

RESEARCH ARTICLE

Deep VMD-attention network for arrhythmia signal classification based on Hodgkin-Huxley model and multi-objective crayfish optimization algorithm

Hang Zhao, Xiongfei Yin^{1*}

School of Physics and Optoelectronic Engineering, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, China

* yinxiongfei888@163.com



OPEN ACCESS

Citation: Zhao H, Yin X (2025) Deep VMD-attention network for arrhythmia signal classification based on Hodgkin-Huxley model and multi-objective crayfish optimization algorithm. PLoS ONE 20(5): e0321484. <https://doi.org/10.1371/journal.pone.0321484>

Editor: Jinran Wu, University of Queensland - Saint Lucia Campus: The University of Queensland, AUSTRALIA

Received: December 22, 2024

Accepted: March 05, 2025

Published: May 14, 2025

Copyright: © 2025 Zhao, Yin. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://physionet.org/content/mitdb/1.0.0/>.

Funding: X.Y. receives the found by the Hangzhou Agricultural and Social Development Initiative Design Project (Grant numbers 2022ZDSJ0846). The funder's website: <https://kj.hangzhou.gov.cn/> Funders did not play

Abstract

Recent research for arrhythmia classification is increasingly based on AI-driven approaches, which are primarily grounded in ECG data, but often neglect the mathematical foundations of cardiac electrophysiology. A finite element model (FEM) of the human heart, grounded in the Hodgkin-Huxley (HH) model was established to simulate cardiac electrophysiology, and ECG signals from 200 representative points were acquired. Two types of arrhythmia characterized by significant anomalies in the variables of the HH model were simulated, and corresponding synthetic ECG signals were generated. A multi-objective optimization method based on non-dominated sorting was integrated into the crayfish optimization algorithm (MOCOA). To optimize the key parameters K and α in variational mode decomposition (VMD), a MOCOA-VMD technique specifically tailored for ECG signal processing was developed. The Pareto optimal front was generated using MOCOA with the indicators of spectral kurtosis and KL divergence, by which the optimal intrinsic mode functions were obtained. A deep VMD-attention network based on MOCOA was developed for ECG signal classification. The ablation study evaluated the effectiveness of the proposed signal decomposition method and deep attention modules. The model based on MOCOA-VMD achieves the highest accuracy of 94.46%, outperforming models constructed using EEMD, VMD, CNN and LSTM modules. Bayesian optimization was employed to fine-tune the hyperparameters and further enhance the performance of the deep model, with the best accuracy of the deep attention model after TPE optimization reaching 96.11%. Moreover, the real-world MIT-BIH arrhythmia database was utilized for further validation to prove the robustness and generalizability of the proposed model. The proposed deep VMD-attention modeling and classification strategy has shown significant promise and may offer valuable inspiration for other signal processing fields as well.

any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Cardiovascular disease (CVD) is a leading global cause of morbidity and mortality, responsible for more than 31% of worldwide deaths in 2020 [1]. Cardiac arrhythmias, defined as deviations from the normal heart rate or rhythm that cannot be explained by physiological factors [2], represent a significant subset of CVD. These arrhythmias can present as irregular heartbeats, excessively fast heart rates (tachycardia), or abnormally slow heart rates (bradycardia), which can be benign or indicate serious underlying heart conditions. Timely and precise identification of arrhythmias not only facilitates appropriate therapeutic interventions but also aids in preventing potential complications and improving patient outcomes [3].

The traditional rule-based approach to diagnosis often struggles with the complexity and heterogeneity of large-scale medical data, necessitating intensive analysis and substantial medical expertise to achieve reliable diagnostic outcomes. In recent years, the application of artificial intelligence (AI) in electrocardiogram (ECG) interpretation has become a transformative development within the cardiovascular domain. Leveraging vast repositories of digitized clinical ECG data, researchers have developed sophisticated AI models designed to identify various cardiac conditions, such as left ventricular dysfunction, asymptomatic atrial fibrillation, and hypertrophic cardiomyopathy. Additionally, these models can accurately predict demographic and phenotypic features, including age, gender, race, and other phenotypes [4–6]. The integration of AI into ECG analysis not only improves diagnostic precision but also enables timely and targeted interventions, ultimately leading to better patient care and improved health outcomes.

Recent advancements in artificial intelligence, particularly deep learning models such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks [7–9], have revolutionized the field of ECG interpretation [10]. These AI-driven approaches have demonstrated exceptional accuracy in detecting nuanced signals and patterns that are typically undetectable by human experts, elevating the ECG to a highly effective, non-invasive biomarker for cardiovascular diagnostics. Despite these achievements, the classification of arrhythmias remains a formidable challenge due to the diverse and complex etiologies underlying ECG signal abnormalities. These abnormalities can be attributed to a multitude of factors, including coronary heart disease, heart failure, cardiac pacing, hypertension, cardiomyopathy, heart valve disease, and electrolyte imbalances. The heterogeneity of these contributing factors significantly complicates the task of accurately classifying arrhythmias. Furthermore, the combination of variational mode decomposition (VMD) and machine learning is an emerging and promising approach for analyzing and modeling complex signals, particularly in time-series analysis and predictive modeling. [11] VMD is a signal decomposition technique that decomposes a signal into intrinsic mode functions (IMFs), while machine learning excels at learning complex patterns from data. Integrating VMD and machine learning can leverage the strengths of both methods to improve model performance and interpretability.

The majority of existing deep-learning techniques for arrhythmia classification are primarily based on ECG data, often neglecting the fundamental principles of cardiac electrophysiology [3]. This limitation underscores the need for a more integrated approach that combines the strengths of AI with a solid understanding of the underlying physiological mechanisms. In this context, this paper seeks to bridge this gap by incorporating the mathematical foundations of cardiac electrophysiology into AI-driven ECG analysis, thereby enhancing the accuracy and interpretability of arrhythmia classification.

This paper presents a framework for modeling mechanisms in arrhythmia signal classification, organized as follows. In Theoretical foundations for electrocardio signal modeling and

processing section, the theoretical foundations for electrocardio signal modeling and processing were elaborated, including Hodgkin-Huxley model, finite element method and variational mode decomposition. The finite element model and various arrhythmia models induced by abnormalities of PDEs were established in Modeling of the human heart and arrhythmia electrocardiosignal acquisition section, and the typical synthetic signals of the models were prepared. In Variational Mode Decomposition based on Multi-objective COA section, the multi-objective crayfish optimization algorithmic variational mode decomposition was proposed and the optimal IMFs were obtained. The deep attention model for classification was constructed in Learning model establishment and optimization for classification section, the modules were validated by ablation studies and the hyperparameters were optimized by the Bayesian algorithm. Finally, the Conclusion section offered the discussions and concluding remarks.

Related works

As noted in the preceding section, a number of AI-based ECG classification methods have been presented in the literature.

Shu Lih Oh et al. [7] propose an automated system using a combination of CNN and LSTM for the diagnosis of normal sinus rhythm, left bundle branch block (LBBB), right bundle branch block (RBBB), atrial premature beats (APB) and premature ventricular contraction (PVC) on ECG signals. The model achieves an accuracy of 98.10%, sensitivity of 97.50% and specificity of 98.70% using a ten-fold cross validation strategy. A similar architecture of the deep model consisting of CNN and LSTM can be seen in [8], which achieved an average accuracy of 97.15%.

Sumanta Kuila et al. [12,13] utilize leading algorithms such as K-nearest neighbor, artificial neural network, and support vector machine to classify different features of ECG signals. In their work, a novel classification algorithm is proposed based on ELM (Extreme Learning Machine) with Recurrent Neural Network (RNN) by using morphological filtering. The experimental results with the MIT-BIH database, using hidden neurons of ELM with RNN, show an accuracy of 96.41%, sensitivity of 93.62% and specificity of 92.66%.

Sayli Siddhasanjay Aphale et al. [14] proposed a novel convolutional neural network named ArrhyNet for MIT-BIH arrhythmia classification. Low pass filter and baseline wander filter, along with the Synthetic Minority Over Sampling (SMOTE) technique are utilized to enhance the performance. The results indicate that the top-1 accuracy of the fiveclass classification system for the database used is 92.73%.

Sadegh Ilbeigipour et al. [15] developed and compared three decision trees, random forest, and logistic regression classifiers, respectively. The performance of the random forest classifier was much higher than that of the decision trees and logistic regression. The model was tested on the MIT-BIH database and achieved 88% accuracy for 3-class classifications.

Qi Meng et al. [16] propose a multi-database integration methodology and a heartbeat self-processing method to eliminate data differences. A 16-classification 5-layer fully connected neural network was built and verified based on the integrated three databases Hercules-3, achieving an accuracy rate of up to 98.67%. Hany El-Ghaish et al. [17] develop a deep learning framework called ECGTransForm by embedding a novel Bidirectional Transformer (BiTrans) mechanism, ensuring a robust spatial feature extraction across various granularities.

Currently, most AI-driven models for arrhythmia classification focus primarily on ECG data, ignoring the fundamentals of cardiac electrophysiology. As a result, this limitation highlights the need for a more integrated approach that combines AI's strengths with a solid understanding of physiological mechanisms. A precise mathematical model of the ECG needs

to be developed, and the accuracy of predictions made by deep models will also significantly improve. Using the mathematical foundations of cardiac electrophysiology, this paper aims to bridge the gap between electrophysiology and AI-driven ECG analysis.

Theoretical foundations for electrocardio signal modeling and processing

Modelling theory of cardiac electrophysiology

Biological systems exhibit numerous electrical activities in response to external stimuli, which are crucial for neuron-based application in information processing analysis. However, the impact of these electrical activities on the encoding and decoding technologies in neuroscience has not yet been fully understood or well established.

There are various neuron models developed to imitate certain biological neuron dynamics [18–21]. The Hodgkin-Huxley (HH) model is a mathematical model that describes how action potentials in neurons are initiated and propagated. This model was developed by Alan Hodgkin and Andrew Huxley in 1952 based on their experimental work on the giant axon of the squid [22]. The HH model is considered one of the most influential models in neuroscience due to its accuracy in describing the electrical properties of neurons and its ability to explain the underlying mechanisms of neuronal excitability. Subsequently, variants based on the HH model emerged, such as the FitzHugh-Nagumo (FHN) model [18,23], a simplified version of the Hodgkin-Huxley model, which captures the essential features of excitability and action potential generation in neurons [24]. The FHN model reduces the complexity of the full Hodgkin-Huxley model while preserving the key dynamics of spike generation and propagation, which facilitates the analysis of fundamental action potential and excitability dynamics, broadening its applicability for studying excitable systems in a more accessible mathematical framework.

The Hodgkin-Huxley model is described by a set of four coupled nonlinear ordinary differential equations. The membrane potential equation is expressed by

$$C_m \frac{dV}{dt} = I - \bar{g}_{Na} m^3 h (V - E_{Na}) - \bar{g}_K n^4 (V - E_K) - \bar{g}_L (V - E_L) \quad (1)$$

Where V is the membrane potential, I is the external applied current, C_m is the membrane capacitance, \bar{g}_{Na} , \bar{g}_K , \bar{g}_L are the maximum conductances for sodium (Na^+), potassium (K^+), and leak currents, respectively. E_{Na} , E_K , E_L are the reversal potentials for sodium, potassium, and leak currents, respectively.

m , h , and n are gating variables representing the sodium activation, sodium inactivation and potassium activation channels, respectively, whose dynamics can be modeled by gating variable equations

$$\begin{aligned} \frac{dm}{dt} &= \alpha_m(V)(1 - m) - \beta_m(V)m \\ \frac{dh}{dt} &= \alpha_h(V)(1 - h) - \beta_h(V)h \\ \frac{dn}{dt} &= \alpha_n(V)(1 - n) - \beta_n(V)n \end{aligned} \quad (2)$$

where α and β are rate functions that describe the opening and closing rates of the ion channels, which are defined by

$$\alpha_m(V) = \frac{0.1(V + 40)}{1 - e^{-(V+40)/10}}$$

$$\begin{aligned}
\beta_m(V) &= 4e^{-(V+65)/18} \\
\alpha_h(V) &= 0.07e^{-(V+65)/20} \\
\beta_h(V) &= \frac{1}{1 + e^{-(V+35)/10}} \\
\alpha_n(V) &= \frac{0.01(V + 55)}{1 - e^{-(V+55)/10}} \\
\beta_n(V) &= 0.125e^{-(V+65)/80}
\end{aligned} \tag{3}$$

