



OPEN An improved ICEEMDAN–depth hybrid network model integrating multimodal data for the screening of diabetic peripheral neuropathy

Mingxia Xiao^{1✉}, Fei Wang¹, Shidong Fang¹, Gaojie Duan¹ & Xiaojing Tang²

Early non-invasive approaches for detecting diabetic peripheral neuropathy (DPN) are crucial to preventing its severe complications. However, these approaches have been limited by insufficient dynamic feature capture, low model efficiency, and poor portability. To improve the non-invasive detection capability for DPN, a novel combined method based on the fusion of PPG and ECG signals is proposed. Firstly, an adaptive denoising method integrating ICEEMDAN-based signal decomposition, wavelet thresholding, and particle swarm optimization is adopted to improve signal quality. Secondly, a combined encoding framework, integrating spatial position encoding, Grampian angular field, and recurrence plot, is employed to transform one-dimensional time-series signal segments into RGB color maps. Finally, an enhanced lightweight network named Afsharid, incorporating multi-branch depth wise convolution and a spatial hybrid self-attention mechanism, is designed to generate fused RGB representations. On the multi-cycle dataset, the proposed model achieved an accuracy of 93.89%, a sensitivity of 93.21%, and a precision of 94.52%. Compared with the best-performing baseline model EfficientNetV2, the accuracy was improved by 6.52%. The results show the feasibility and potential of the combined method as a new solution for early detection and daily monitoring of DPN.

Keywords ICEEMDAN, SHSA, Multimodal, Multibranch inception, DPN

Early non-invasive diagnosis of DPN is a major clinical challenge¹. Conventional diagnostic methods are limited by invasiveness, complexity, and high cost, restricting their accessibility and repeatability. Therefore, non-invasive detection techniques based on PPG and ECG signals have attracted extensive attention due to their complementary information and convenient acquisition². However, effectively fusing these two heterogeneous signals and building a robust automatic diagnosis system still represents a critical technical bottleneck.

In DPN detection, the parameter configuration of existing signal decomposition algorithms lacks stability, limiting their effective application in real-world scenarios with complex noise. This challenge stems primarily from the inherent weakness and high susceptibility to contamination of PPG and ECG signals. To the above demands, a substantial body of research has focused on adaptive signal processing using Empirical Mode Decomposition (EMD) and its variants. However, these methods are hampered by inherent limitations. For instance, while Ensemble EMD (EEMD) alleviates mode mixing by adding Gaussian white noise, it often leaves residual noise in the reconstructed signal, compromising the accuracy of subsequent feature extraction³. Another study employed the Tuna Swarm Optimization (TSO) algorithm to optimize the parameters of Complete EEMD with Adaptive Noise (CEEMDAN). Although specific metrics improved, concerns remain regarding its convergence stability on large-scale datasets, partly due to the small-sample design of the initial experiments⁴. The Improved CEEMDAN (ICEEMDAN) enhances decomposition consistency and stability, yet its performance is sensitive to parameter settings tailored to specific noise types, thus lacking sufficient adaptability in complex, real-world clinical environments⁵. Therefore, a preprocessing method that offers stronger parameter self-adaptability while achieving a better balance between noise suppression and signal fidelity is critically needed to establish a reliable foundation for subsequent analysis.

In feature representation, transforming one-dimensional time-series signals into two-dimensional images for deep learning is a mainstream approach, whose core advantage is converting abstract temporal dependencies

¹School of Electrical and Information Engineering, North Minzu University, No. 204 North Wenchang Street, Yinchuan 750021, Ningxia, China. ²College of Medical Information and Engineering, Ningxia Medical University, No. 1160 Shengli Street, Yinchuan 750004, Ningxia, China. ✉email: xiao_mx@nmu.edu.cn

and dynamic patterns into spatial structures recognizable by convolutional neural networks. However, existing methods still struggle to fully capture the complementary dynamic relationships when fusing PPG and ECG signals. Specifically, GAF preserves temporal dependencies via polar coordinate transformation but increases data dimensionality and computational cost, with performance vulnerable to acquisition device and electrode placement variations⁶. RP characterizes nonlinear dynamic features but relies heavily on empirical critical parameter settings (recurrence threshold and time delay), lacking adaptive optimization and leading to insufficient representational stability⁷. Regional Markov Random Fields (MRF) effectively model local spatial correlations via super pixel segmentation and energy function optimization but poorly capture global temporal information for highly dynamic PPG and ECG signals, with complexity increasing significantly with signal length⁸. Thus, a single image encoding method cannot comprehensively and efficiently extract the complementary value of the two signals.

With the penetration of deep learning technology, it has demonstrated excellent performance in the auxiliary diagnosis of DPN. However, existing studies still have limitations when migrated to the time-series signal analysis of PPG and ECG. For instance, although the Conv-LSTM model⁹ can alleviate the gradient vanishing problem of traditional recurrent neural networks (RNNs) and excel at capturing long-term dependencies of a single signal, its feature extraction is not adapted to the inherent differences between PPG and ECG signals. During fusion, it only relies on simple integration through temporal layers, failing to deeply explore the complementary correlations between the two signals in the pathological process of DPN. This results in the uniqueness and synergy of cross-modal features not being fully utilized. The SAE-CNN mode¹⁰ enhances unimodal features through sparse autoencoders (SAEs) and combines convolutional neural networks (CNNs) to extract local details. Nevertheless, it only processes the feature extraction of the two signals independently without establishing a dynamic synergy mechanism. Moreover, it lacks a dedicated module during fusion and merely relies on shallow concatenation for integration, making it difficult to fully reflect the synergistic correlation value of ECG and PPG signals in reflecting DPN lesions. Even when combined with variational autoencoder (VAE)-based data augmentation to adapt to small-sample scenarios, the generalization performance of the fused features has not been fully verified. Models integrated with interpretability components such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive explanations (SHAP)¹¹ have improved the transparency of decision-making. However, they have not optimized core flaws such as insufficient differentiation in multi-modal feature extraction and shallow-level fusion, failing to solve the problem of synergistic utilization of cross-modal information. Although the quantum machine learning (QML) architecture¹² provides a new perspective for classification tasks, it lacks a differentiated feature capture mechanism and a dedicated fusion module, making it difficult to exert its potential advantages. Although these models have achieved certain results in aspects such as time-series feature capture, unimodal feature enhancement, and improvement of decision-making transparency, they still fail to solve the problems of differentiated adaptation and deep fusion of multi-modal feature extraction, and thus cannot fully exploit the complementary value of the two signals.

To address the limitations of the studies, this research proposes a deep learning framework for fusing physiological signals for non-invasive DPN detection:

1. An adaptive denoising method integrating ICEEMDAN-based signal decomposition, wavelet thresholding, and PSO is adopted to improve signal quality.
2. An SGR algorithm integrated with spatial encoding is proposed. By synergistically combining spatial position encoding, Gramian Angular Field, and Recurrence Plot, one-dimensional time-series signals are converted into two-dimensional image representations that retain dynamic correlation features, thereby providing efficient and information-complete inputs for deep models;
3. An enhanced lightweight network (Eff_SHSAIDC) is constructed, which integrates multi-branch convolution and spatial hybrid self-attention mechanism to achieve efficient extraction and fusion of two-dimensional spatial graph features, and ultimately realize automatic high-precision classification of DPN.

These contributions address three critical technical bottlenecks in non-invasive DPN detection: unstable signal decomposition under complex noise, insufficient dynamic feature encoding during image transformation, and shallow cross-modal fusion in deep learning models. The proposed framework fully exploits the complementary information between PPG and ECG signals via adaptive preprocessing, refined feature encoding, and strengthened cross-modal fusion, thus providing an effective and reliable technical solution for early DPN diagnosis.

The organization of this paper is as follows: Section “[Related work](#)” reviews related work on DPN detection, machine learning, deep learning, and reference methods. Section “[Materials and Methods](#)” describes the dataset, feature fusion, experimental setup, and the proposed deep learning framework based on fused PPG and ECG signals. Section “[Experimental results](#)” presents the experimental results and performance comparisons. Section “[Discussion](#)” discusses the findings, limitations, and future work. Section “[Conclusion](#)” concludes the paper.

Related work

In existing research on DPN assessment, early work primarily focused on the analysis of single physiological signals. For instance, ECG-based QTc interval analysis has been used to evaluate autonomic nerve function¹³, while PPG waveform feature extraction has been employed to reflect peripheral vascular status¹⁴. These studies preliminarily established the association between physiological signals and DPN pathology.

With the advancement of machine learning technologies, researchers began to explore the use of traditional models to mine diagnostic information from signals. For example, Bayesian classifiers have been applied to identify neuropathy risk from ECG features¹⁵, and shallow neural networks have been utilized for classifying parameters derived from PPG¹⁶. These methods provided preliminary solutions for the automated assessment of DPN.

In recent years, deep learning has demonstrated significant advantages in the field of biomedical signal processing. CNN¹⁷ and Recurrent Neural Networks (RNN)¹⁸, along with their variants, have been successfully applied to tasks such as arrhythmia detection¹⁹, sleep stage classification²⁰, and medical image analysis²¹. Their powerful capability for nonlinear modeling provides a novel tool for characterizing complex physiological and pathological relationships. This trend has extended to DPN research, prompting scholars to explore more advanced architectures.

Notably, the success of multimodal physiological signal fusion and complex deep learning models in related fields provides important references for DPN assessment. In the brain-computer interface domain, frameworks that fuse electroencephalography (EEG) and functional near-infrared spectroscopy (firs) signals combined with optimized CNNs have significantly improved classification performance²². In the field of cardiovascular monitoring, hybrid methods that fuse ECG and PPG signals, incorporating Windkessel models, autoregressive integrated moving average (ARIMA) models, and long short-term memory (LSTM) networks, have effectively enhanced the accuracy of non-invasive blood pressure estimation²³. Furthermore, advancements in Explainable Artificial Intelligence (XAI) technologies for improving model trustworthiness²⁴ also offer insights for the clinical application of medical diagnostic models.

In summary, while methods based on single-signal analysis or traditional machine learning have laid the foundation for DPN assessment, fully leveraging the complementary information from PPG and ECG, along with drawing upon advanced deep learning fusion architectures and explainability techniques, holds significant promise for advancing the field. Building upon this foundation, the present study proposes a deep learning framework for physiological signal fusion designed for the non-invasive detection of DPN.

