES OF AUTHORS

Giovanni Dimas Prenata earned his bachelor's degree in electrical engineering, specializing in power systems, from Petra Christian University in Surabaya in 2009. His undergraduate research focused on energy management and efficiency enhancement, carried out as an applied study at PT Tri Tunggal Manunggal Sejati, where he evaluated operational performance and formulated recommendations to optimize energy consumption. Driven by a strong interest in intelligent systems and signal processing, he continued his academic journey by completing a master's degree in electrical engineering with a specialization in telematics at Institut Teknologi Sepuluh Nopember (ITS) Surabaya, graduating in 2018. His master's thesis explored the use of support vector machine (SVM) techniques for electrocardiogram (ECG) signal classification to detect and analyze negative emotional states in elderly individuals, combining principles of biomedical signal processing and machine learning. In addition to his formal education, he actively participates in interdisciplinary research, with particular attention to artificial intelligence, biomedical signal analysis, and computational modeling. His current research endeavors focus on designing intelligent algorithms in C++ for real-world applications in both energy and healthcare domains. Through this continuous line of research, he seeks to connect theoretical advancements with practical engineering solutions, contributing to the progress of data-driven approaches for intelligent decision-making and predictive system development. He can be contacted at email: gprenata@untag-sby.ac.id.

Ahmad Ridho'i is a graduate of the 1995 cohort of the electrical engineering undergraduate program with a specialization in electronics at Universitas 17 Agustus 1945 Surabaya. He completed his bachelor's studies with a thesis focused on the design and realization of a speech aid system for individuals with speech disabilities, employing an IBM personal computer as the main processing unit, which reflected his early interest in human–machine interaction and assistive technology development. He subsequently pursued a master's degree in electrical engineering with a concentration in electrotechnics at the Bandung Institute of Technology (ITB), completing his studies in 2000. His master's research centered on the development of an intelligent control strategy to optimize elevator service queuing systems, combining control engineering principles with computational intelligence to enhance the performance of automated systems. Over the course of his academic and professional career, he has consistently engaged in research on disaster mitigation and early warning systems. Given Indonesia's position as an archipelagic nation along the pacific ring of fire, where natural hazards such as earthquakes and floods occur frequently, his research efforts focus on the design and implementation of early disaster detection and monitoring technologies. Through this work, he aims to support the creation of community-oriented systems capable of reducing material damage and loss of life, thereby strengthening societal resilience and disaster preparedness through technological innovation. He can be contacted at email: ridhoi@untag-sby.ac.id.