Finite element method for numeric solving

Developed in the early 1960s, the finite element method (FEM) has become one of the most widely used numerical analysis approaches in engineering. By simulating the physical behavior of objects and systems, engineers are able to optimize designs, ensure safety, and reduce development costs [25]. Consider a multidimensional steady-state heat conduction process as an example, which can be described by the Poisson equation with homogeneous boundary conditions

$$\begin{aligned}
-\nabla^2 u &= f \quad \text{in } \Omega, \\
u &= 0 \quad \text{on } \partial\Omega,
\end{aligned} \tag{4}$$

with domain $\Omega \subset \mathbb{R}^d$. Here, u is the unknown function to be solved, and f is the known function (source term). The weak form of Eq 4 can be obtained by choosing a function v from a space U of smooth functions by performing the inner product of both sides with v , i.e.,

$$-\langle \nabla^2 u, v \rangle = \langle f, v \rangle \tag{5}$$

Assume that, in addition to having the necessary smoothness, the functions that are to be the solutions also satisfy the boundary conditions. The space U of test functions is of the form $U = \{v : v \in C^2(\Omega), v = 0 \text{ on } \partial\Omega\}$. More specifically, let $d = 2$, the weak form of Eq 4 can be written as

$$\iint_{\Omega} \nabla u \cdot \nabla v \, dx \, dy = \iint_{\Omega} f v \, dx \, dy. \tag{6}$$

To obtain a numerical method, it requires U to be finite-dimensional with basis $\{u_1, \dots, u_n\}$. Then the approximate solution u^h of Eq 4 can be represented as

$$u^h = \sum_{j=1}^n c_j u_j. \tag{7}$$

where the coefficients c_j are to be determined once a basis has been chosen for the approximation space U . By inserting u^h into the weak form Eq 6, and selecting the basis functions of U as trial functions v , a system of equations is obtained

$$\iint_{\Omega} \nabla \left[\sum_{j=1}^n c_j u_j \right] \cdot \nabla u_i \, dx \, dy = \iint_{\Omega} f u_i \, dx \, dy, \quad i = 1, \dots, n \tag{8}$$

Which is known as the Ritz-Galerkin method and can be written in matrix form, $A\mathbf{c} = \mathbf{b}$, The expression of stiffness matrix A is

$$A_{i,j} = \iint_{\Omega} \nabla u_j \cdot \nabla u_i dx dy \quad (9)$$

Theoretical basis of variational mode decomposition

Variational Mode Decomposition (VMD) [26] represents a recent advancement in signal processing, designed to decompose a multi-component signal f into a series of K quasi-orthogonal intrinsic mode functions (IMFs) u_k . Each IMF exhibits specific sparsity characteristics and limited bandwidth, obtained through a non-recursive decomposition process. The variational formulation and the optimization-based approach make VMD a powerful and flexible tool for multi-component signal analysis and decomposition, with advantages over traditional data-driven methods like Empirical Mode Decomposition (EMD) [27], Ensemble Empirical Mode Decomposition (EEMD) [28], etc., which recursively decomposes a non-stationary signal into IMFs thus facing robustness and mode mixing problems [29].

The essential aspect of the VMD is solving a constrained variational formulation written as

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \text{ s.t. } \sum_k u_k = f \quad (10)$$

Where $\{u_k\} := \{u_1, \dots, u_K\}$ and $\{\omega_k\} := \{\omega_1, \dots, \omega_K\}$ are shortening designations for the set of all modes and their center frequencies, respectively. ∂_t is the partial derivative of the function at the time t , and $\delta(t)$ is the unit impulse function. $\sum_k := \sum_{k=1}^K$ denotes the summation over all modes. In order to render the problem unconstrained, both a quadratic penalty term α and a Lagrangian multiplier λ are considered, and the augmented Lagrangian \mathcal{L} is introduced as

$$\begin{aligned} \mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) := & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (11)$$

Therefore, the original minimization problem Eq 10 can be solved by finding the saddle point of the augmented Lagrangian in a sequence of iterative sub-optimizations known as the alternating direction method of multipliers (ADMM) [30]. After pre-setting the decomposition mode number K and the quadratic penalty term α , the decomposed mode \hat{u}_k and its associated center frequency ω_k^{n+1} , along with the Lagrangian multiplier $\hat{\lambda}$, are initialized. Then the updating strategy of modes $u_k(t)$ for the k -th mode is

$$\hat{u}_k^{n+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i < k} \hat{u}_i^{n+1}(\omega) - \sum_{i > k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (12)$$

where the $\hat{u}_k(\omega)$ is the Fourier transform of the k -th mode, $\hat{f}(\omega)$ is the Fourier transform of the original signal, $\hat{\lambda}(\omega)$ is the Fourier transform of the Lagrangian multiplier, and n is the

iteration index. The center frequencies ω_k are updated to minimize the bandwidth of each mode by

$$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega |\widehat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\widehat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (13)$$

The Lagrangian multiplier $\lambda(t)$ is updated to enforce the constraint that the sum of the modes equals the original signal by

$$\widehat{\lambda}^{n+1}(\omega) \leftarrow \widehat{\lambda}^n(\omega) + \tau \left(\widehat{f}(\omega) - \sum_k \widehat{u}_k^{n+1}(\omega) \right) \quad (14)$$

where τ is the update parameter for the Lagrangian multiplier. The criterion of convergence is

$$\sum_k \frac{\|\widehat{u}_k^{n+1} - \widehat{u}_k^n\|_2^2}{\|\widehat{u}_k^n\|_2^2} < \varepsilon \quad (15)$$

where ε is typically set to 10^{-7} , the value of k is constrained by the complexity of the signal and computational resources, normally less than 10. Finally, the K modes of the original signal are obtained.

The selection of the number of intrinsic mode functions K and the penalty term α is paramount when implementing the variational mode decomposition method. These parameters significantly influence the decomposition process and the quality of the resulting intrinsic mode functions. K influences the granularity of the decomposition, while α controls the smoothness of the resulting modes. Therefore, optimizing the number of intrinsic mode functions and the penalty term in Variational Mode Decomposition is crucial for achieving accurate and meaningful signal decomposition, which balances the need to capture all relevant signal components against the risk of introducing noise or artifacts. Various optimization methods, such as grid search, Bayesian optimization, and meta-heuristic optimization can be used to achieve the best decomposition results [31–33].

Popular meta-heuristic optimization techniques are typically inspired by physical phenomena, species' behaviors, biological evolution, or laws governing social operations [34,35]. Many emergent behavioral patterns in organisms have optimized overtones due to the evolutionary process of natural selection and have been identified separately and used in nature-inspired algorithms. There are plenteous evolutionary algorithms and swarm algorithms built on the foundations of naturally occurring mechanisms, such as Genetic Algorithm (GA) [36], Particle Swarm Optimization (PSO) [37], Gray Wolf Optimization (GWO) [38], Gold Rush Optimizer (GRO) [39], Rime Optimization Algorithm (RIME) [40], Crayfish optimization algorithm (COA) [41], Beluga Whale Optimization (BWO) [42] and so on. Meta-heuristic optimization algorithms are playing an increasingly important role in optimization and decision-making problems in various engineering fields and will also be employed in this paper.

Attention scheme for sequencing signal feature extraction

The Transformer model [43] has proven a great success in sequencing signal feature extraction, making a significant contribution to natural language processing (NLP) tasks, and has become the foundation for many state-of-the-art deep network models, such as BERT [44], GPT [45] and so on. Effectively capturing long-range dependencies and relationships in the

input sequence is made possible by the attention scheme, which is a key component and a powerful mechanism of the Transformer model. Instead of processing the entire input sequence uniformly, the attention mechanism enables the model to selectively focus on relevant parts of the input when generating the output.

The attention scheme in the Transformer model mainly consists of three matrix components, i.e., Query (Q), Key (K), and Value (V), which is linearly transformed by the input sequences. The dot product between Q and K is calculated and is divided by $\sqrt{d_k}$ to prevent the result from being too large, where d_k is the dimension of the Key vector. After implementing the Softmax operation to normalize the results, the probability distribution is obtained, which is finally multiplied by the matrix V to yield the weighted sum representation as the final output, as shown in Eq 16.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (16)$$

The Transformer model utilizes multiple attention heads, each with its own set of Query, Key, and Value matrices, which facilitate the model to capture different types of relationships and dependencies in the input sequence.

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \\ \text{where head}_i &= \text{Attention}(W_i^Q, KW_i^K, VW_i^V) \end{aligned} \quad (17)$$

where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$.

The position-wise feed-forward network performs two linear transformations with a ReLU activation function [46] that can be modeled as

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \quad (18)$$

where W_1 , W_2 are the weights and b_2 is the bias of linear transformations, which is the same across different positions but vary from layer to layer. Each layer of the Transformer Encoder module consists of two primary components, multi-head self-attention mechanisms and position-wise feed-forward networks. All the parts work together to interpret the input sequence into a high-level abstraction, prospering the transformer model a powerful architecture for dealing with sequential data.

Modeling of the human heart and arrhythmia electrocardiosignal acquisition

Finite element model establishment

The task of modeling the three-dimensional human heart is intricate and requires the integration of data from multiple imaging modalities, anatomical knowledge, and computational techniques. In specific, firstly, acquire high-resolution imaging from Magnetic Resonance Imaging (MRI) data, then improve image quality by applying cleaning and preprocessing algorithms to prepare the data for segmentation, and extract the heart structures from the imaging data. Finally, a 3D model is converted by the segmented 2D slices by stacking and interpolating between the slices, which can be rendered by visualization tools. In this paper, a full-scale, three-dimensional model of the human heart is established by MRI data, as shown in Fig 1(A).

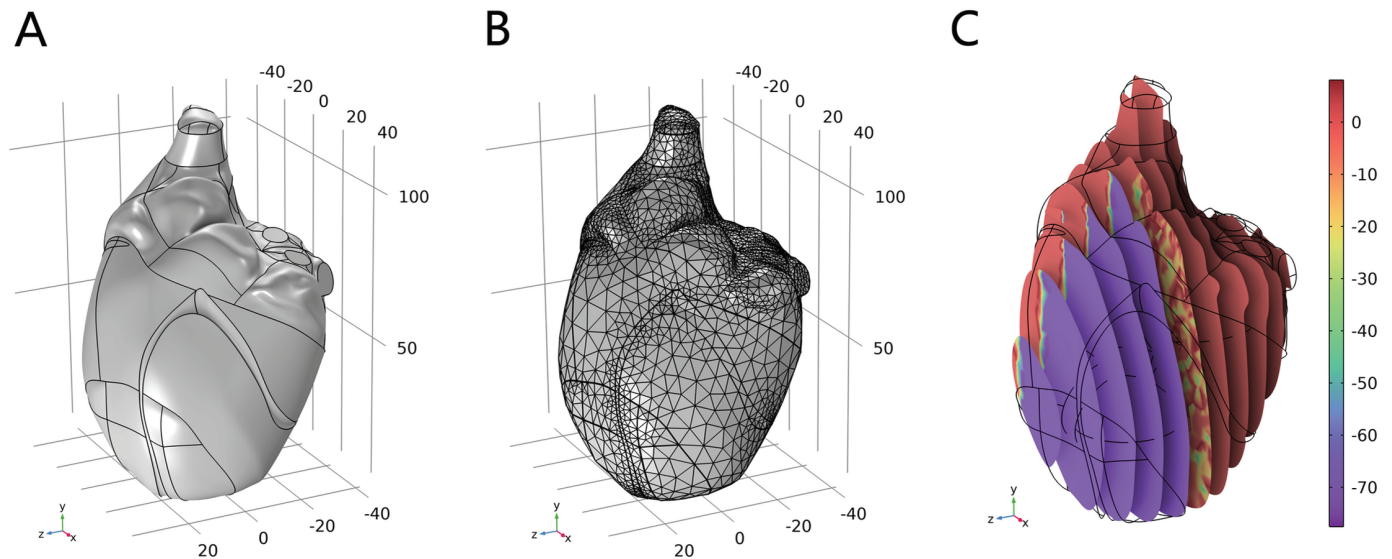


Fig 1. Model establishment of the human heart. (A) 3D geometry. (B) Finite element model. (C) Initial condition.

<https://doi.org/10.1371/journal.pone.0321484.g001>

The common values of the parameters in HH model are examined and determined by the following. $C_m \approx 1 \mu\text{F}/\text{cm}^2$, $\bar{g}_{\text{Na}} \approx 120 \text{ mS}/\text{cm}^2$, $\bar{g}_K \approx 36 \text{ mS}/\text{cm}^2$, $\bar{g}_L \approx 0.3 \text{ mS}/\text{cm}^2$, $E_{\text{Na}} \approx 50 \text{ mV}$, $E_K \approx -77 \text{ mV}$, $E_L \approx -54.4 \text{ mV}$.