Materials and methods

Participants

This study was approved by the Biomedical Research Ethics Committee of North Minzu University (Approval No. 2024-2). A total of 120 participants were recruited, comprising 43 healthy volunteers, 32 patients without DPN, and 45 patients with DPN, as detailed in Table 1. All patients were diagnosed according to the classification criteria outlined in the 2021 edition of the Chinese Expert Consensus on the Diagnosis and Treatment of Diabetic Neuropathy. Participants with conditions such as cardiac arrhythmias or neuropathies from other causes were excluded. All participants provided informed consent, and this study was conducted in strict accordance with the ethical principles of the Declaration of Helsinki.

Equipment and collection methods

The experimental setup is shown in Fig. 1. A self-developed six channel ECG PWV synchronous acquisition system was used. Before data collection, participants abstained from caffeine and theophylline for at least 12 h and took a morning fasting blood test. ECG and PPG signals were synchronously recorded in a 26 ± 1 °C consultation room between 8:00 and 10:00 AM to reduce motion noise. Signal acquisition utilized an infrared sensor with a wavelength of 940 nm, fixed in a clip-on form to the participant's left index finger to obtain waveforms by sensing blood volume changes. After converting the received optical signal into an analog electrical signal, filtered by a 0.48–10 Hz second order bandpass filter, amplified by a 1–10 mV circuit, finally digitized at 500 Hz using a USB 6008 DAQ card, ultimately preserving the raw ECG and PPG signals for each participant.

For each dataset obtained from the subject groups, the raw signals were first segmented into cycles using a 6 s short-term sliding window, followed by normalization. The processed data then underwent signal denoising via the PSO-optimized ICEEMDAN combined with wavelet thresholding algorithm. Subsequently, based on the SGR algorithm, position, phase, and period-related information from the ECG and PPG sequence points were utilized to construct a three-channel image representation. Finally, the Eff_SHSAIDC hybrid network model was employed, which replaces traditional convolutions with a multi-branch Inception structure to capture multi-directional features, integrates SHSA, and utilizes a dynamic feature fusion strategy to adaptively integrate multi-scale features and attention-enhanced representations. This enables the model to balance lightweight characteristics with effective extraction of deep signal features relevant to DPN. The system provides an auxiliary tool for clinical DPN diagnosis that combines high performance with interpretability.

Significant parameters	Group 1 (n = 43)	Group 2 (n = 32)	Group (n = 45)
Age (years)	57.53 ± 8.96	64.44 ± 9.80	62.06 ± 10.99
BMI (kg/m ²)	25.00 ± 3.45	26.13 ± 3.19	26.65 ± 5.73
HbA1c (%)	5.92 ± 0.38	6.45 ± 0.45**	8.53 ± 1.72**
Blood Sugar AC (mg/dL)	100.44 ± 14.49	120.13 ± 23.50**	169.24 ± 55.48**
Cholesterol (mg/dL)	200.91 ± 39.14	170.25 ± 38.52**	182.51 ± 39.50
Triglyceride	116.37 ± 80.38	121.50 ± 80.03	145.70 ± 67.00*

Table 1. Basic human physiological parameters of the participants in the three groups, n: number of people. Where * indicates $p < 0.05$, meaning there is a statistical difference between groups; ** indicates $p < 0.001$, meaning there is a significant statistical difference between groups.

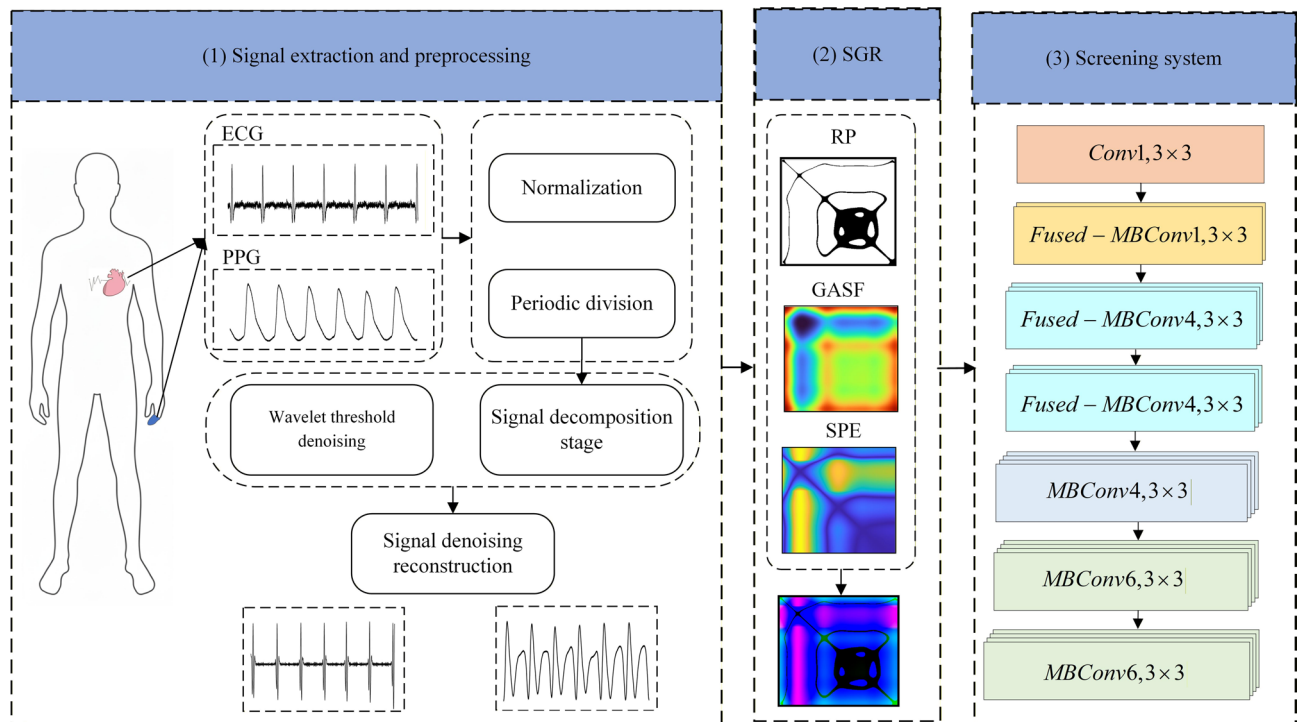


Fig. 1. Architecture diagram of DPN detection classification. The framework comprises three key modules: (1) signal extraction and preprocessing for denoising and structuring raw ECG and PPG signals; (2) the SGR module, which fuses SPE, GASF, and RP to convert 1D signals into 2D image representations; and (3) a multi-branch convolutional screening system for DPN classification.

PSO_ICEEMDAN combined with wavelet threshold algorithm

Acquired ECG and PPG signals are often corrupted by baseline drift, power line interference and electromyographic noise, severely undermining the reliability of subsequent analysis, making effective denoising indispensable. Traditional Ensemble Empirical Mode Decomposition (EEMD)²⁵ decomposes signals via repeated random noise addition and ensemble averaging for denoising, yet still suffers from mode mixing; direct elimination of high-frequency noise-containing components also easily causes useful information loss. To address this, the Complete Ensemble EMD with Adaptive Noise (CEEMDAN)²⁶ was proposed, which mitigates mode mixing by adding adaptive-intensity Gaussian white noise to residual signals, but retains residual noise due to raw noise introduction, impairing decomposition accuracy. The Improved CEEMDAN (ICEEMDAN)²⁷ remedied this by a key refinement: instead of raw Gaussian white noise, it introduces specific K-th order Intrinsic Mode Function (IMF) components²⁸ derived from EMD decomposition of white noise. This structured noise component integration boosts noise suppression capability, enhancing denoising performance and signal fidelity; leveraging the ensemble averaging framework, ICEEMDAN also ensures decomposition stability and accuracy. Thus, the performance of ICEEMDAN is highly dependent on the proper selection of its key parameters in practical applications.

PSO algorithm is an evolutionary algorithm²⁹. It initializes a swarm of random particles, where each particle cooperatively explores and moves through the search space based on its own historical best experience and the global best experience of the swarm. Through iterative optimization, each particle adjusts its velocity and position according to the following formulas²⁹.

$$v_i(t+1) = \omega \times v_i(t) + c_1 \times r_1 \times (p_i - x_i(t)) + c_2 \times r_2 \times (g - x_i(t)) \quad (1)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2)$$

where $i = 1, 2, 3, \dots, N$, and N is the total number of particles in the swarm; v_i is the velocity of the particle; x_i is the current position of the particle; c_1 is the individual learning factor (cognitive coefficient); c_2 is the social learning factor (social coefficient); p_i is the personal best position of the i -th particle; g is the global best position of the entire swarm; ω is the inertia factor, whose value controls the optimization capability.

Therefore, this study integrates the PSO optimization algorithm with ICEEMDAN. First, the PSO algorithm is employed to optimize the key parameters of ICEEMDAN and select their optimal values. Next, the ICEEMDAN algorithm is used to decompose the signal. Subsequently, the correlation coefficient between each IMF component and the original signal is calculated. A higher correlation coefficient indicates that the IMF component contains more features of the original signal, suggesting less influence from noise. IMF components with moderately high correlation coefficients undergo wavelet threshold denoising, while those with correlation

coefficients close to zero are discarded. The screened IMF components are then reconstructed to obtain the high-quality signals required for subsequent analysis. The flowchart of this algorithm is shown in Fig. 2 below.

Spatial encoding fusion feature

After obtaining denoised, high-quality PPG and ECG signals, single-signal analysis of PPG alone reveals that its conventional time-domain features lack sufficient sensitivity for distinguishing DPN-related pathological states, with significant feature overlap across different subject groups. Thus, multi-signal collaborative analysis is essential: this entails deeply mining the dynamic and nonlinear pathological characteristics of PPG signals, while integrating complementary cardiac electrical activity information from ECG³⁰. Their synergy enables a more comprehensive characterization of DPN-associated pathological mechanisms. To this end, three spatial encoding algorithms are introduced in this study.

Spatial position encoding (SPE) fuses spatial location information with signal features to boost the model's discriminatory capability³¹. Different spatial positions within the same signal may exhibit distinct feature patterns that reflect blood flow dynamics; the transformed SPE matrix helps the model capture more global and local sequential dependencies, thus enhancing feature generalizability. Specifically, the spatial position

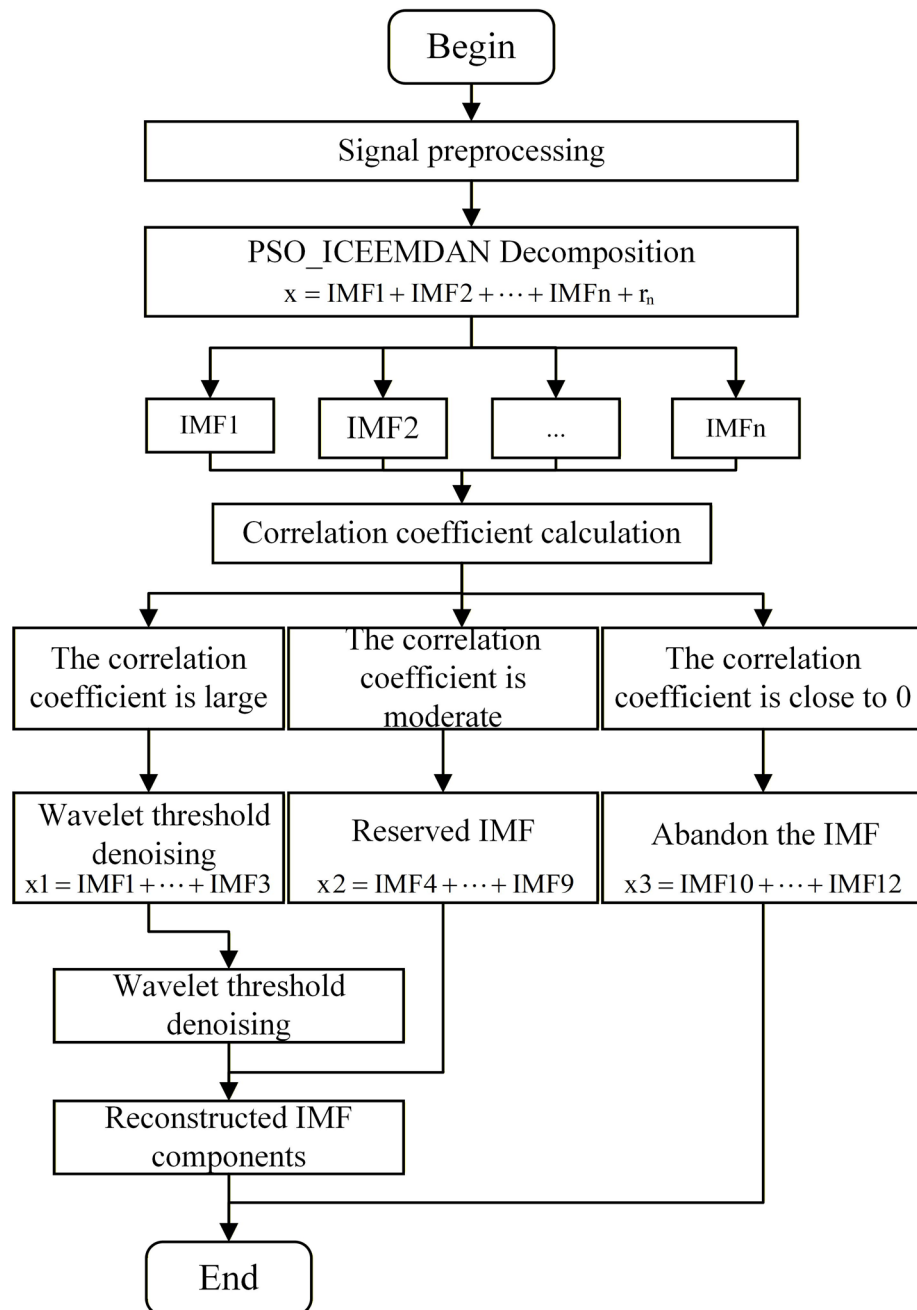


Fig. 2. PSO_ICEEMDAN combined with wavelet threshold algorithm denoising flowchart.

information between any two points in the sequence is calculated sequentially using the Euclidean norm, as shown in the following formula³¹:

$$\text{SPE}_{ij} = \|\vec{x}_i - \vec{x}_j\| = \sqrt{(\vec{x}_i - \vec{x}_j)^T (\vec{x}_i - \vec{x}_j)}, i, j \in [0, m] \text{ SPE} \in \mathbb{R}^{m \times m} \quad (3)$$

Gramian angular field (GAF) maps each time point to the corresponding angular and polar radius values by calculating the differences and relative angles between time points. It uses a polar coordinate system to reveal dynamic evolutionary patterns across time points, thus enabling in-depth exploration of the implicit vascular blood flow fluctuation characteristics in PPG signals³².

$$\phi_i = \arccos(\vec{x}_i), r_i = \frac{i}{m}, i \in [0, m] \quad (4)$$

Among them, ϕ_i serves as the angular vector, and r_i serves as the radius. By utilizing the angles between the different points, the following DAF can be obtained³².

$$\text{GASF}_{ij} = [\cos(\phi_i + \phi_j)] = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_j) \\ \cos(\phi_2 + \phi_1) & \cdots & \cos(\phi_2 + \phi_j) \\ \vdots & \vdots & \vdots \\ \cos(\phi_i + \phi_1) & \cdots & \cos(\phi_i + \phi_j) \end{bmatrix} \quad (5)$$

Recurrence plot (RP) maps 1D time series to a high-dimensional phase space via phase space reconstruction, and is well-suited for non-stationary, short-period time series signals. Converting 1D ECG and PPG signals into 2D recurrence plots not only preserves intrasignal dynamic information and uncovers its hidden structures, but also characterizes the inherent nonlinear features of the signals³³.

$$\text{RP}_{ij} = \Phi(\lambda - \|\vec{x}_i - \vec{x}_j\|), i, j \in [0, m] \quad (6)$$

$$\Phi(\cdot) = \begin{cases} 1, (\lambda - \|\vec{x}_i - \vec{x}_j\|) \geq 0 \\ 0, (\lambda - \|\vec{x}_i - \vec{x}_j\|) \leq 0 \end{cases} \quad (7)$$

Here, the threshold λ is set to 0.1 (with the normalized peak value being 1, representing 10% of the peak value) and serves as the parameter for the step function.

In summary, this paper takes short-term electrocardiogram and pulse wave signal sequences as the original input, and fuses the above three spatial encoding strategies into the three channels of an RGB image to construct the SGR encoding, thereby achieving effective multi-dimensional dynamic image fusion. Subsequently, the generated images are uniformly resized to a standard resolution to satisfy the input requirements of the subsequent model. The preprocessed signals can be represented as $X = \{x_1, x_2, \dots, x_n\}$, where n denotes the sequence length, and the values are normalized to the range $[0, 1]$.

Detection of diabetic peripheral neuropathy based on Eff_SHSAIDC

Convolutional neural network

Using the preprocessed image dataset, to develop an efficient and accurate classification model, this study selects and modifies the lightweight convolutional neural network EfficientNetV2³⁴ as the core architecture. Derived from EfficientNetV1, this network retains the Mobile Inverted Bottleneck (MBCConv) module and introduces the Fused Mobile Inverted Bottleneck (Fused-MBCConv) module, which effectively mitigates the reduced training speed caused by depthwise separable convolutions in the network's shallow layers. The EfficientNetV2 improvements in this study are illustrated in Fig. 3, and the modified network structure is divided into three components: Fused-MBCConv layers, MBCConv layers, and the output layer.

Multiscale feature extraction and reuse collaborative optimization

The original Fused-MBCConv structure has inherent limitations in feature reuse and multi-scale information fusion, especially in deep networks where it suffers from gradient attenuation and insufficient feature representation capacity. This hampers its adaptability to extracting multi-scale, directional features from physiological signal images for DPN detection tasks. Traditional convolutional kernels face a trade-off between computational complexity and feature capture capability when modeling long-range dependencies and local details: large kernels capture global features but drastically increase computational load, while small kernels are computationally efficient but lack long-range dependency modeling ability. To address this, this study designs a multi-branch Inception depthwise convolutional module, as shown in Fig. 4, which integrates the lightweight nature of depthwise separable convolutions with the multi-scale capture advantages of asymmetric strip kernels, following the Inception parallel multi-scale processing paradigm. It decomposes standard convolution into a combination of differently oriented strip convolutions and depthwise convolutions, with a three-step workflow: first, the input feature map is split along the channel dimension into one identity branch and three convolutional branches $X = [X_{id}, X_{hw}, X_w, X_h]$; next, depthwise convolutions are used to capture multi-dimensional features separately, where a 3×3 kernel is applied for the local spatial feature branch, a 1×11 kernel for the horizontal long-range branch, and an 11×1 kernel for the vertical long-range branch. The corresponding formulation is given as follows:

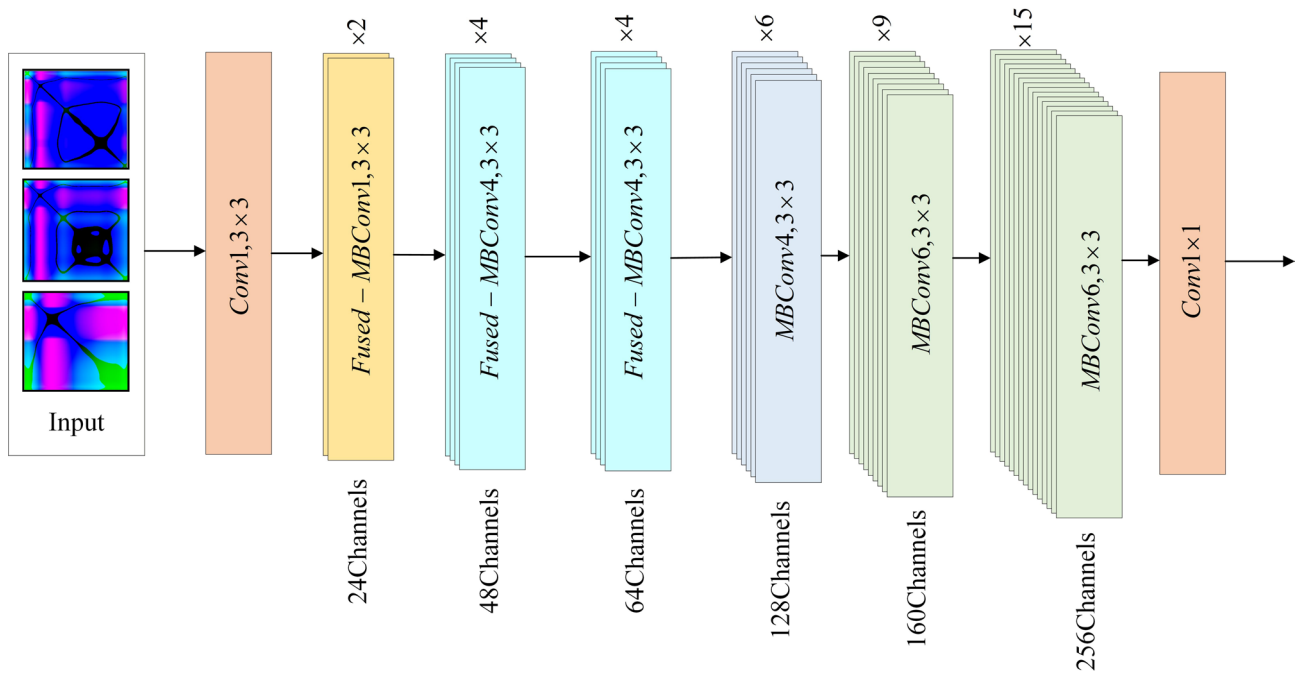


Fig. 3. EfficientNetV2 model structure diagram.