For simulations, the typical initial condition of the membrane potential V_0 is set to -65 mV , and the initial gating variables are usually set to their steady-state values at the resting potential V_0 calculated by

$$\begin{aligned} m_0 &= \frac{\alpha_m(V_0)}{\alpha_m(V_0) + \beta_m(V_0)} \\ h_0 &= \frac{\alpha_h(V_0)}{\alpha_h(V_0) + \beta_h(V_0)} \\ n_0 &= \frac{\alpha_n(V_0)}{\alpha_n(V_0) + \beta_n(V_0)} \end{aligned} \quad (19)$$

The potential distribution V is initialized by setting one quadrant of the heart at a constant, elevated potential, V_0 , while the rest remains at zero. The logical expression of the boundary condition can be written as

$$V(0, x, y, z) = V_0((x + d) > 0) \cdot ((z + d) > 0) \quad (20)$$

Where TRUE evaluates to 1 and FALSE to 0. d is set to 10^{-5} , which is included in the expressions to shift the elevated potential slightly off the main axes.

The simulation platform *COMSOL Multiphysics* is adopted for partial differential equation (PDE) solving and result visualization, which allows for coupling various physical fields, including cardiac electrophysiology and mechanics. Due to a plethora of irregular and complex surfaces in the 3D model, the free tetrahedron mesh is employed as the partitioning strategy. The finite element mesh model that consists of 30245 elements across all domains is shown in Fig 1(B), and the initial condition of Eq 19 is visualized in Fig 1(C).

Through FEM simulation results, the temporal variation pattern of the action potential throughout the heart model region is revealed. Five representative points within the heart region are extracted for further analysis. As shown in Fig 2, the action potential signals at these representative points exhibit typical and homogeneous electrocardiographic characteristics. The different phases demonstrate the propagation of electrical signals through the heart tissues. For convenience, the original model is named model 0, and the electrocardiographic signals are identified as signal 0.

Various arrhythmias modelings induced by PDE parameters

According to the Hodgkin-Huxley model, it can be deduced that the parameters in partial differential Eq 1 and gating variable Eq 2 play essential roles in the propagation behaviors of electrocardiographic signals. Therefore, the key anomalies characterized by I and h in the mathematical model of HH are examined in detail. An exponentially attenuated sinusoidal function is well-designed for simulating the HH model's abnormal dynamics caused by PDE variables, which is defined as

$$\Psi(t) = e^{-4t} * \sin(2\pi t) \quad (21)$$

The function consists of two main components: an exponentially decaying term and a sinusoidal term with a frequency of 2π . Fig 3 depicts the varying pattern of the function value over time, which exhibits a sinusoidal oscillation whose amplitude decays exponentially.

The variable I is a critical parameter representing the external current applied to the neuron. The underlying dynamics of action potential generation and propagation in cardiac cells can be understood by adjusting this parameter to simulate and study different excitability conditions. Here, consider multiplying I by the exponentially attenuated sinusoidal function $\Psi(t)$, then the HH model given by Eq 1 transforms to

$$C_m \frac{dV}{dt} = I * \Psi(t) - \bar{g}_{Na} m^3 h (V - E_{Na}) - \bar{g}_K n^4 (V - E_K) - \bar{g}_L (V - E_L) \quad (22)$$

The arrhythmia model described by Eq 22 is denoted as model 1. After mesh generation and FEM calculation, the results of five representative points within the heart region are

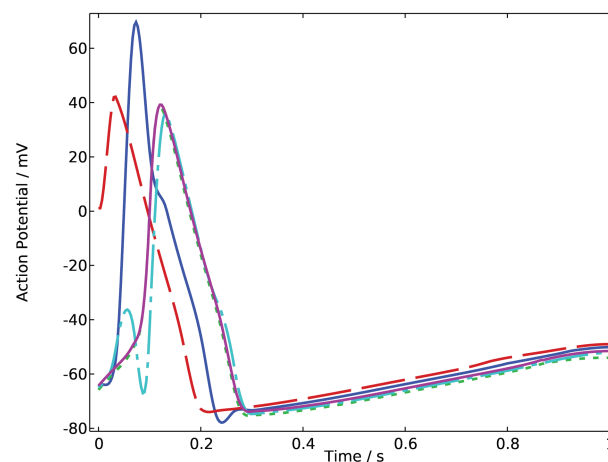


Fig 2. Action potential of sample points in model 0.

<https://doi.org/10.1371/journal.pone.0321484.g002>

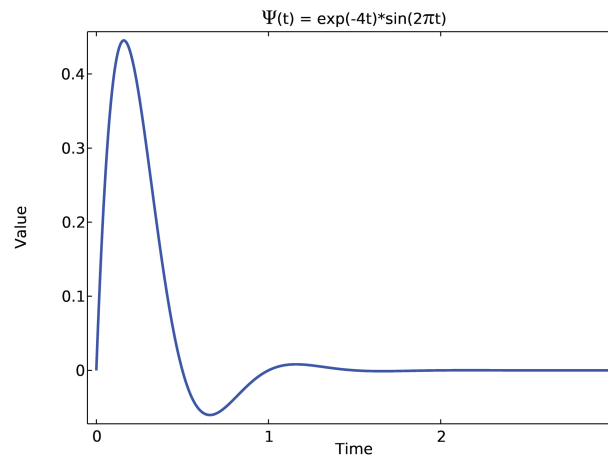


Fig 3. Exponentially attenuated sinusoidal function.

<https://doi.org/10.1371/journal.pone.0321484.g003>

displayed in Fig 4, which are referred to as signal 1. The curves of action potential changing with time show a homogeneous pattern with different amplitudes and phases, which exhibit more fluctuated characteristics compared with signal 0.

The sodium inactivation gate h plays a critical role in regulating the availability of sodium channels during and after an action potential. By inactivating the channels during depolarization and allowing them to recover during repolarization, it ensures that action potentials are well-spaced and that the neuron can return to its resting state properly. The exponentially attenuated sinusoidal function $\Psi(t)$ is encapsulated in Eq 2 of the sodium inactivation variable h , which is regarded as the arrhythmia model 2 and can be represented by

$$\frac{dh}{dt} = (\alpha_h(V)(1-h) - \beta_h(V)h) * \Psi(t) \quad (23)$$

Similarly, the arrhythmia model 2 represented by Eq 23 is modeled and calculated using the finite element method. The results for five representative points within the heart region are illustrated in Fig 5. The curves of action potential changing over time demonstrate a homogeneous pattern but show different characteristics compared with signal 0 and signal 1 in terms of amplitude.

In order to construct the final synthetic signals, the prototypical signals derived from the three models are collected and processed through periodization. Moreover, Gaussian white noise, with a range of signal-to-noise ratios (SNR), is introduced to the synthetic signals to enhance their realism and robustness in various analytical contexts. Fig 6 exhibits the comparison of the typical synthetic signals of the three models with a noise standard deviation of 5 and a mean of -65, which are prepared for the next signal decomposition and classification.

Variational mode decomposition based on multi-objective COA

Crayfish optimization algorithm and comparison with competitive meta-heuristic algorithms

The Crayfish Optimization Algorithm (COA) represents a novel meta-heuristic optimization method inspired by the foraging behaviors and navigational strategies observed in

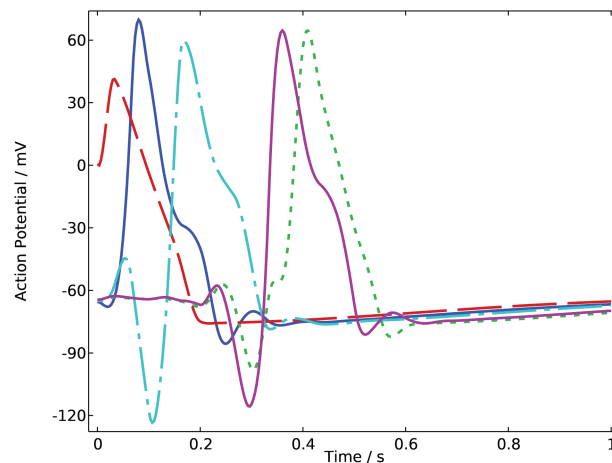


Fig 4. Action potential of sample points in model 1.

<https://doi.org/10.1371/journal.pone.0321484.g004>

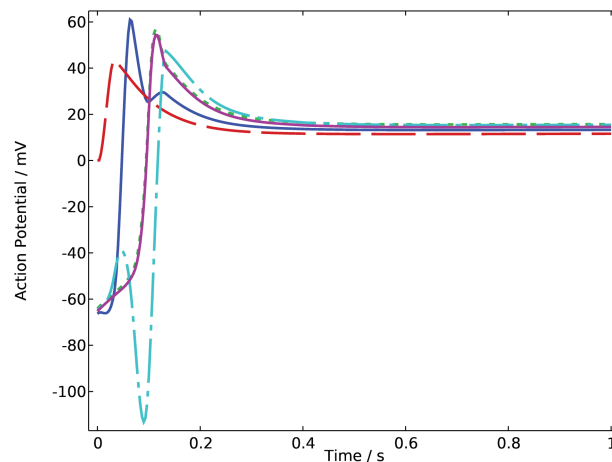


Fig 5. Action potential of sample points in model 2.

<https://doi.org/10.1371/journal.pone.0321484.g005>

crayfish [41]. Similar to other nature-inspired algorithms (such as Genetic Algorithms, Particle Swarm Optimization, and Ant Colony Optimization), COA aims to solve complex optimization problems by mimicking the natural processes observed in crayfish. The feeding amount of crayfish is influenced by temperature, with the ideal feeding range for crayfish being between 20°C and 30°C, and 25°C being the best temperature. Thus, COA defines temperature range from 20 to 35 °C

$$temp = rand \times 15 + 20 \quad (24)$$

where *temp* denotes the ambient temperature of the crayfish's location, and *rand* is the random scalar obtained from the uniform distribution of the interval (0,1). The mathematical

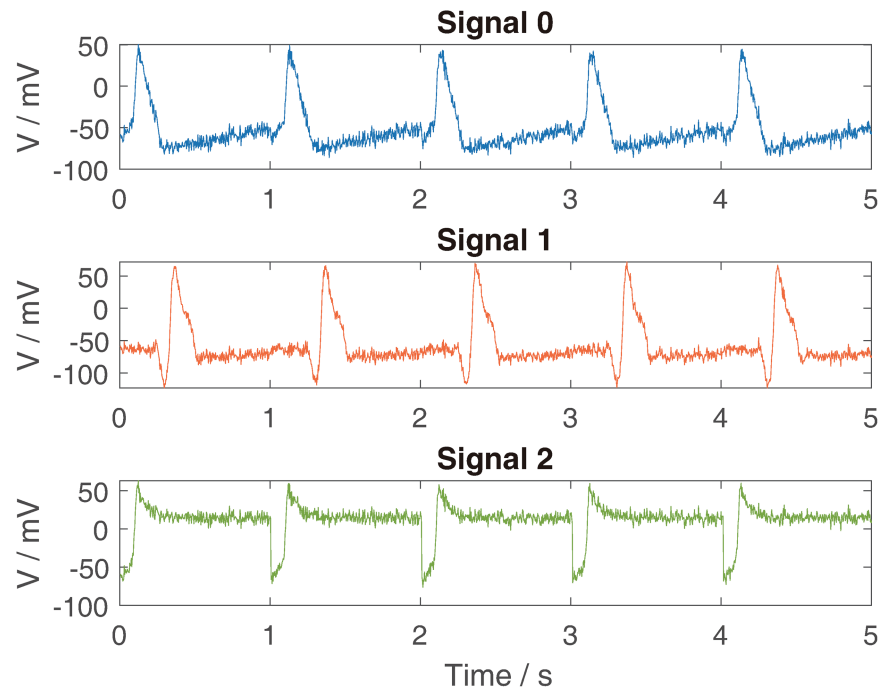


Fig 6. Different patterns of electrocardio signals with Gaussian noise. X-axis represents time, Y-axis represents the magnitude of the membrane potential.

<https://doi.org/10.1371/journal.pone.0321484.g006>

model for crayfish intake is characterized by a Gaussian-like distribution

$$p = C_1 \times \left(\frac{1}{\sqrt{2\pi} \times \sigma} \times \exp \left(-\frac{(\text{temp} - \mu)^2}{2\sigma^2} \right) \right) \quad (25)$$

where μ denotes the temperature most suitable for crayfish, σ and C_1 are parameters controlling the intake of crayfish at different temperatures.