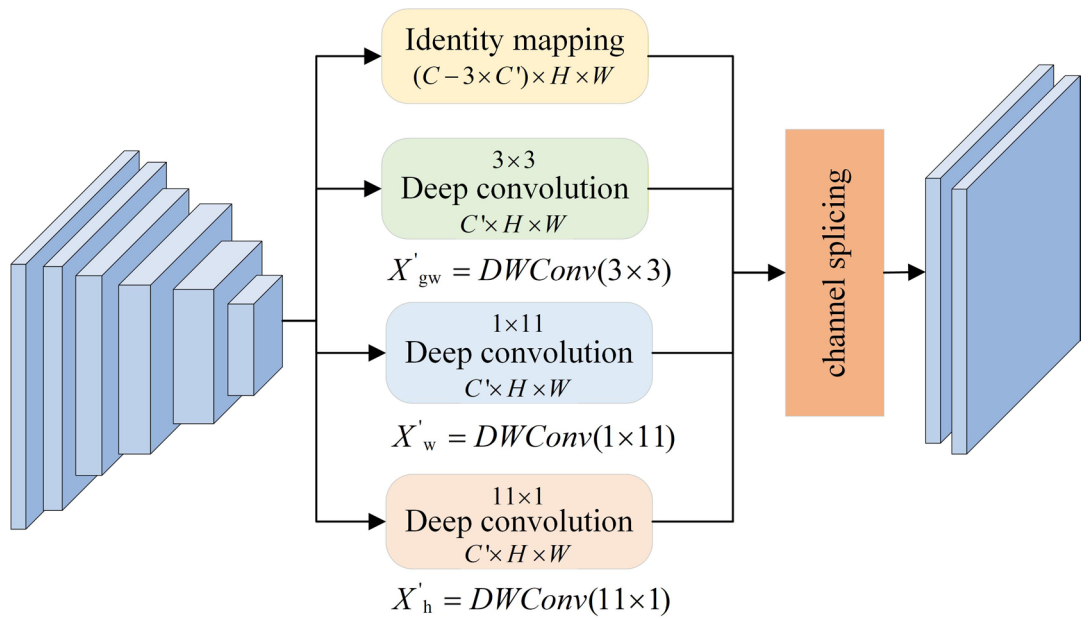


Fig. 4. Multi-branch inception deep convolutional structure diagram.

$$X'_{hw} = DWConv(X_{hw}, k = (3, 3), p = (1, 1)) \tag{8}$$

$$X'_w = DWConv(X_w, k = (1, 11), p = (0, 5)) \tag{9}$$

$$X'_h = DWConv(X_h, k = (11, 1), p = (5, 0)) \tag{10}$$

DWConv represents the depthwise convolution operation, where k denotes the kernel size and p denotes the padding. Finally, the output $X_{out} = Cat[X_{id}, X'_{hw}, X'_w, X'_h]$ is obtained through channel concatenation.

This figure depicts the multi-branch Inception deep convolution architecture. The input feature $X \in R^{C \times H \times W}$ is split into one identity branch and three convolutional branches (using 3×3 , 1×11 , and

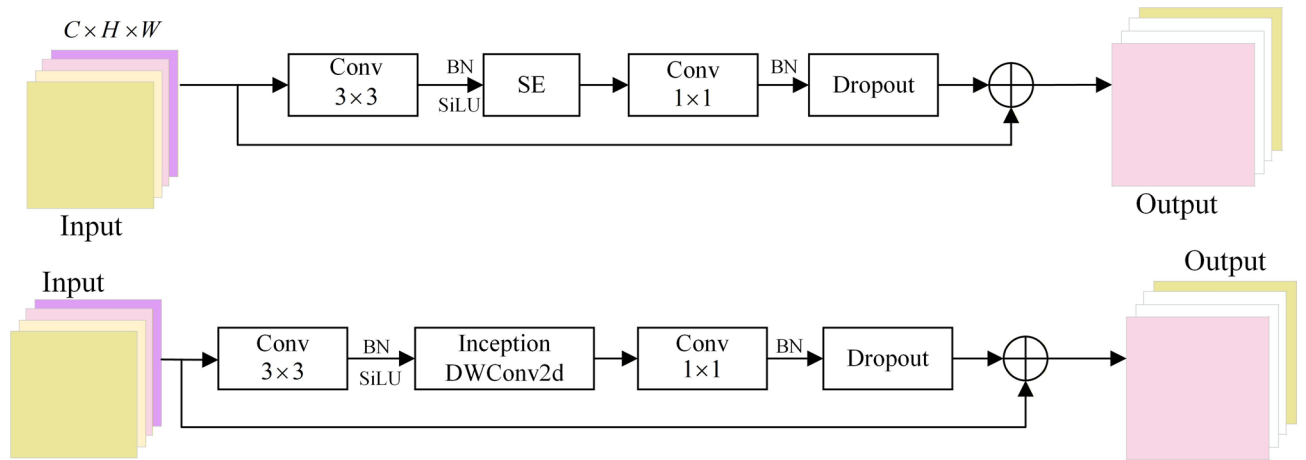


Fig. 5. Comparison chart of the improved Fused-MBConv layer structure.

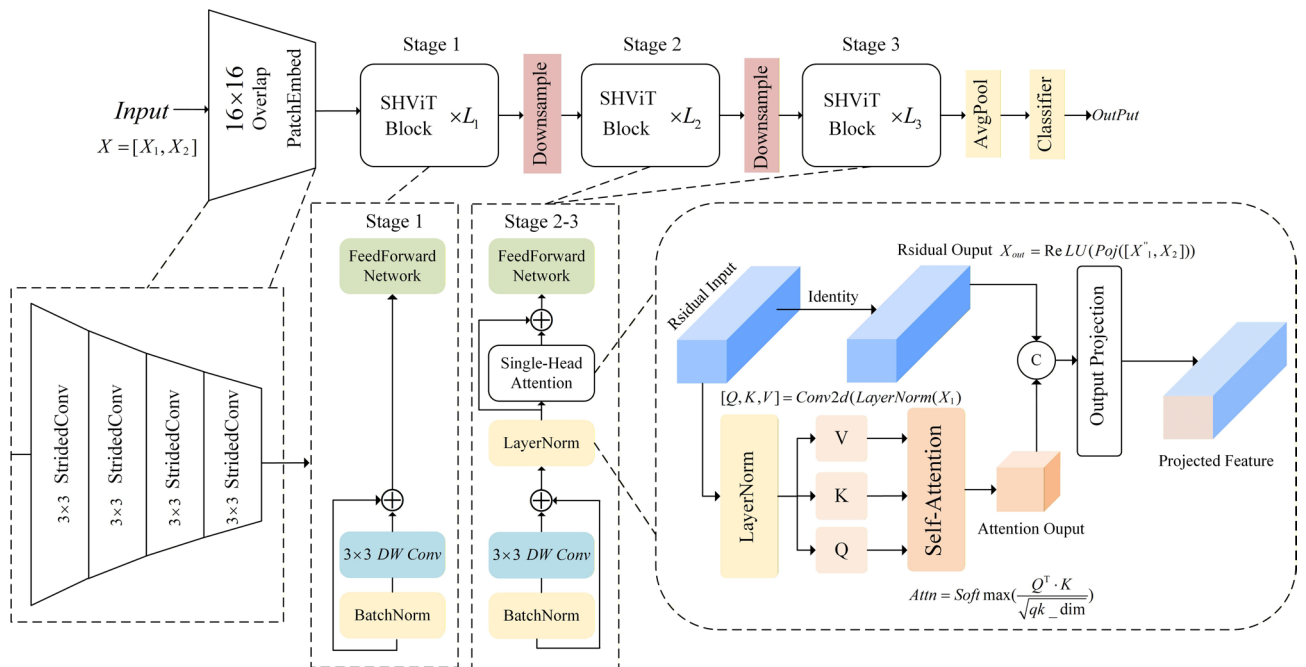


Fig. 6. Structure diagram of SHSA space and hybrid self-attention mechanism.

11 × 1 depthwise convolutions, respectively). After multi-scale spatial feature extraction, branch outputs are concatenated along the channel dimension to produce the final output with C channels.

To further enhance the performance of the Fused-MBConv, this study replaces its original partial structure with the aforementioned multi-branch Inception depthwise convolutional module. The improved structure, as shown in Fig. 5, retains the efficient feature flow mechanism of Fused-MBConv while strengthening its capability to capture features across different scales and orientations. This achieves a significant improvement in model performance with only a limited increase in parameter count and computational cost, thereby better meeting the requirements of the DPN detection task.

Context awareness and multi-scale feature enhancement in synergy

SHSA is a self-attention mechanism that balances lightweight design with multi-dimensional feature enhancement. Its core objective is to simultaneously capture local spatial details and global semantic correlations, ensuring comprehensive feature representation while controlling computational overhead. Its structure is illustrated in Fig. 6. Specifically, the module divides the input features into two branches along the channel dimension: the core feature branch (X_1) and the identity branch (X_2). X_2 directly participates in subsequent feature concatenation, while X_1 , after normalization, is mapped via a 1 × 1 convolution to generate the Query (Q), Key (K), and Value (V) vectors. The corresponding formulation is given as follows:

$$(Q, K, V) = \text{Conv2d}(\text{LayerNorm}(X_1)) \quad (11)$$

where $Q, K \in R^{B \times qk_dim \times HW}$, qk_dim denote the attention head dimension, $V \in R^{B \times p_dim \times HW}$. A scaled dot-product mechanism is employed to compute the attention weights to mitigate gradient vanishing, i.e.,

$$\text{Attn} = \text{Softmax}\left(\frac{Q^T \cdot K}{\sqrt{qk_dim}}\right) \quad (12)$$

After weighted fusion and further refinement of local spatial features via a 3×3 depthwise convolution, the result is concatenated with X_2 along the channel dimension. Finally, it is passed through a projection layer to output $X_{out} = \text{ReLU}(\text{Proj}([X_{H1}, X_2]))$, achieving dual enhancement of both global semantic correlation and local spatial structure.

This figure shows the overall architecture of SHSA. The input is divided into 16×16 overlapping patches, processed by stacked SHViT Blocks and downsampling layers, and fed into a classifier via global average pooling. The sub-modules detail the structure of SHViT Blocks and the core pipeline of the single-head self-attention module in SHSA, including channel splitting, QKV generation, and attention computation.

While the MBConv structure maintains lightweight characteristics through depthwise convolution, its inverted residual design inadequately captures the global contextual information of input features, and the single-scale depthwise convolution limits the richness of spatial features. To enhance global perception and multi-scale spatial representation, we introduce a lightweight SHSA module after the expansion layer, strengthening the model's ability to model long-range dependencies. Simultaneously, a multi-branch Inception design is integrated during the depthwise convolution stage, leveraging different convolutional kernels to extract multi-scale features in parallel. This improvement effectively enhances the module's capability for global context integration and spatial feature diversity without significantly increasing computational cost. The improved structure is shown in Fig. 7.

The proposed module begins by applying a 1×1 convolution to expand the input channels to a higher dimension, enhancing feature representation capacity. Depthwise separable convolution is then employed for efficient feature extraction. This structure decomposes standard convolution into depthwise (channel-wise) and pointwise convolution, significantly reducing parameter count. After depthwise convolution, the SHSA module is inserted to further refine the extracted features. The spatial attention mechanism focuses on local spatial structures, enabling the model to emphasize target geometry, while the hybrid self-attention captures long-range dependencies and contextual information, such as inter-object interactions. This multi-scale, multi-dimensional enhancement strategy allows the model to understand image content more comprehensively. Finally, a 1×1 convolution compresses the feature channels back to the original dimension, completing the module.

Experimental setup and evaluation indicators

Hyperparameter configuration and training scheme

To optimize performance while maintaining computational efficiency, we combined Bayesian optimization³⁵ with grid search to tune key hyperparameters, including initial learning rate, weight decay, batch size, and augmentation magnitude. The final hyperparameter configuration, as presented in Table 2, was selected based on the highest validation accuracy achieved after 100 epochs. The model was trained for 100 epochs using the AdamW optimizer, which enhances generalization through decoupled weight decay³⁶. A cosine annealing learning rate schedule³⁷ was applied, decaying the learning rate from an initial value of 0.001 to 10^{-8} over the training course. To effectively leverage pre-trained weights and stabilize fine-tuning, a progressive unfreezing strategy³⁸ was employed. Specifically, only the classification head and attention modules were updated for the first 10 epochs. After the 10th epoch, intermediate feature layers were unfrozen, and after the 30th epoch, the full backbone network was made trainable. The entire training process was conducted in automatic mixed precision³⁹ to accelerate computation and maintain numerical stability.

To ensure the statistical robustness and generalization ability of the final reported results, we adopted a fivefold cross-validation strategy³⁹. The entire dataset was initially partitioned into five equally sized and non-overlapping subsets, ensuring that each data point was used exactly once for validation. The model was then trained and evaluated five times, with each fold serving as the validation set in turn while the remaining four

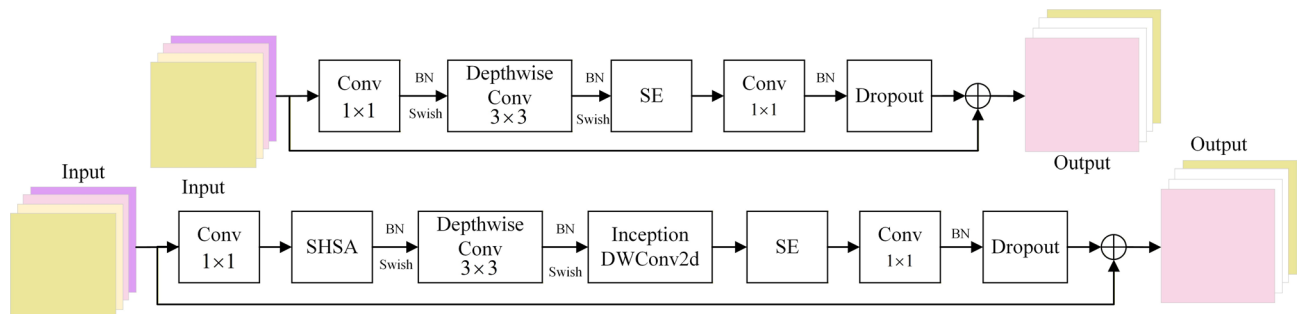


Fig. 7. Comparison chart of the improved MBConv layer structure.

Category	Parameter	Value
Training and Optimization	Optimizer Type	AdamW
	Initial Learning Rate	0.001
	Final Learning Rate Factor	1e-05
	Final Learning Rate	1e-08
	Weight Decay	0.001
	Total Epochs	100
	Batch Size	16
Training strategy	Number of work Processes	8
	Learning Rate Scheduler	Cosine Annealing
	Gradient Scaler Growth Interval	100
	Gradient Scaler Backoff Factor	0.5
	Learning Rate Decay Factor	0.5
	Freeze layer strategy	True
	Unfreeze Epoch 1	10
Unfreeze Epoch 2	30	

Table 2. Comprehensive hyper-parameter configuration for model training.

folds were used for training. This rigorous methodology ensures that the reported performance is not a fortuitous outcome of a single train-test split, thereby enhancing the credibility of the research results.

Dataset division and evaluation metrics

To ensure the model's generalization ability and prevent data leakage, this study adopted a subject-wise data partitioning strategy. The original dataset comprised 120 subjects, who were divided into training, validation, and test sets following a 7:2:1 ratio. Specifically, the training set included 84 subjects, the validation set 24 subjects, and the test set 12 subjects. All partitioning was performed at the subject level, ensuring that all data segments from the same subject appeared exclusively in one of the sets, thereby guaranteeing their independence. This approach effectively prevented data leakage and provided a solid foundation for the reliability and reproducibility of the experimental results.

The model's performance was evaluated using standard detection metrics, including Accuracy (ACC), Sensitivity (SEN), Precision (PRE), F1-Score, and Receiver Operating Characteristic (ROC). These metrics are defined as follows³³:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (13)$$

$$Sensitivity/Recall = \frac{TP}{TP + FN} \quad (14)$$

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$F1_Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (16)$$

In the above, TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative respectively. These are used to show the classification detection performance of the model.

Experimental results

Signal preprocessing analysis

To optimize the performance of the ICEEMDAN algorithm, this study proposes an improved algorithm based on Particle Swarm Optimization (PSO-ICEEMDAN), designed to automatically search for key parameters: the optimal number of decomposition modes and the standard deviation of the noise. The effectiveness of the algorithm was evaluated through a comprehensive quantitative assessment using three metrics. Envelope entropy served as the objective function to quantify the complexity and disorder degree of the signal envelope, directly reflecting the quality of the decomposed Intrinsic Mode Function components. Signal-to-noise ratio (SNR) and root mean square error (RMSE) were used as independent validation metrics, assessing the final denoising effectiveness from the dimensions of noise suppression level and signal fidelity, respectively.

This optimization algorithm simulates the collective intelligent search of a particle swarm within the parameter space, where each particle iteratively updates its position based on its own historical best position and the swarm's global best position, ultimately converging to the global optimal parameter combination that minimizes the envelope entropy. Figure 8 illustrates the dynamic trajectory of the objective function value descending and eventually converging with increasing iteration count, providing intuitive verification of the algorithm's effectiveness and stability in optimization.

The convergence trajectory of the envelope entropy (objective function) versus iteration number is shown, verifying the stability and effectiveness of the parameter search process.

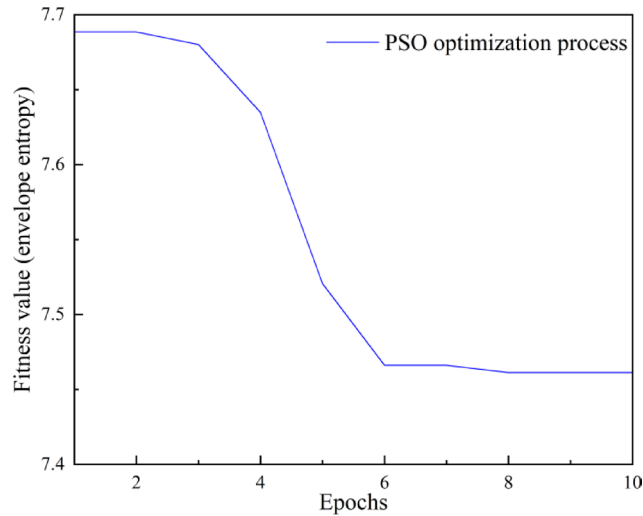


Fig. 8. PSO optimization curve chart.