The goal of the crayfish at the summer resort stage is to get to the cave, which stands for the ideal solution. By doing so, individuals get closer to the ideal solution and improves COA's exploitation potential. The crayfish will enter the cave for summer resort by

$$X_{i,j}^{t+1} = X_{i,j}^t + C_2 \times \text{rand} \times (X_{\text{shade}} - X_{i,j}^t) \quad (26)$$

where t , $t + 1$ represent the current generation number and the next generation iteration number, respectively. C_2 is a decreasing coefficient. The cave $X_{\text{shade}} := (X_G + X_L) / 2$, where X_G denotes the optimal position obtained so far by the number of iterations, and X_L refers to the optimal position of the current population.

In the Competition stage, crayfish X_i engage in competition with each other and modify their positions in response to the position X_z of another crayfish. Adjusting the position expands the search range of COA. The crayfish compete for the cave through

$$X_{i,j}^{t+1} = X_{i,j}^t - X_{z,j}^t + X_{\text{shade}} \quad (27)$$

where z represents the random individual of crayfish calculated by $z = \text{round}(\text{rand} \times (N - 1)) + 1$, N denoting the population size.

In the Foraging stage, crayfish use different feeding methods based on the size of their food Q

$$Q = C_3 \times \text{rand} \times \left(\text{fitness}_i / \text{fitness}_{\text{food}} \right) \quad (28)$$

where C_3 is the food factor, representing the largest food. fitness_i represents the fitness value of the i -th crayfish, and $\text{fitness}_{\text{food}}$ represents the fitness value of the food location. The crayfish will approach the food Q when it is the right size for eating. When Q is too large ($Q > (C_3 + 1)/2$), it indicates that there is a significant difference between the crayfish and the optimal solution. At this time, the crayfish will tear the food with the first claw and the mathematical equation is

$$X_{\text{food}} = \exp\left(-\frac{1}{Q}\right) \times X_{\text{food}} \quad (29)$$

The food obtained by crayfish is also related to the food intake, so the equation for foraging is as follows

$$X_{ij}^{t+1} = X_{ij}^t + X_{\text{food}} \times p \times (\cos(2 \times \pi \times \text{rand}) - \sin(2 \times \pi \times \text{rand})) \quad (30)$$

where food X_{food} represents the optimal solution, which is supposed to be reduced and brought closer to the food. When the size of the food is not too large ($Q \leq (C_3 + 1)/2$), the crayfish can simply move towards the food and eat directly

$$X_{ij}^{t+1} = (X_{ij}^t - X_{\text{food}}) \times p + p \times \text{rand} \times X_{ij}^t \quad (31)$$

Through the foraging stage, COA will approach the optimal solution, improving the algorithm's exploitation ability and allowing it to have excellent convergence capabilities.

A well-designed test function, named Ackley's Function, is built and used to verify the performance of COA. Transformations such as shifts and rotations are introduced to increase the complexity of the function landscape. As a result, this function poses significant challenges for global optimization algorithms due to its numerous local minima and a single global minimum. The expression of the test function is

$$f_{\text{Ackley}}(\mathbf{x}) = -a \exp\left(-b \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2}\right) - \exp\left(\frac{1}{d} \sum_{i=1}^d \cos(cx_i)\right) + a + e \quad (32)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_d)$ is a vector of variables, d is the dimension of the search space. a , b , and c are constants, here $a = 20$, $b = 0.2$, and $c = 2\pi$. e is the base of the natural logarithm.

The function landscape is displayed in Fig 7(A), where the range of \mathbf{x} is $[-32, 32]$ in each dimension, and the colormap corresponds to the function value. Several state-of-the-art algorithms mentioned before, i.e., GWO, GRO, RIME, PSO, GA, and BWO were chosen as competitors, the convergence comparison of which is shown in Fig 7(B). After 1000 iterations, GA has the best precision followed by COA, whereas the other algorithms exhibit premature convergence and remain steady. However, it is evident that COA converges faster than GA, which falls into the local optimum in hundreds of iterations during optimization. In summary, COA shows the greatest potential among other algorithms and is worth further development for multi-objective optimization.

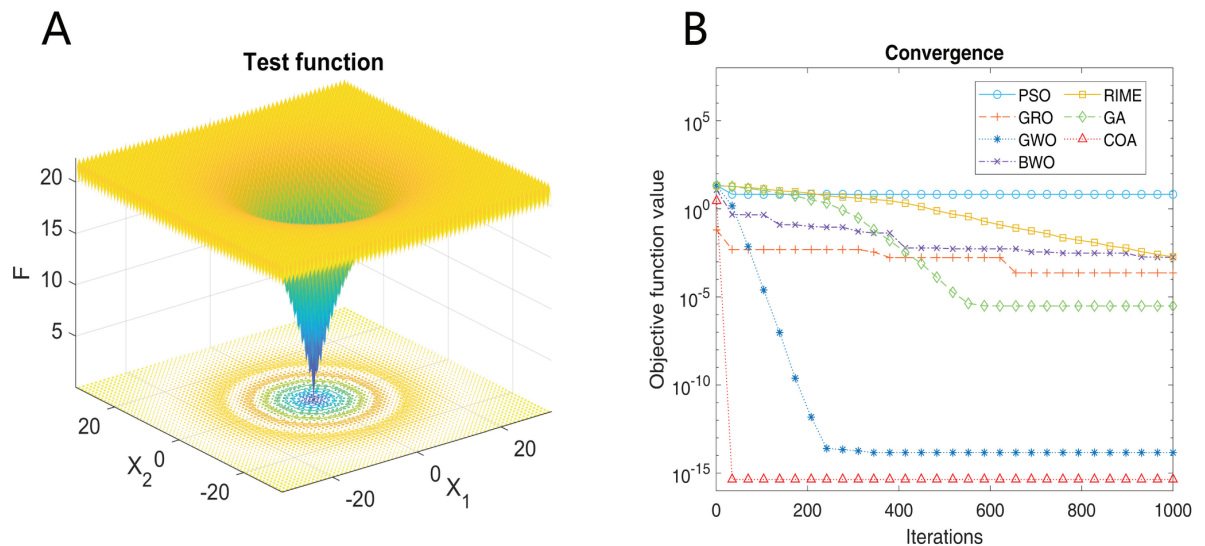


Fig 7. Comparison of different optimization algorithms. (A) Schematic diagram of test function. (B) Convergence comparison.

<https://doi.org/10.1371/journal.pone.0321484.g007>

Multi-objective COA based on non-dominated sorting

The original Crayfish Optimization Algorithm (COA) is formulated for scalar objective optimization, where a single objective function is optimized. However, in many practical engineering applications, multi-objective optimization problems are prevalent. These problems involve the simultaneous optimization of multiple, often conflicting, objectives, which adds a layer of complexity not present in single-objective optimization. For example, solutions to a problem with two objectives are not straightforward, as improving one objective may lead to the deterioration of another. Therefore, it is imperative to enhance the COA to broaden its applicability for multi-objective optimization problems. Regarding this point, a Multi-objective COA (MOCO) based on non-dominated sorting is proposed in this paper.

Non-dominated sorting is an essential concept in multi-objective optimization, particularly in evolutionary algorithms. It is used to classify solutions based on Pareto dominance, helping to identify the set of optimal solutions known as the Pareto front [47].

For the minimization optimization problem with a feasible region Ω , a vector $U = [u_1, u_2, \dots, u_k] \in \mathbb{R}^k$ is dominant to vector $V = [v_1, v_2, \dots, v_k] \in \mathbb{R}^k$ (denoted by $U < V$) if and only if $\forall i \in \{1, 2, \dots, k\}, u_i \leq v_i \wedge \exists j \in \{1, 2, \dots, k\} : u_j < v_j$. That is to say, there is at least one u_j which is smaller than v_j whilst the remaining u 's are either smaller or equal to corresponding v 's. The solution vector X^* is considered Pareto optimal (minimal) if no other solution can be found to dominate X^* according to the definition of Pareto dominance, i.e., $\forall i \in \{1, 2, \dots, k\}, \forall X \in \Omega - \{X^*\}, f_i(X^*) \leq f_i(X) \wedge \exists j \in \{1, 2, \dots, k\} : f_j(X^*) < f_j(X)$. A Pareto set is a set in the decision variable space consisting of all the Pareto optimal vectors $P^* = \{X \in \Omega | \nexists X' \in \Omega : F(X') < F(X)\}$. The Pareto front PF^* is a set of vectors of objective functions generated from the vectors of decision variables in the Pareto set $PF^* = \{F(X) = (f_1(X), f_2(X), \dots, f_k(X)) : X \in P^*\}$.

Non-dominated sorting is a fundamental technique in multi-objective optimization to classify solutions based on Pareto dominance. Solutions are grouped into different fronts based on their dominance relationships, and each front is assigned a rank, which denotes the level of Pareto dominance it belongs to. For example, solutions in the same rank are equally

non-dominated with respect to each other. Lower ranks (e.g., Rank 1) indicate solutions that are closer to the Pareto front, representing more optimal trade-offs among the objectives.

The concept of crowding degree is further introduced, which represents the density of the surrounding individuals. As the crowding degree decreases, the density of surrounding individuals increases. Consequently, individuals with greater crowding distance at the same rank are selected. The crowding degree is computed by

$$\text{distance}[i] = \frac{f(x_{j-i+1}) - f(x_{j-i-1})}{f_{j_max} - f_{j_min}} \quad (33)$$

where $f(x_{j-i+1})$ and $f(x_{j-i-1})$ refer to the two adjacent fitness values of the i -th individual in the j -th rank, respectively. f_{j_max} is the maximum fitness value, and f_{j_min} is the minimum fitness value in the j -th rank.

The non-dominated sorting method in this paper can be mainly divided into two steps. The first step is to sort all individuals by using a fast non-dominated sorting algorithm. Then the crowding of the decision space ($CD_{i,x}$) and the object space ($CD_{i,f}$) of the i -th individual are calculated, respectively. The i -th individuals in the same rank are rated based on the special crowding distance

$$SCD_i = \begin{cases} \max(D_{i,x}, CD_{i,f}) & CD_{i,x} > CD_{avg,x} \text{ or } CD_{i,f} > CD_{avg,f} \\ \min(D_{i,x}, CD_{i,f}) & \text{otherwise} \end{cases} \quad (34)$$

where $CD_{avg,f}$ and $CD_{avg,x}$ denote the average crowding degree of the objective space and the decision space, respectively.

Step 1. Define the number of iterations T , population size N , dimension of design space dim , upper bound ub , and lower bound lb . For optimization target F_k with K different objective functions, initialize population X_{ij} derived from the upper and lower bounds, respectively

$$P_F(X) = \begin{bmatrix} X_{1,1} & \cdots & X_{1,j} & \cdots & X_{1,dim} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{i,1} & \cdots & X_{i,j} & \cdots & X_{i,dim} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{N,1} & \cdots & X_{N,j} & \cdots & X_{N,dim} \end{bmatrix} \quad (35)$$

X_{ij} is calculated by $X_{ij} = lb_j + (ub_j - lb_j) \times rand$, where lb_j , ub_j represent the lower bound and the upper bound of the j -th dimension, respectively. Every crayfish is a matrix with dimensions of $1 \times dim$, where each column matrix represents a solution to a problem. Moreover, the ambient temperature of crayfish is determined by Eq 24 to induce different stages of COA.

Step 2. When $temp > 30$ and $rand < 0.5$, COA activates the summer resort stage, a new position X_{ij}^{t+1} is generated according to the cave position X_{shade} and crayfish position X_{ij}^t by Eq 26. When $temp > 30$ and $rand \geq 0.5$, COA enters the competitive stage, in which two crayfish will compete for the cave, a new position X_{ij}^{t+1} is obtained based on the cave position (X_{shade}) and the positions (X_{ij}^t , $X_{z_j}^t$) of the two crayfish by Eq 27.

Step 3. When $temp \leq 30$, COA proceeds to the foraging stage, the food intake p and the food size Q are specified by Eq 25 and 28, respectively. If $Q > (C_3 + 1)/2$, shred the food according to Eq 29, then update the position through Eq 30. Otherwise, the position of crayfish is updated by Eq 31.