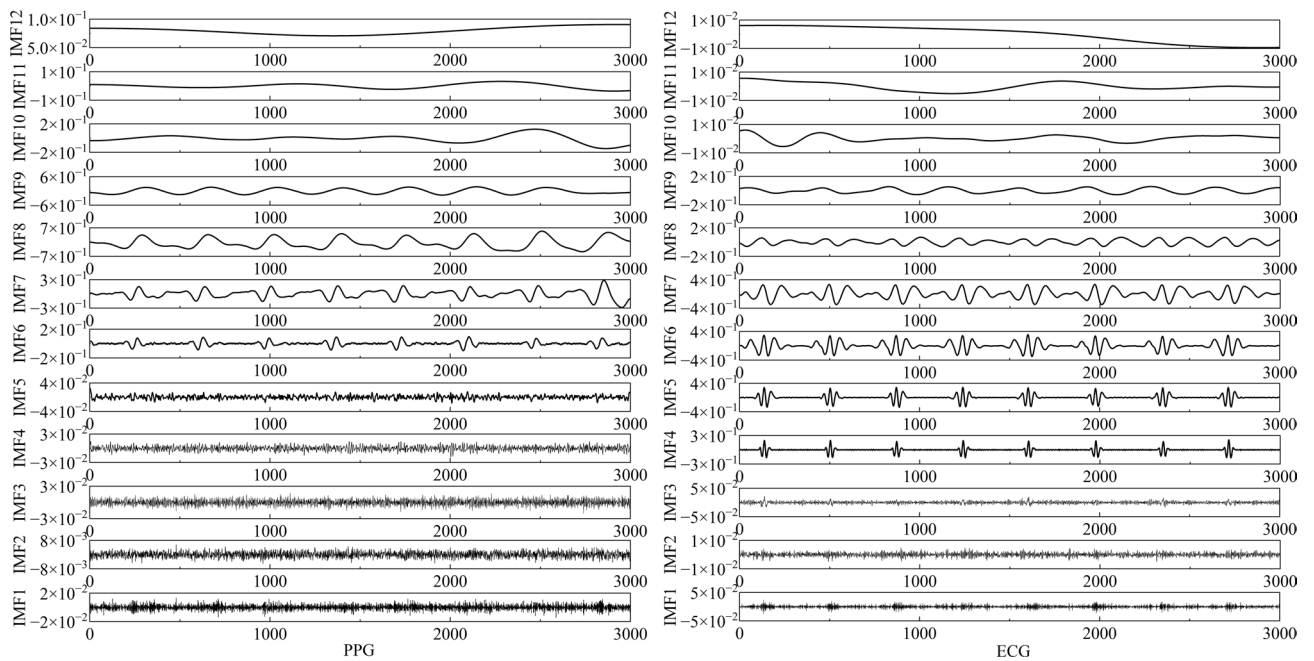


Fig. 9. Signal decomposition diagram.

During the initial iteration phase, the envelope entropy decreased with the number of iterations. Subsequently, the slope of the curve gradually flattened, and the envelope entropy stabilized around 7.4612, forming a steady plateau. This indicates that the particle swarm had converged to the global optimal state, with the corresponding parameters representing the optimal solution. The experimental results show that the optimal parameters converged to $N_{std}=0.1$ and $K=7$, with a minimum envelope entropy of 7.4612. The optimized ICEEMDAN algorithm decomposed the signal into 12 IMF components, as shown in Fig. 9. Among them, the high-frequency IMFs (IMF1–IMF4) primarily contain noise, with correlation coefficients to the original signal all < 0.03 . The mid-frequency IMFs (IMF5–IMF9) capture physiological fluctuation information, with correlation coefficients ranging from 0.03 to 0.95. Among these, IMF8 exhibited the highest correlation coefficient of 0.9534, indicating it contains the core pulse signal. The residual component (IMF12) reflects the overall trend of the signal.

The left figure shows the IMF components of PPG decomposition, and the right figure shows the IMF components of ECG.

To systematically evaluate the denoising performance of the proposed PSO_ICEEMDAN wavelet threshold algorithm, this study conducted comparative experiments against four benchmark algorithms: EEMD, VMD,

Decomposition algorithm	ECG		PPG	
	SNR/dB	RMSE	SNR/dB	RMSE
EEMD	3.8235	0.2439	2.3510	0.4033
VMD	6.7606	0.1767	8.6400	0.1589
ICEEMDAN	12.4482	0.0516	11.7579	0.1270
PSO_ICEEMDAN	15.6542	0.0357	19.1254	0.0508
PSO_ICEEMDAN wavelet	16.7898	0.0258	20.1372	0.0477

Table 3. Analysis of the performance indicators of biological signal denoising for different decomposition algorithms.

Dataset	Input sequence	ACC (%)	SEN (%)	PRE (%)	F1_Score (%)
Train set	One	88.33 ± 0.59	88.62 ± 0.61	88.55 ± 0.60	88.29 ± 0.62
	Three	92.22 ± 0.61	92.90 ± 0.60	92.12 ± 0.58	92.12 ± 0.31
Validation set	One	92.35 ± 0.63	91.089 ± 0.59	81.89 ± 0.57	91.89 ± 0.58
	Three	93.89 ± 0.63	93.21 ± 0.58	94.52 ± 0.67	93.87 ± 0.61

Table 4. Comparison of performance indicators between single cycle and multi cycle signals.

ICEEMDAN, and PSO_ICEEMDAN. The experiments were performed on both PPG and ECG signals, using signal-to-noise ratio and root mean square error as quantitative evaluation metrics.

The results are presented in Table 3. For ECG signals, the proposed algorithm significantly outperformed all comparative methods in both evaluation metrics. Its root mean square error was reduced by 0.0672, 0.1251, 0.0159, and 0.0099 compared to the other four algorithms, respectively. Correspondingly, its signal-to-noise ratio increased by 2.9371, 5.6876, 3.206, and 1.1356. These results validate the synergistic advantage of combining PSO optimization with ICEEMDAN and wavelet threshold denoising, demonstrating performance superior to any single technique or partial combination. Furthermore, the experimental results reveal differences in the algorithm's adaptability to signal types, with the denoising performance for PPG signals being generally superior to that for ECG signals. PPG signals typically exhibit stronger nonlinearity and non-stationarity, and the adaptive decomposition and thresholding mechanisms within this algorithm demonstrate a better capability to capture and process such complex signal characteristics.

Performance analysis on different datasets

In this study, we conducted a comparative analysis of single-cycle and multi-cycle signal segments, and presented the corresponding experimental process and results. In the pulse fluctuation scenario, the oscillation of a single cycle is easily influenced by the previous cycle, and has a strong correlation with the subsequent adjacent cycle; meanwhile, through the processing of waveform data, it can be known that the multi-cycle signal waveform is more stable and complete. Therefore, for the evaluation of multi-cycle signal segments, the current segment needs to be combined with the adjacent previous and subsequent segments to form a new $X_{new} = Combined(X_{n-1}, X_n, X_{n+1})$ comprehensive segment for analysis.

To provide a clearer and more intuitive comparison of the differences, the number of iterations was set to 100, and the learning curves for loss and accuracy were plotted, each distinguished by different colors. The experiments employed the same Eff_SHSAIDC model, with identical parameter selections and variable settings, ensuring the authenticity and rationality of the experimental results.

To verify the effectiveness of periodic sequence signals with different configurations in DPN detection, this study conducted experiments on both single-cycle and multi-cycle sequence datasets separately. The experimental results are presented in Table 4. When trained on single-cycle sequences, the model achieved an accuracy of 88.33% on the training set and 92.35% on the validation set, respectively. In contrast, with multi-cycle sequences as the input, the model performance was further improved, reaching an accuracy of 92.22% on the training set and 93.89% on the validation set. Overall, the Eff_SHSAIDC model exhibited superior performance in terms of multiple evaluation metrics including specificity, sensitivity, and F1-score on the multi-cycle dataset, representing a 1.54% improvement compared with the results on the single-cycle dataset. This verifies the effectiveness of the multi-cycle sequence processing strategy. Figure 10 presents a performance comparison of the Eff_SHSAIDC network model across different datasets.

The confusion matrix intuitively reflects the classification performance of the Eff_SHSAIDC model on different datasets. The rows of the matrix represent true labels, and the columns represent predicted labels. The categories include healthy individuals, non-DPN patients, and DPN patients. As shown in Fig. 11, the dark blue squares on the diagonal represent correctly classified samples, whereas the off-diagonal squares represent misclassified samples. The analysis results indicate that although the model based on the single-cycle dataset performs well overall, it exhibits misclassification when distinguishing non-DPN patients. In contrast, using multi-cycle datasets enhances the model's discriminative ability, particularly for the easily confused non-DPN patient category.

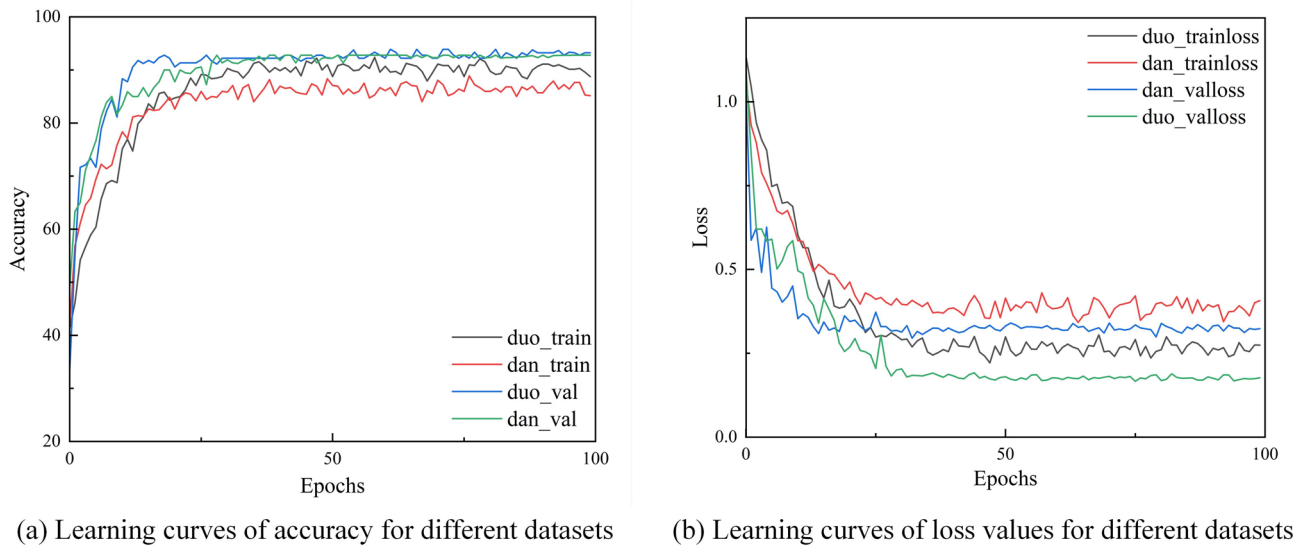


Fig. 10. The accuracy-loss curves for single-cycle and multi-cycle, (a) is the accuracy curve; (b) is the loss value curve.

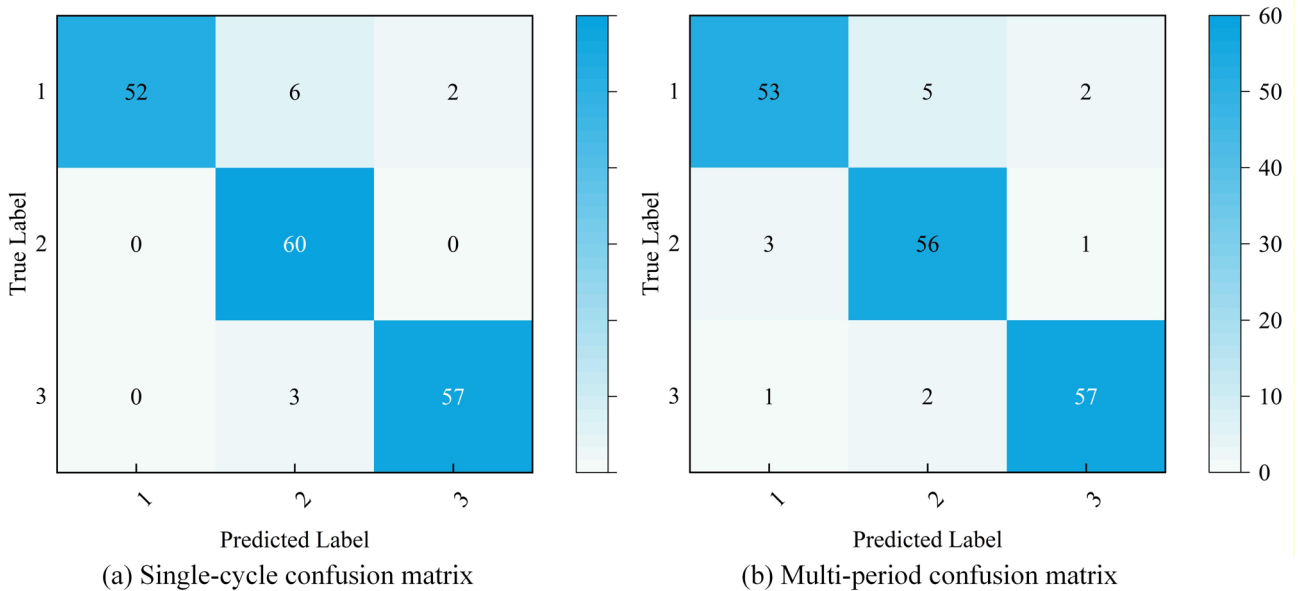


Fig. 11. Confusion matrices for different datasets: (a) Confusion matrix for the single-cycle dataset; (b) Confusion matrix of multi-period dataset.

Figure 12 presents the ROC curves for evaluating the Eff_SHSAIDC model on both single-cycle and multi-cycle datasets. These ROC curves demonstrate excellent classification performance. Specifically, the curve corresponding to the multi-cycle sequences is closer to the top-left corner of the coordinate plot, and its area under the curve (AUC) is significantly larger than that of the single-cycle sequences. This result clearly indicates that the model based on multi-cycle sequences exhibits superior classification performance, with the increase in AUC value directly reflecting enhanced generalization capability of the model. This improvement is primarily because multi-cycle sequences can encompass more comprehensive dynamic characteristics and long-range dependency information between heartbeats and pulse waves, providing the model with more discriminative learning material. Consequently, the model demonstrates higher accuracy and robustness in distinguishing between DPN patients and healthy individuals.

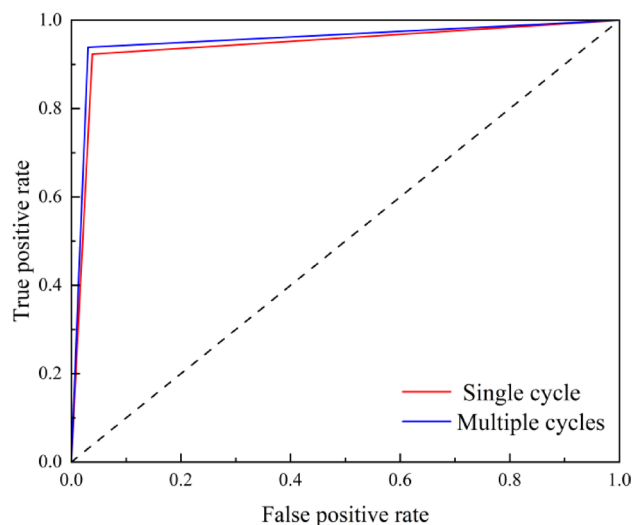


Fig. 12. Single-cycle and multi-cycle ROC curves.

Model structure	ACC (%)	SEN (%)	PRE (%)	F1_Score
Densenet121	76.67 ± 0.62	76.23 ± 0.59	76.07 ± 0.56	76.15 ± 0.58
Googlenet	72.59 ± 0.61	73.82 ± 0.63	71.85 ± 0.58	72.82 ± 0.59
ResNet_18	77.04 ± 0.59	77.18 ± 0.61	77.59 ± 0.63	78.38 ± 0.58
ViT-B	88.43 ± 0.73	89.10 ± 0.67	88.43 ± 0.82	88.76 ± 0.67
Hybrid ViT	91.11 ± 0.56	91.21 ± 0.58	91.11 ± 0.60	91.16 ± 0.57
SwinTransformer	95.12 ± 0.71	92.11 ± 0.68	93.72 ± 0.74	92.75 ± 0.64
EfficientnetV2	87.37 ± 0.58	87.22 ± 0.61	88.41 ± 0.64	87.56 ± 0.59
Eff_SHSAIDC	93.89 ± 0.63	93.21 ± 0.58	94.52 ± 0.67	93.87 ± 0.61

Table 5. Comparison of fivefold cross-validation performance of different classification models on multi-period datasets.

Performance analysis on different models

In this study, the hyperparameters were uniformly configured. Based on the same dataset and using multi-cycle sequence structures as input data, six common image classification models were employed for the experiments. The evaluation metrics for the three-class DPN classification are presented in Table 5.

The results in Table 5 show that GoogLeNet performed relatively poorly. For complex image classification tasks, its stacked convolutional layer architecture demonstrates average classification performance. The dense connections between its layers may lead to issues such as gradient vanishing. In contrast, DenseNet121, with its narrower network and relatively fewer parameters, and ResNet_18, with its residual structure, achieved slightly higher classification accuracy. Compared to the four common classification models, the Eff_SHSAIDC model achieved notably better results, with all four evaluation metrics exceeding 93%. This further enhances the model's generalization capability and improves the accuracy of medical image classification assessment.

Ablation experiment

To verify the necessity of the model network architecture proposed in this experiment, we conducted ablation experiments to evaluate the performance of the entire model, as shown in Table 6.

This study proposes an improved Eff_SHSAIDC network model. It utilizes short-term segments of physiological signals as samples and converts them into three-dimensional fused images using the SGR spatial encoding algorithm. By testing various models on the self-collected dataset and comparing the improved Eff_SHSAIDC model with its base model and ablated models, the experimental results demonstrate that the improved model exhibits superior performance in DPN classification, achieving a classification accuracy of 93.89%. The combination of its enhancements broadens the model's capacity to mine data features, strengthens its recognition and understanding of complex patterns, and significantly optimizes classification performance.

Discussion

In this study, a deep learning framework based on PPG and ECG fusion is proposed for DPN screening. With an accuracy of 93.89%, a sensitivity of 93.21%, and a specificity of 94.52%, the framework achieves superior performance compared with conventional methods. By fusing non-invasive PPG and ECG signals, it

Model architecture						
IDC	SHSA	Model	ACC (%)	SEN (%)	PRE (%)	F1_Score (%)
		EfficientnetV2	87.37	87.22	88.41	87.56
√		Eff_IDC	86.11	87.48	86.11	85.84
	√	Eff_SHSA	88.43	89.10	88.43	88.76
√	√	Eff_SHSAIDC	93.89	93.21	94.52	93.87

Table 6. Ablation experiments on the components of the Eff_SHSAIDC network.

comprehensively captures peripheral perfusion and cardiac electrical information, which is more conducive to clinical translation. In particular, the model delivers a high sensitivity of 94.12%, enabling effective identification of early DPN lesions with low false-negative rates. This non-invasive, high-precision framework satisfies clinical requirements for DPN screening and provides a promising auxiliary diagnostic tool for DPN.

DPN pathological progression simultaneously impairs peripheral vasodilation and cardiac electrical rhythm, leading to abnormalities in the hemodynamic characteristics of PPG signals and the rhythmic parameters of ECG signals. Synergistic analysis of these two modalities can overcome the information limitations of single-signal analysis, providing a more comprehensive physiological basis for DPN diagnosis⁴⁰. However, physiological signals are easily disturbed by motion artifacts and environmental noise, and existing fixed-parameter denoising methods cannot satisfy the demands of DPN-related signals. This study employs PSO-ICEEMDAN combined with wavelet threshold denoising, which dynamically optimizes parameters through particle swarm optimization to reduce noise-induced distortion and ensure high-quality signal input for feature extraction. On this basis, the SGR algorithm transforms time-series signals into structured feature maps, effectively capturing the spatiotemporal correlation patterns across modalities.

While multimodal fusion has shown promise for DPN diagnosis, existing methods rely mostly on shallow concatenation or static weighting, which cannot fully capture dynamic physiological correlations⁴¹. Inadequate preprocessing further limits feature discrimination, and traditional deep learning models suffer from high computational complexity, hindering clinical rapid screening⁴². To address these issues, this study establishes a complete technical pipeline from signal preprocessing to feature fusion: using PSO-ICEEMDAN combined with wavelet threshold denoising for signal quality improvement, the SGR algorithm for structured correlation transformation, and the lightweight Eff_SHSAIDC network for efficient feature extraction. Experimental results validate that this system improves deep feature fusion while supporting convenient clinical application.

Building upon the strengths of existing research, this study constructs a complete technical pipeline spanning from signal preprocessing to feature fusion, which is specifically designed to meet the practical requirements of clinical DPN screening workflows. The combination of PSO-ICEEMDAN and wavelet threshold denoising ensures robust signal quality under diverse acquisition conditions, reducing manual signal screening and re-acquisition in real clinical scenarios. The SGR algorithm enables structured conversion of signal correlation information into image representations, supporting intuitive interpretation by clinicians and smooth integration with current image-based diagnostic pipelines. Furthermore, the lightweight Eff_SHSAIDC network achieves efficient feature extraction with low computational cost, allowing real-time inference on portable point-of-care devices. Within existing DPN screening protocols, this integrated framework can act as an auxiliary screening tool in primary care settings, helping non-specialist nurses and general practitioners identify high-risk individuals for timely specialist referral, thus improving resource allocation and facilitating early intervention. The experimental results verify the effectiveness of the proposed system, showing its potential to connect advanced deep learning methods with practical clinical deployment.