Step 4: Evaluate the fitness of X_1, \dots, X_N at step $t + 1$ using the different objective functions, respectively, and merge the populations by non-dominated sorting. Specifically, the population at step $t + 1$ is expanded as $X = \{X_1^{F_1}, \dots, X_N^{F_1}, \dots, X_1^{F_k}, \dots, X_N^{F_k}\}$, and then non-dominated sorting is conducted based on optimization fitness values $F(X)$. The special crowding distance of each individual is further computed as the sort criterion for individuals at the same rank. Eventually, the population consisting of kN individuals is fully sorted, and only the top N individuals are retained as the updated population for the subsequent step.

Step 5: Determine whether to end the cycle based on whether the maximum number of iterations T has been reached. If the maximum number of iterations has not been reached, continue with Step 2. Otherwise, output the Pareto optimal front PF^* .

COA is remarkable in its simulations of the summer resort mechanism, competition mechanism, and foraging mechanism of crayfish, showing significant advantages over other meta-heuristic algorithms. The non-dominated sorting and crowding degree sorting algorithm is innovatively introduced in the MOCOA proposed in this paper, thereby enhancing its applicability and efficacy in resolving multi-objective optimization problems.

Optimal VMD based on MOCOA

MOCOA-VMD (optimized VMD based on MOCOA) is proposed to search for the best number of intrinsic mode functions K and the penalty term α in Variational Mode Decomposition. In order to evaluate the effectiveness of VMD, the kurtosis and KL divergence are introduced as indicators of performance. The spectral kurtosis (SK) is a dimensionless time series statistic used to identify and quantify non-stationary changes in a signal that may reflect the random distribution of time series data. For a given frequency bin, SK measures the deviation of the power spectral density (PSD) from that expected for a Gaussian random process. A high SK level corresponds to a high level of nonstationary or non-Gaussian behavior [48,49]. The Kurtogram is a graphical representation that shows the spectral kurtosis values of various frequency bands in a signal, and it is used to identify the frequency range where the signal has the highest kurtosis, which can indicate the presence of faults or abnormalities in the signal [50].

The spectral kurtosis is calculated by

$$SK(f) = \frac{\langle |X(f)|^4 \rangle}{\langle |X(f)|^2 \rangle^2} - 2 \quad (36)$$

where X_f is the Fourier transform of a signal at frequency f , and $\langle \cdot \rangle$ is the operator that computes the expectation of a series.

Fig 8 shows the Kurtograms of different signals. Among them, the optimal window length of signals are 32, 4, 12, respectively, and their kurtograms exhibit similar but differentiated characteristics. The frequency bands with the highest kurtosis values occur at a frequency of around 20 Hz under the sampling frequency of 300 Hz, which indicates the presence of impulses or transients. Kurtosis effectively extracts transient signal changes, making it a suitable performance index.

Kullback–Leibler (KL) divergence quantifies how one probability distribution diverges from another and is a fundamental concept in the realms of information theory and statistics. The discrete form of KL divergence is determined as the index for measuring the similarity between different IMFs, which is defined as

$$D_{KL}(p||q) = \sum_{i=1}^N [p(x_i) \log p(x_i) - p(x_i) \log q(x_i)] \quad (37)$$

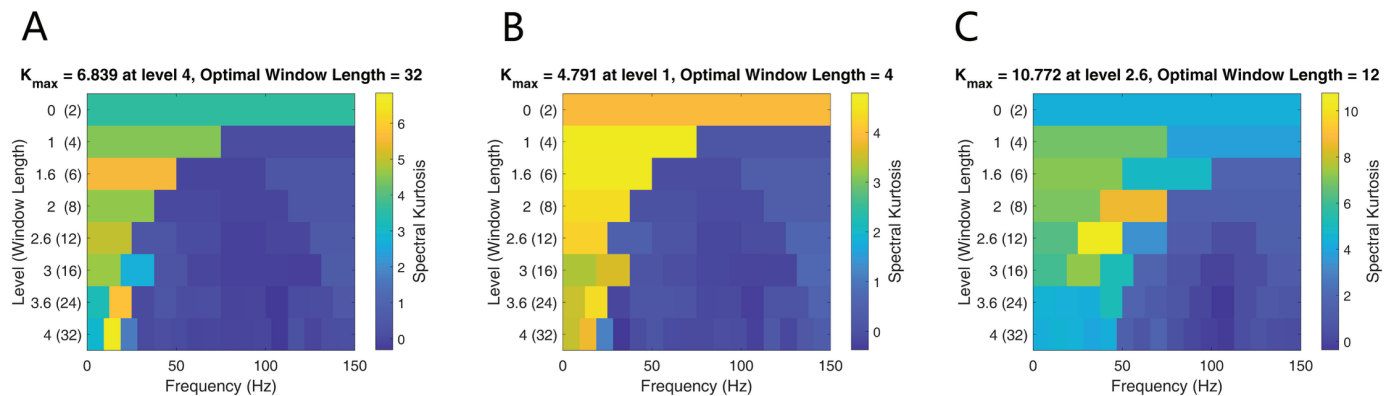


Fig 8. Kurtograms of different signals. (A) Kurtogram of signal 0. (B) Kurtogram of signal 1. (C) Kurtogram of signal 2. X-axis represents the frequency bands, Y-axis represents the resolution levels, color intensity represents the kurtosis value (higher intensity indicates higher kurtosis). Kurtograms exhibit similar impulses components but differentiated characteristics of the kurtosis values across three signals.

<https://doi.org/10.1371/journal.pone.0321484.g008>

where $P(x)$ and $Q(x)$ are the probabilities of x in the distribution of P (the true distribution) and Q (the approximate distribution), respectively.

A sensitivity analysis of K and α in VMD was conducted to illustrate how parameter variations affect signal decomposition performance. As shown in Fig 9, the influence of K on the spectral kurtosis and KL divergence is detailed under a fixed value of parameter $\alpha = 2500$; the same applies to α under a fixed $K = 6$. The KL divergence presents a contrary trend as K and α rise, and the values of spectral kurtosis fluctuate as K and α increase, which reveals the coupling effect of the VMD parameters on the performance indexes.

Sensitivity analysis indicates that K and α have varying and coupled impacts on decomposition performance and may interact with each other. Therefore, further optimization is required. The optimization goal for MOCOIA is to minimize the spectral kurtosis and KL divergence of the decomposed IMFs by VMD, with 2 design variables K and α , to be optimized. Their ranges of design space are integer $K \in [3, 8]$ and real number $\alpha \in [500, 5000]$, according to previous knowledge. The optimization problem can be written as

$$\begin{aligned} \arg \min_x \quad & O(x) = [SK_f(x), D_{KL}(x)]^T \\ \text{s.t.} \quad & x_1 = K \in [3, 8] \cap \mathbb{Z}, \\ & x_2 = \alpha \in [500, 5000] \cap \mathbb{R} \end{aligned} \quad (38)$$

Specifically, the MOCOIA configuration starts with an initial population of 300 based on the design space, and the maximum number of generations is set to 1000. The spectral kurtosis and KL divergence of each IMFs decomposed by VMD are calculated to evaluate the individual's performance at each generation continuously. The crowding degree sorting algorithm is employed to rank the individuals in each generation, and the best individuals are selected based on the Pareto dominance relationship. The optimal solution is determined by the Pareto optimal front, which contains the best trade-off solutions between the two objectives. Fig 10 displays the Pareto optimal front after 1000 generations of evolution in the case of signal 0.

Since all the optimal points in the Pareto optimal front's solution set are mutually non-dominant, any point in the front set could be the optimal candidate. The summary table of the key parameters and performance indices of the VMD is provided in Table 1. The optimal parameters $K = 5$ and $\alpha = 1928$ are determined after subtly selecting, which minimize the

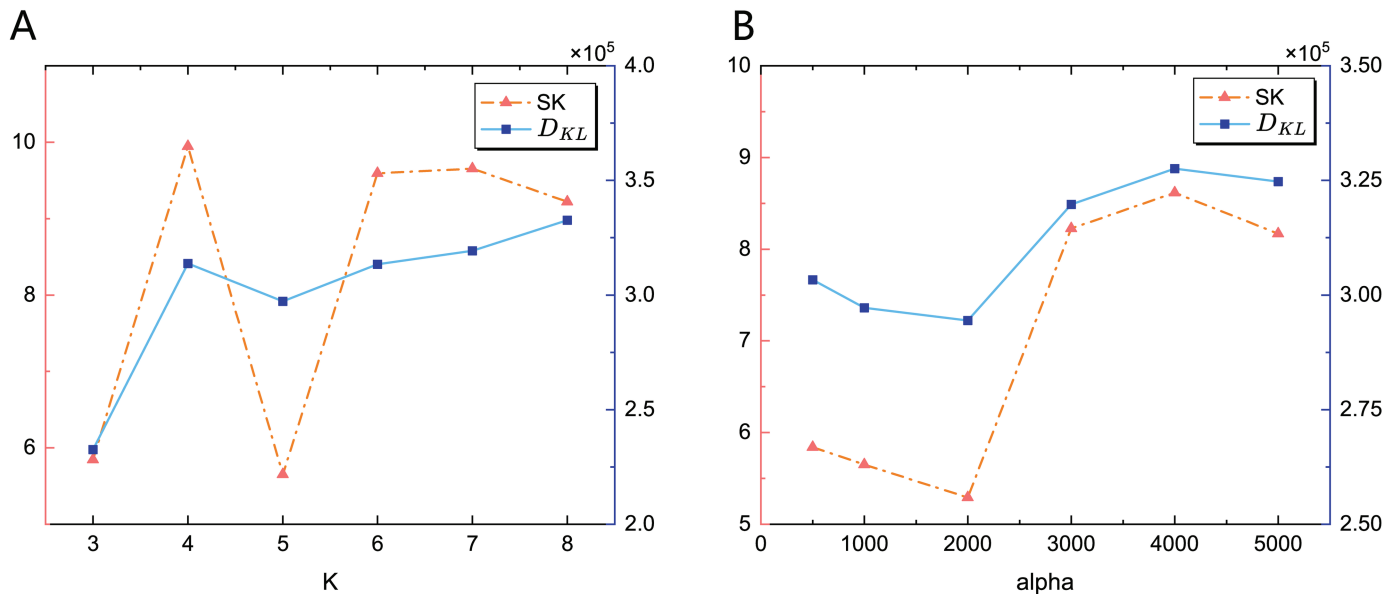


Fig 9. Sensitivity analysis of parameter K and α in VMD. (A) Influence of K on the VMD performance. (B) Influence of α on the VMD performance. X-axis represents the value of K or α , left Y-axis represents spectral kurtosis, right Y-axis represents KL divergence. The parameters K and α have significant and coupled impacts on decomposition performance.

<https://doi.org/10.1371/journal.pone.0321484.g009>

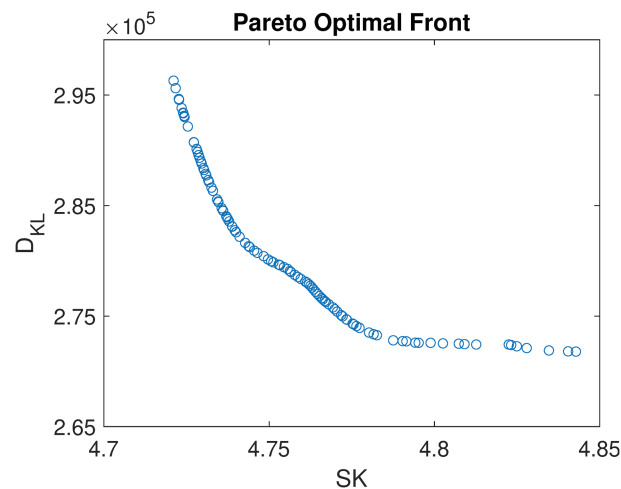


Fig 10. Pareto optimal front result of MOCOA.

<https://doi.org/10.1371/journal.pone.0321484.g010>

spectral kurtosis and KL divergence of the decomposed IMFs. The optimal VMD results are then fed into the deep attention model for arrhythmia classification.

Fig 11 shows the decomposition results of MOCOA-VMD compared with the original VMD with parameters $K = 6$, $\alpha = 2500$. It can be observed that the IMFs decomposed by MOCOA-VMD offers a balance between computational efficiency, decomposition stability, and interpretability, especially in IMF1, where the main component or trend of the signal is detected. The outcome validates the effectiveness of the proposed MOCOA-VMD method,

Table 1. Summary of key parameters and performance indices of VMD.

Design Params/Metrics	Name	Meaning	Optimal value
Design variable	K	number of modes	5
	α	bandwidth control parameter	1928
Performance index	SK	spectral kurtosis	4.74
	D_{KL}	KL divergence	2.82×10^5

<https://doi.org/10.1371/journal.pone.0321484.t001>

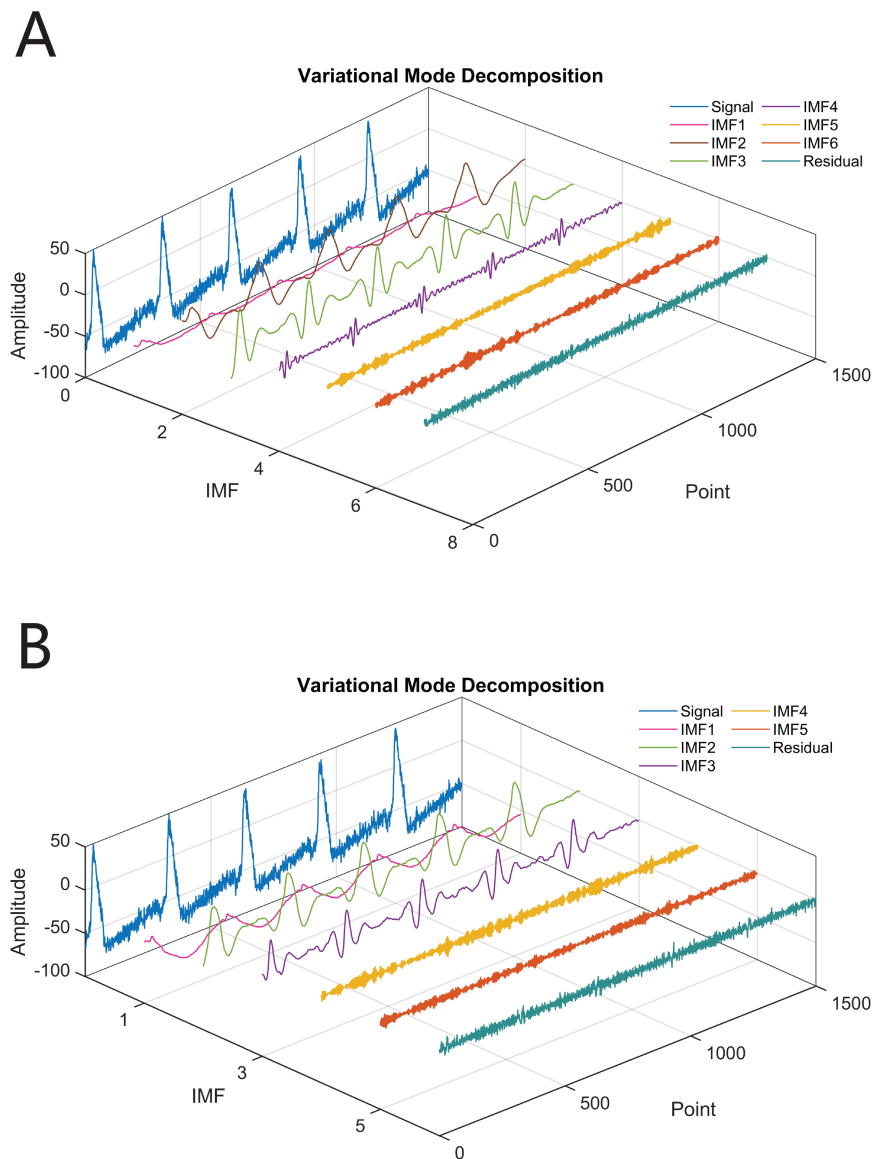


Fig 11. Comparison of original VMD and MOCO-VMD results. (A) Original VMD result. (B) Optimal VMD result based on MOCO-VMD. IMF1's MOCO-VMD result demonstrates superior performance in capturing the main component of the signal compared to the original VMD.

<https://doi.org/10.1371/journal.pone.0321484.g011>

which strikes a compromise between the necessity to capture all pertinent signal components and the potential for the introduction of noise or artifacts.

Learning model establishment and optimization for classification

Model establishment based on MOCOA-VMD attention scheme and ablation study

The classification of arrhythmias from electrocardiogram signals presents several significant challenges. These challenges often arise from the complex and often non-stationary nature of ECG signals, the substantial variability between patients, and the presence of noise and artifacts that can obscure relevant features. Facing these difficulties, a deep attention model based on MOCOA-VMD is established for arrhythmia signal classification, whose architecture is shown in Fig 12.

During the data acquisition stage, the Hodgkin-Huxley model is simulated and solved using FEM to obtain the ECG signals based on the simulation results. Two different types of arrhythmia signals are artificially generated by introducing the abnormalities in the PDEs. In each case, 200 representative points are uniformly and randomly selected throughout the heart region. The signals from these points are coupled with Gaussian noise at different SNRs. As a result, the total data consists of 3000 signals with 3 different ECG patterns, which are divided into 90% for training and 10% for testing. In the process of signal decomposition, the signals are processed using the proposed MOCOA-VMD method, which employs non-dominated sorting by calculating K_f and D_{KL} of IMFs and crowding degree sorting. The best decomposition results with the optimal parameters K and α in VMD are achieved and the IMF signals are fed to the deep attention model. In deep attention construction, the multi-head encoder consists of matrices Q , K , and V computed by decomposed signals, and the matrices are handled by linear transformation, scaled dot-product attention, and concatenation operation. Then the fully connected layer is employed to fit the features and corresponding categories, and ultimately the softmax layer outputs classification results in the form of a probability distribution.

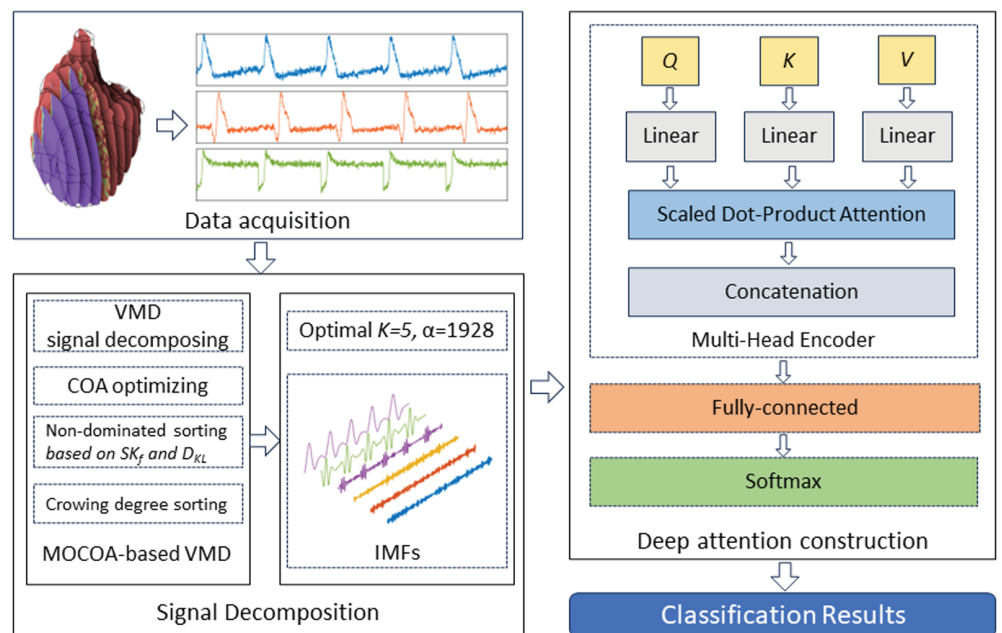


Fig 12. Deep attention network architecture for arrhythmia classification.

<https://doi.org/10.1371/journal.pone.0321484.g012>

In order to verify the effectiveness of the proposed signal decomposition method and deep attention modules, an ablation study is implemented, which is a critical part of evaluating the performance and importance of different components in a deep learning model or system. Specifically, the various components in the model are systematically removed or altered to understand their contribution to the overall performance and which parts might be redundant or less important.

The ablation study is carried out to help optimize the modules, and the corresponding results are shown in Table 2, where the ensemble empirical mode decomposition (EEMD), original variational mode decomposition (VMD), and the convolutional neural network (CNN), long short-term memory (LSTM) are chosen as competitors for the decomposition method and feature extraction, respectively. Moreover, paired t-tests with 10 samples of each model are implemented, and p-values are calculated to validate whether the performance differences between model variants are statistically significant. The p-values reflect that the performance improvements observed in ablation studies are significant.

It can be concluded that the attention scheme shows the best performance, followed by LSTM, which performs slightly worse, while CNN ranks last in terms of feature extraction. The attention scheme and LSTM generally exhibit superior performance in signal processing tasks due to their ability to handle temporal dependencies, sequence length flexibility, memory, and contextual information, adaptability to non-stationary data. Besides, the mode decomposition operation significantly improves the model performance. The performance of the model based on MOCO-VMD reaches the best accuracy of 94.46%, followed by VMD, and EEMD last, which verifies the superiority of the deep attention model based on MOCO-VMD proposed in this paper. Since the experimental values are adopted for setting the hyperparameters in the established deep attention model, to further improve the performance for classification, Bayesian optimization for hyperparameters will be studied and implemented in the next step.

Bayesian optimization for hyperparameters

Bayesian optimization is a powerful technique that efficiently seeks an approximate optimal solution while minimizing the evaluation cost, which is especially advantageous for scenarios where the computation of fitness functions is resource-intensive or time-consuming [51]. Gaussian Process [52] (GP) is a commonly used non-parametric probabilistic surrogate

Table 2. Comparison of ablation study. The deep model which consists of MOCO-VMD and attention scheme achieves the best accuracy.

Number	Signal decomposition	Feature extraction	Best Accuracy(%)	P-value
1	None	CNN	79.90	1.20×10^{-19}
2	None	LSTM	81.95	1.40×10^{-18}
3	None	Attention scheme	83.63	1.00×10^{-17}
4	EEMD	CNN	88.20	2.82×10^{-14}
5	EEMD	LSTM	89.09	4.19×10^{-13}
6	EEMD	Attention scheme	88.60	3.51×10^{-12}
7	VMD	CNN	91.14	1.35×10^{-09}
8	VMD	LSTM	91.95	1.89×10^{-08}
9	VMD	Attention scheme	92.99	5.49×10^{-05}
10	MOCO-VMD	CNN	93.56	9.76×10^{-03}
11	MOCO-VMD	LSTM	93.98	3.28×10^{-02}
12	MOCO-VMD	Attention scheme	94.46	1.00

<https://doi.org/10.1371/journal.pone.0321484.t002>

model, widely applied in regression, classification, and many other domains that require inference of black-box function. Formally, a GP is specified by a mean function $m: \mathcal{X} \rightarrow \mathbb{R}$ and a semi-positive definite covariance function (or kernel) $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \quad (39)$$

where $m(x) = \mathbb{E}[f(x)]$ is the mean function, $k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$ is the covariance function, which defines the covariance between the values of $f(x)$ and $f(x')$ at two points x and x' .

Given training data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$, where $y_i = f(x_i) + \epsilon$ and $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ is Gaussian noise, GP regression is trying to predict the function value at new points x_* . The joint distribution of the observed values y and the function values at the test points f_* is

$$\begin{pmatrix} y \\ f_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m(x) \\ m(x_*) \end{pmatrix}, \begin{pmatrix} K(x, x) + \sigma_n^2 I & K(x, x_*) \\ K(x_*, x) & K(x_*, x_*) \end{pmatrix}\right) \quad (40)$$

where $K(x, x)$ is the covariance matrix of the training points, $K(x_*, x)$ is the covariance matrix between test points and training points, $K(x_*, x_*)$ is the covariance matrix of the test points. The predictive distribution at the test points x_* is Gaussian with mean and covariance

$$\begin{aligned} \mu_* &= K(x_*, x) [K(x, x) + \sigma_n^2 I]^{-1} y \\ \Sigma_* &= K(x_*, x_*) - K(x_*, x) [K(x, x) + \sigma_n^2 I]^{-1} K(x, x_*) \end{aligned} \quad (41)$$

where the mean μ_* provides the predicted value, while Σ_* represents the uncertainty.

The acquisition function, often denoted as α , maps from the input space \mathcal{X} , the observation space \mathbb{R} , and the hyperparameter space Θ to the real number space, i.e., $\alpha: \mathcal{X} \times \mathbb{R} \times \Theta \rightarrow \mathbb{R}$. The function constructs the posterior distribution obtained from the observed data set $D_{1:t}$ and guides the selection of the next evaluation point x_{t+1} by maximizing α at time step t

$$x_{t+1} = \max_{x \in \mathcal{X}} \alpha_t(x; D_{1:t}) \quad (42)$$

One popular acquisition strategy is the Expected Improvement (EI), which is often used in conjunction with a Gaussian Process model. According to the Bayesian formula [53], EI at an acquisition point y^* can be written as

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} dy \quad (43)$$

Given the GP's predictive mean $\mu(x)$ and variance $\sigma^2(x)$ at a candidate point x and step t , the EI acquisition strategy based on the Gaussian process is specified as

$$\alpha_t(x; D_{1:t}) = \begin{cases} (v^* - \mu_t(x)) \phi\left(\frac{v^* - \mu_t(x)}{\sigma_t(x)}\right) + \sigma_t(x) \phi\left(\frac{v^* - \mu_t(x)}{\sigma_t(x)}\right), & \sigma_t(x) > 0 \\ 0, & \sigma_t(x) = 0 \end{cases} \quad (44)$$

where v^* is the best value observed so far, and $\phi(\cdot)$ is the normal distribution probability density function.

The Tree-structured Parzen estimator [54] (TPE) models the objective function using two probability density functions

$$p(x|y) = \begin{cases} \ell(x) & \text{if } y < y^* \\ g(x) & \text{if } y \geq y^* \end{cases} \quad (45)$$

where $\ell(x)$ represents the density formed by using the observations $\{x^{(i)}\}$ such that the corresponding loss $f(x^{(i)})$ was less than a threshold of y^* , $g(x)$ is the density generated by collecting the remaining observations.

By constructing $\gamma = p(y < y^*)$, then $p(x) = \int_{\mathbb{R}} p(x|y) p(y) dy = \gamma \ell(x) + (1 - \gamma) g(x)$, so we have

$$\int_{-\infty}^{y^*} (y^* - y) p(x|y) p(y) dy = \gamma y^* \ell(x) - \ell(x) \int_{-\infty}^{y^*} p(y) dy \quad (46)$$

finally

$$\text{EI}_{y^*}(x) = \frac{\gamma y^* \ell(x) - \ell(x) \int_{-\infty}^{y^*} p(y) dy}{\gamma \ell(x) + (1 - \gamma) g(x)} \propto \left(\gamma + \frac{g(x)}{\ell(x)} (1 - \gamma) \right)^{-1} \quad (47)$$

EI's goal is to balance exploration and exploitation by taking into account both the predicted mean and uncertainty of the model. Consequently, to fully optimize point X, it is essential to maximize the function $\ell(x)$ and minimize the function $g(x)$. During each iteration, the TPE algorithm returns the candidate x^* with the highest expected improvement.

In Bayesian optimization, batch size, learning rate, number of epochs, and momentum are determined as key parameters to optimize. Batch size defines the number of training samples used in one iteration of training a neural network, which calls for a compromise between computational efficiency and model convergence behavior. While trying to minimize the loss function, the learning rate controls the step size at each iteration. The number of epochs specifies how many times the learning algorithm will work through the entire training dataset, which directly impacts the model's performance and generalization ability. Momentum is a technique used in the optimization process of training neural networks to accelerate the gradient vectors in the right directions, thus leading to faster convergence and helping to smooth out the gradients and avoid getting stuck in local minima.

Referring to the general experience of fine-tuning a deep model, the ranges of values for the key design parameters are determined as batch size $\{4, 8, 16, 32, 64\}$, learning rate $[0.0001, 0.1] \cap \mathbb{R}$, epochs $[200, 2000] \cap \mathbb{Z}$, momentum $[0.8, 0.99] \cap \mathbb{R}$. The design space is composed of all the combinations of the above key parameters. The Gaussian Process surrogate model and TPE algorithm are put into operation for the deep attention models that demand high evaluation costs. The results of the top 10 performance combinations are shown in Fig 13. The best accuracy of 96.11% occurs with a combination of batch size 32, learning rate 0.00092, epochs 1070, and momentum 0.966, which is higher by 1.65% than the model before optimization.

The performance measurement tool confusion matrix is utilized to evaluate the accuracy of a classification model, as shown in Fig 14. It can be seen that the proposed model achieves excellent success in the classification of the different patterns of ECG signals, with accuracies up to 97.06% and 96.47% for signal 0 and signal 1, respectively, and slightly lower at 94.91% for signal 2. We speculate that it is possibly due to the varying characteristics modeled by PDEs for arrhythmia ECG patterns.

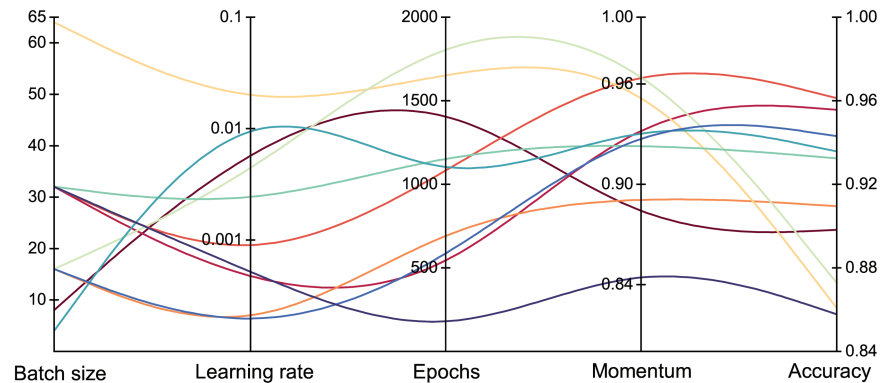


Fig 13. Bayesian optimization results of key hyperparameters.

<https://doi.org/10.1371/journal.pone.0321484.g013>

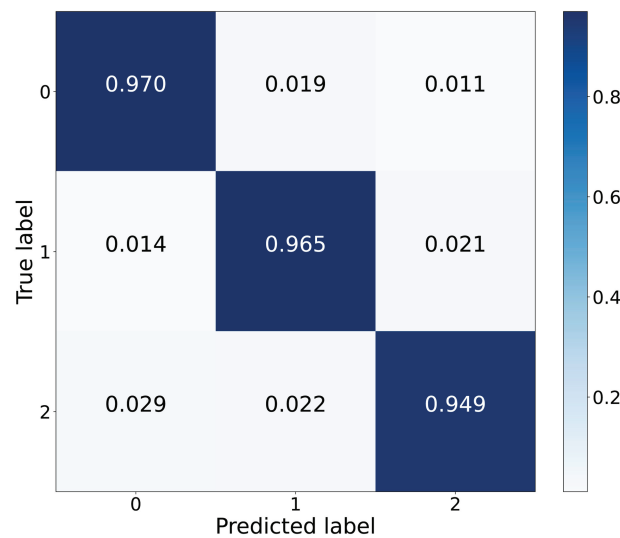


Fig 14. Confusion matrix.

<https://doi.org/10.1371/journal.pone.0321484.g014>

Model validation on MIT-BIH database

To assess the robustness and generalizability of the proposed model, we utilize the well-known MIT-BIH Arrhythmia Database for real-world ECG signal analysis. This database, provided by PhysioNet, includes annotated ECG recordings from 47 patients, totaling approximately 25 hours of recording time. Each record contains two lead ECG signals (MLII and V1) sampled at 360 Hz with 12-bit resolution. The comprehensive annotations facilitate the evaluation of our algorithm's performance in detecting various arrhythmias.

Three different type of ECGs are chosen for model validation, i.e., 75052 samples of normal sinus beat (N), 7130 samples of premature ventricular contraction (V) and 2546 samples of premature atrial contraction (A). The dataset was divided into 80% for training and 20% for testing. The results of the performance validation conducted on the MIT-BIH Arrhythmia Database after 1000 epochs are shown in Fig 15. The training state curve in Fig 15(A) shows that the accuracy reaches 95.46% on the training set and 92.87% on the test set. The confusion matrix in Fig 15(B) reveals that the model performs less accurately on normal sinus beats

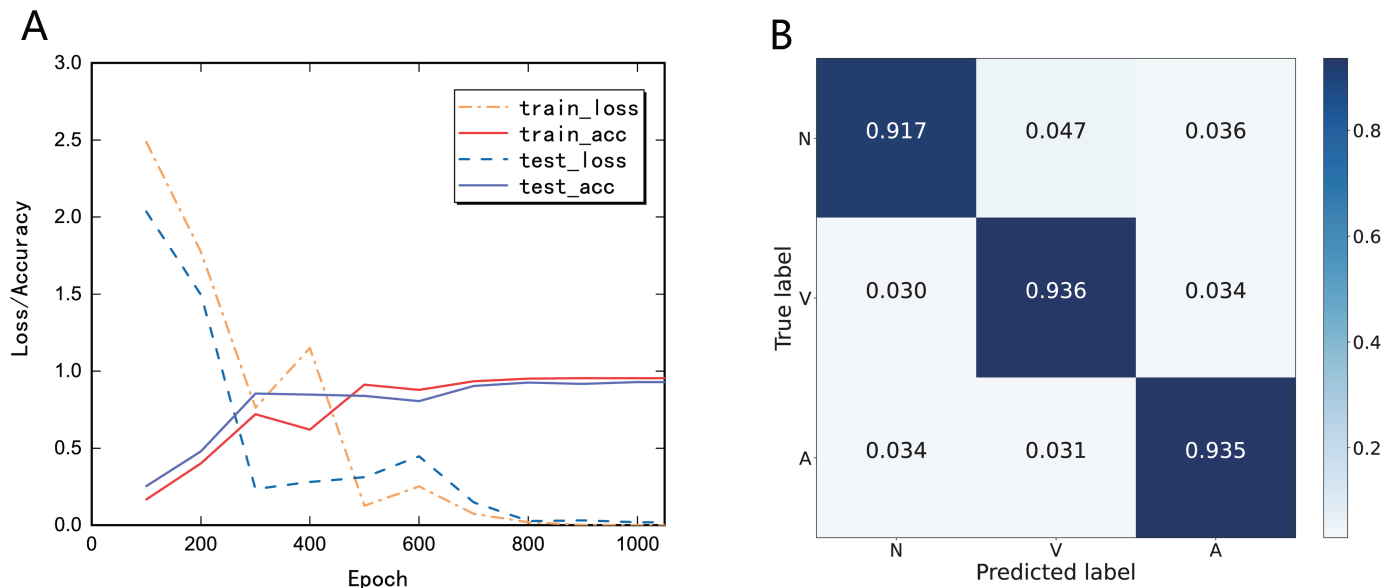


Fig 15. Performance validation on the MIT-BIH database. (A) Training state. (B) Confusion matrix.

<https://doi.org/10.1371/journal.pone.0321484.g015>

Table 3. Performance comparison with other papers.

Paper	Number of classes	Models	Accuracy(%)
[8]	5	CNN+LSTM	97.15
[13]	14	ELM-RNN	96.41
[14]	5	SMOTE+CNN	92.73
[15]	3	random forest	88.70
this paper	3	VMD+attention scheme	92.87

<https://doi.org/10.1371/journal.pone.0321484.t003>

compared to premature ventricular contractions and premature atrial contractions, which may necessitate further investigations in the future.

The experiment results are compared with other related studies based on the MIT-BIH database, as shown in Table 3. It can be seen that the deep model proposed in this paper performs well but needs to be further improved in future studies. Integrating different deep modules, such as CNN, RNN and LSTM for feature extraction may be a promising direction.

In summary, the deep VMD-attention network based on the multi-objective crayfish optimization algorithm in this paper is highly effective for classification tasks and is believed to have the potential to be utilized in other signal processing and classification issues.

Conclusion

This paper proposes a novel deep attention model for arrhythmia signal processing based on multi-objective crayfish optimization algorithmic variational mode decomposition. The main conclusions are as follows.

1. A finite element model of the human heart based on the Hodgkin-Huxley model was established for cardiac electrophysiology simulation, and the ECG signals were obtained from the FEM results of representative points. Two distinct types of arrhythmia, characterized by major anomalies of parameters I and h in the HH model, were examined

in detail, and the corresponding ECG signals were also obtained through simulation. The typical synthetic signals from the three models with built-in Gaussian noise, were prepared for signal decomposition and classification.

2. A variational mode decomposition technique for ECG signal processing based on a multi-objective crayfish optimization algorithm (MOCOA-VMD) was proposed. A multi-objective optimization method based on non-dominated sorting was incorporated into the crayfish optimization algorithm to search for the optimal K and α in VMD processing, and the spectral kurtosis and KL divergence were determined as the indicators for decomposition. The Pareto optimal front was generated by MOCOA and the intrinsic mode functions of VMD with the best combination of K and α were obtained.
3. A deep VMD-attention model based on MOCOA was constructed for ECG signal classification. To verify the effectiveness of the proposed signal decomposition method and deep attention modules, an ablation study was implemented. The performance of the model based on MOCOA-VMD achieves the best accuracy of 94.46%, much higher than the model constructed by modules of EEMD, VMD, CNN and LSTM. Moreover, Bayesian optimization was carried out to fine-tune the hyperparameters batch size, learning rate, epochs, and momentum. The best accuracy of the deep attention model after TPE optimization reaches 96.11%, which is 1.65% higher than the original model. The real-world ECG signals from the MIT-BIH arrhythmia database were utilized for further validation, and the accuracy reached 95.46% on the training set and 92.87% on the test set, proving the robustness and generalizability of the proposed model.

In conclusion, the deep VMD-attention network based on the HH model and MOCOA has demonstrated significant success in the classification of ECG signals. The method proposed in this paper shows substantial potential for both mathematical modeling and practical applications. Looking forward, on the one hand, the VMD-attention network needs to be further optimized to improve performance, and the mathematical models for different arrhythmias' electrophysiology are to be established. On the other hand, we anticipate expanding this ECG signal modeling and processing strategy to various other biological time-dependent and non-stationary signal processing domains, such as electroencephalogram (EEG) signals, respiratory signals, blood pressure signals, and electrooculography (EOG).

Acknowledgments

The authors would like to express gratitude to Shenghua Zhu for his professional suggestions about the 3D model establishment.

Author contributions

Conceptualization: Hang Zhao, Xiongfei Yin.

Data curation: Hang Zhao.

Formal analysis: Hang Zhao.

Funding acquisition: Xiongfei Yin.

Investigation: Hang Zhao.

Methodology: Hang Zhao.

Supervision: Xiongfei Yin.

Visualization: Hang Zhao, Xiongfei Yin.

Writing – original draft: Hang Zhao.

Writing – review & editing: Hang Zhao, Xiongfei Yin.

References

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. *Circulation*. 2020;141(9):e139–e596.
2. Antzelevitch C, Burashnikov A. Overview of basic mechanisms of cardiac arrhythmia. *Card Electrophysiol Clin*. 2011;3(1):23–45. <https://doi.org/10.1016/j.ccep.2010.10.012> PMID: 21892379
3. Ebrahimi Z, Loni M, Daneshdalan M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X*. 2020;7:100033. <https://doi.org/10.1016/j.eswx.2020.100033>
4. Feeny AK, Chung MK, Madabhushi A, Attia ZI, Cikes M, Firouznia M, et al. Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. *Circ Arrhythm Electrophysiol*. 2020;13(8):e007952. <https://doi.org/10.1161/CIRCEP.119.007952> PMID: 32628863
5. Liu J, Li Z, Jin Y, Liu Y, Liu C, Zhao L, et al. A review of arrhythmia detection based on electrocardiogram with artificial intelligence. *Expert Rev Med Devices*. 2022;19(7):549–60. <https://doi.org/10.1080/17434440.2022.2115887> PMID: 35993248
6. Singhal S, Kumar M. A systematic review on artificial intelligence-based techniques for diagnosis of cardiovascular arrhythmia diseases: challenges and opportunities. *Arch Computat Methods Eng*. 2022;30(2):865–88. <https://doi.org/10.1007/s11831-022-09823-7>
7. Oh SL, Ng EYK, Tan RS, Acharya UR. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput Biol Med*. 2018;102:278–87. <https://doi.org/10.1016/j.compbiomed.2018.06.002> PMID: 29903630
8. Chen C, Hua Z, Zhang R, Liu G, Wen W. Automated arrhythmia classification based on a combination network of CNN and LSTM. *Biomedical Signal Processing Control*. 2020;57:101819. <https://doi.org/10.1016/j.bspc.2019.101819>
9. Ullah A, Rehman SU, Tu S, Mehmood RM, Ehatisham-UI-Haq M. A hybrid deep CNN model for abnormal arrhythmia detection based on cardiac ECG signal. *Sensors (Basel)*. 2021;21(3):951. <https://doi.org/10.3390/s21030951> PMID: 33535397
10. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. 2021;18(7):465–78. <https://doi.org/10.1038/s41569-020-00503-2> PMID: 33526938
11. Pachori D, Tripathy RK, Jain TK. Detection of atrial fibrillation from ppg sensor data using variational mode decomposition. *IEEE Sens Lett*. 2024.
12. Kuila S, Dhanda N, Joardar S. Feature extraction of electrocardiogram signal using machine learning classification. *IJECE*. 2020;10(6):6598. <https://doi.org/10.11591/ijece.v10i6.pp6598-6605>
13. Kuila S, Dhanda N, Joardar S. ECG signal classification and arrhythmia detection using ELM-RNN. *Multimed Tools Appl*. 2022;81(18):25233–49. <https://doi.org/10.1007/s11042-022-11957-6>
14. Aphale SS, John E, Banerjee T. ArrhyNet: a high accuracy arrhythmia classification convolutional neural network. In: 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS). 2021. p. 453–7. <https://doi.org/10.1109/mwscas47672.2021.9531841>
15. Ilbeigipour S, Albadvi A, Akhondzadeh Noughabi E. Real-time heart arrhythmia detection using apache spark structured streaming. *J Healthc Eng*. 2021;2021:6624829. <https://doi.org/10.1155/2021/6624829> PMID: 33968352
16. Qi M, Shao H, Shi N, Wang G, Lv Y. Arrhythmia classification detection based on multiple electrocardiograms databases. *PLoS One*. 2023;18(9):e0290995. <https://doi.org/10.1371/journal.pone.0290995> PMID: 37756278
17. El-Ghaish H, Eldele E. ECGTransForm: Empowering adaptive ECG arrhythmia classification framework with bidirectional transformer. *Biomed Signal Process Control*. 2024;89:105714. <https://doi.org/10.1016/j.bspc.2023.105714>
18. Fitzhugh R. Impulses and physiological states in theoretical models of nerve membrane. *Biophys J*. 1961;1(6):445–66. [https://doi.org/10.1016/s0006-3495\(61\)86902-6](https://doi.org/10.1016/s0006-3495(61)86902-6) PMID: 19431309
19. Morris C, Lecar H. Voltage oscillations in the barnacle giant muscle fiber. *Biophys J*. 1981;35(1):193–213. [https://doi.org/10.1016/S0006-3495\(81\)84782-0](https://doi.org/10.1016/S0006-3495(81)84782-0) PMID: 7260316

20. Chay TR. Chaos in a three-variable model of an excitable cell. *Physica D: Nonl Phenomena*. 1985;16(2):233–42. [https://doi.org/10.1016/0167-2789\(85\)90060-0](https://doi.org/10.1016/0167-2789(85)90060-0)
21. Panahi S, Jafari S, Khalaf AJM, Rajagopal K, Pham V, Alsaadi FE. Complete dynamical analysis of a neuron under magnetic flow effect. *Chin J Phys*. 2018;56(5):2254–64. <https://doi.org/10.1016/j.cjph.2018.08.008>
22. Anderson AE. Transistors in switching circuits. *Bell Syst Tech J*. 1952;31(6):1207–49. <https://doi.org/10.1002/j.1538-7305.1952.tb03709.x>
23. Nagumo J, Arimoto S, Yoshizawa S. An active pulse transmission line simulating nerve axon. *Proc IRE*. 1962;50(10):2061–70. <https://doi.org/10.1109/jrproc.1962.288235>
24. Xu Q, Chen X, Chen B, Wu H, Li Z, Bao H. Dynamical analysis of an improved FitzHugh-Nagumo neuron model with multiplier-free implementation. *Nonl Dyn*. 2023;111(9):8737–49. <https://doi.org/10.1007/s11071-023-08274-4>
25. Reddy JN. Introduction to the finite element method. McGraw-Hill Education; 2019.
26. Dragomiretskiy K, Zosso D. Variational mode decomposition. *IEEE Trans Signal Process*. 2014;62(3):531–44. <https://doi.org/10.1109/tsp.2013.2288675>
27. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond A*. 1998;454(1971):903–95. <https://doi.org/10.1098/rspa.1998.0193>
28. Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal*. 2009;01(01):1–41. <https://doi.org/10.1142/s1793536909000047>
29. Zeiler A, Faltermeier R, Keck IR, Tomé AM, Puntonet CG, Lang EW. Empirical mode decomposition - an introduction. In: The 2010 International Joint Conference on Neural Networks (IJCNN). 2010. p. 1–8.
30. Bertsekas DP. Constrained optimization and Lagrange multiplier methods. Academic Press; 2014.
31. Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N. Taking the human out of the loop: a review of Bayesian optimization. *Proc IEEE*. 2016;104(1):148–75. <https://doi.org/10.1109/jproc.2015.2494218>
32. Abdel-Basset M, Abdel-Fatah L, Sangaiah A. Metaheuristic algorithms: a comprehensive review. In: Computational intelligence for multimedia big data on the cloud with engineering applications. 2018. p. 185–231.
33. Wang Z, He G, Du W, Zhou J, Han X, Wang J, et al. Application of parameter optimized variational mode decomposition method in fault diagnosis of gearbox. *IEEE Access*. 2019;7:44871–82. <https://doi.org/10.1109/access.2019.2909300>
34. Slowik A, Kwasnicka H. Evolutionary algorithms and their applications to engineering problems. *Neural Comput Appl*. 2020;32(16):12363–79. <https://doi.org/10.1007/s00521-020-04832-8>
35. SS VC, HS A. Nature inspired meta heuristic algorithms for optimization problems. *Computing*. 2022;104(2):251–69.
36. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Computat*. 2002;6(2):182–97. <https://doi.org/10.1109/4235.996017>
37. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks, vol. 4. IEEE; 1995. p. 1942–8. <https://doi.org/10.1109/icnn.1995.488968>
38. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Adv Eng Softw*. 2014;69:46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
39. Zolf K. Gold rush optimizer: a new population-based metaheuristic algorithm. *Oper Res Decis*. 2023;33(1).
40. Su H, Zhao D, Heidari AA, Liu L, Zhang X, Mafarja M, et al. RIME: a physics-based optimization. *Neurocomputing*. 2023;532:183–214. <https://doi.org/10.1016/j.neucom.2023.02.010>
41. Jia H, Rao H, Wen C, Mirjalili S. Crayfish optimization algorithm. *Artif Intell Rev*. 2023;56(S2):1919–79. <https://doi.org/10.1007/s10462-023-10567-4>
42. Zhong C, Li G, Meng Z. Beluga whale optimization: a novel nature-inspired metaheuristic algorithm. *Knowl-Based Syst*. 2022;251:109215. <https://doi.org/10.1016/j.knosys.2022.109215>
43. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30.
44. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1 (Long and Short Papers); 2019. p. 4171–86.
45. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman F. Gpt-4 technical report. arXiv preprint. 2023. <https://arxiv.org/abs/2303.08774>

46. Sharma S, Sharma S, Athaiya A. Activation functions in neural networks. *Towards Data Sci.* 2017;6(12):310–6.
47. Deng W, Zhang X, Zhou Y, Liu Y, Zhou X, Chen H, et al. An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. *Inf Sci.* 2022;585:441–53. <https://doi.org/10.1016/j.ins.2021.11.052>
48. Antoni J. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mech Syst Signal Process.* 2006;20(2):282–307. <https://doi.org/10.1016/j.ymssp.2004.09.001>
49. Antoni J. Fast computation of the kurtogram for the detection of transient faults. *Mech Syst Signal Process.* 2007;21(1):108–24. <https://doi.org/10.1016/j.ymssp.2005.12.002>
50. Udmale SS, Singh SK. A mechanical data analysis using kurtogram and extreme learning machine. *Neural Comput Appl.* 2019;32(8):3789–801. <https://doi.org/10.1007/s00521-019-04398-0>
51. Hutter F, Hoos HH, Leyton-Brown K. Sequential model-based optimization for general algorithm configuration. In: *Learning and Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17–21, 2011. Selected Papers 5.* Springer; 2011. p. 507–23.
52. Damianou A, Lawrence N. Deep gaussian processes. *Artificial intelligence and statistics.* PMLR; 2013. p. 207–15.
53. Garnett R. Bayesian optimization. Cambridge University Press; 2023.
54. Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. *Adv Neural Inf Process Syst.* 2011;24.