Despite the certain progress achieved, this study still has some limitations. All the samples were collected from a single clinical center, with a relatively concentrated distribution of patients' ages and disease courses; the lack of coverage of heterogeneous populations and diverse clinical acquisition conditions may affect the generalization performance of the model in multi-center and real-world clinical scenarios. In the process of framework construction, clinical pathological indicators were not fully integrated, which to a certain extent affected the clinical relevance and interpretability of the model output results. Additionally, the practical deployment of the proposed framework still faces potential challenges. For one thing, the model's performance in extreme signal quality scenarios needs further improvement, a common technical challenge in PPG and ECG signal fusion analysis. For another, adapting the framework to various portable acquisition devices and optimizing its real-time inference efficiency require further exploration to enable clinical bedside application.

Future research will focus on the clinical translation and performance optimization of the proposed model. Multi-center and large-sample clinical data will be collected to further verify the model's generalization ability across heterogeneous populations with different ages, disease stages and comorbidities. Meanwhile, clinical pathological and imaging data will be integrated to improve model interpretability via feature visualization and mechanism analysis. Adaptive signal completion and enhancement algorithms will be developed to enhance robustness against low-quality signals. In addition, the network will be optimized and lightweighted for wearable devices, promoting DPN screening from single-point detection toward full-cycle health management.

Conclusion

This study has successfully constructed a deep learning framework based on the fusion of PPG and ECG signals for the non-invasive screening of DPN. By converting time-series signals into images through the innovative

SGR algorithm and combining them with a lightweight and efficient network specifically designed for DPN screening, the model achieved a classification accuracy of 93.89% on the multi-cycle dataset, with performance significantly superior to that of the baseline model. This fully confirms the technical feasibility of realizing non-invasive DPN detection using multimodal physiological signals. This study provides a complete technical solution with a solid theoretical foundation and clinical translation potential for the development of low-cost, convenient early screening tools suitable for community and home scenarios.

Data availability

The data that support the findings of this study are available upon reasonable request from the corresponding author.

Received: 10 November 2025; Accepted: 23 March 2026

Published online: 27 March 2026

References

- Selvarajah, D. et al. Diabetic peripheral neuropathy: Advances in diagnosis and strategies for screening and early intervention. *Lancet Diabetes Endocrinol.* **7**(12), 938–948 (2019).
- Rehman, R. Z. U. et al. Assessment of physiological signals from photoplethysmography sensors compared to an electrocardiogram sensor: A validation study in daily life. *Sensors* **24**(21), 6826 (2024).
- Kærgaard, K., Jensen, S. H. & Puthusserypady, S. A comprehensive performance analysis of EEMD-BLMS and DWT-NN hybrid algorithms for ECG denoising. *Biomed. Signal Process. Control* **25**, 178–187 (2016).
- Li, S., Li, J., Mao, J. D., Hongm, W. & Sun, A. A denoising method for ECG signals based on CEEMDAN-TSO and stacked sparse autoencoders. *Comput. Biol. Med.* **25**, 178–187 (2024).
- Altuve, M., Suárez, L. & Ardila, J. Fundamental heart sounds analysis using improved complete ensemble EMD with adaptive noise. *Biocybern. Biomed. Eng.* **40**(1), 426–439 (2020).
- Camara, C., Peris-Lopez, P., Safkhani, M. & Bagheri, N. ECG identification based on the Gramian angular field and tested with individuals in resting and activity states. *Sensors* **23**(2), 937 (2023).
- Shankar, A., Khaing, H. K. & Dandapat, S. Analysis of epileptic seizures based on EEG using recurrence plot images and deep learning. *Biomed. Signal Process. Control* **69**, 102854 (2021).
- Yang, X., Yang, X., Zhang, C. & Wang, J. SAR image classification using markov random fields with deep learning. *Remote Sens.* **15**(3), 617 (2023).
- Rahman, M., Islam, D., Mukti, R. J. & Saha, I. A deep learning approach based on convolutional LSTM for detecting diabetes. *Comput. Biol. Chem.* **88**, 107329 (2020).
- García-Ordás, M. T., Benavides, C., Benítez-Andrades, J. A., Alaiz-Moretón, H. & García-Rodríguez, I. Diabetes detection using deep learning techniques with oversampling and feature augmentation. *Comput. Methods Programs Biomed.* **202**, 105968 (2021).
- Tanim, S. A. et al. Explainable deep learning for diabetes diagnosis with DeepNetX2. *Biomed. Signal Process. Control* **99**, 106902 (2025).
- Gupta, H., Varshney, H., Sharma, T. K., Pachauri, N. & Verma, O. P. Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. *Complex Intell. Syst.* **8**, 3073–3087 (2022).
- Jiang, A. J. et al. Heart rate-corrected QT interval: A novel diagnostic biomarker for diabetic peripheral neuropathy. *J. Diabet. Investig.* **13**(5), 850–885 (2022).
- Wei, H. C. et al. Prognosis of diabetic peripheral neuropathy via decomposed digital volume pulse from the fingertip. *Entropy* **22**(7), 754 (2020).
- Keikhosravi, A., Aghajani, H. & Zahedi, E. Discrimination of bilateral finger photoplethysmogram responses to reactive hyperemia in diabetic and healthy subjects using a differential vascular model framework. *Physiol. Meas.* **34**(5), 513–525 (2013).
- Chowdhury, A., Chowdhury, M. H., Das, D., et al. FPGA implementation of PPG-based cardiovascular diseases and diabetes classification algorithm. *Arabian J. Sci. Eng.*, 1–13 (2024).
- Kumar, D. S. R., Maram, B. & Kshirsagar, P. R. Deep belief parallel forward harmonic network for cardiovascular disease detection using ECG images. *Eur. Phys. J. Plus* **140**, 1113 (2025).
- Şentürk, Ü., Yücedağ, I., Polat, K. Repetitive neural network (RNN) based blood pressure estimation using PPG and ECG signals. In *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 1–4. (2018).
- Ansari, Y., Omar, M., Qaraqe, K. & Serpedin, E. Deep learning for ECG arrhythmia detection and classification: An overview of progress for period 2017–2023. *Front. Physiol.* <https://doi.org/10.3389/fphys.2023.1246746> (2023).
- Santaji, S. & Desai, V. Analysis of EEG signal to classify sleep stages using machine learning. *Sleep Vigil.* **4**, 145–152 (2020).
- Wang, J. et al. A review of deep learning on medical image analysis. *Mobile Netw. Appl.* **26**, 351–380 (2021).
- Nour, M., Öztürk, Ş & Polat, K. A novel classification framework using multiple bandwidth method with optimized CNN for brain-computer interfaces with EEG-fNIRS signals. *Neural Comput. Applic* **33**, 15815–15829 (2021).
- Mahajan, P. & Kaul, A. Enhanced cuffless blood pressure estimation using ECG and PPG signals: A hybrid approach with Windkessel, ARIMA, and LSTM. *Turk. J. Electr. Eng. Comput. Sci.* **33**(3), 282–305 (2025).
- Eren, E., Yildirim, O. F. & Özdemir, S. Unveiling anomalies: A survey on XAI-based anomaly detection for IoT. *Turk. J. Electr. Eng. Comput. Sci.* **32**(3), 358–381 (2024).
- Wu, Z. H. & Huang, N. E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Data Sci. Adapt. Anal.* **1**, 1–41 (2009).
- Huang, N. E. et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Royal Soc. London Series A: Math., Phys. Eng. Sci.* **454**(1971), 903–995. <https://doi.org/10.1098/rspa.1998.0193> (1998).
- Colominas, M. A., Schlotthauer, G. & Torres, M. E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Process. Control* **14**, 19–29 (2014).
- Wang, G., Chen, X. Y., Qiao, F. L., Wu, Z. H. & Huang, N. E. On intrinsic mode function. *Adv. Adapt. Data Anal.* **2**(3), 277–293 (2010).
- Ramirez-Ochoa, D.-D., Pérez-Domínguez, L. A., Martínez-Gómez, E.-A. & Luviano-Cruz, D. PSO, a swarm intelligence-based evolutionary algorithm as a decision-making strategy: A review. *Symmetry* **14**(3), 455 (2022).
- Chiu, I. M. et al. Utilization of personalized machine-learning to screen for dysglycemia from ambulatory ECG, toward noninvasive blood glucose monitoring. *Biosensors (Basel)* **13**(1), 23 (2022).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkorei, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010. (2017).
- Li, Z., Wang, T., Song, X., He R, He, F. Coronary artery disease detection based on photoplethysmography via gramian angular field transformation and deep learning. *J. Intell. Med.*, 1–10. (2025).

33. Ouyang, C. et al. Inter-patient classification with encoded peripheral pulse series and multi-task fusion CNN: Application in type 2 diabetes. *IEEE J. Biomed. Health Inform.* **25**(8), 3130–3140 (2021).
34. Tan, M., Le, Q. EfficientNetV2: Smaller models and faster training. In *Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research*, 139: 10096–10106 (2021).
35. Kushner, H. J. A new method of locating the maximum point of an arbitrary multiplex curve in the presence of noise. *J. Basic Eng.* **86**(1), 97–106 (1964).
36. Zhou, P., Xie, X., Lin, Z. & Yan, S. Towards understanding convergence and generalization of AdamW. *IEEE Trans. Pattern Anal. Mach. Intell.* **46**(9), 6486–6493 (2024).
37. Wang, C. L., Wang, J. H., Xie, J. H. Cosine annealing weights in knowledge distillation. In *Proceedings of the 2025 6th International Conference on Computer Information and Big Data Applications. New York: Association for Computing Machinery*, 222–228 (2025).
38. Rohitha, K., Supriya, K., Kuchoor, S. K. & Bodapati, J. D. Optimizing deep convolutional neural networks with progressive unfreezing for enhanced sports activity recognition. *Int. Conf. Integr. Intell. Commun. Syst. (ICIICS)* **2024**, 1–6 (2024).
39. Rastogi, D., Johri, P., Tiwari, V. & Elngar, A. A. Multi-class classification of brain tumour magnetic resonance images using multi-branch network with inception block and five-fold cross validation deep learning framework. *Biomed. Signal Process. Control* **88**(A), 105602 (2024).
40. Pau, M. et al. Cyclograms reveal alteration of inter-joint coordination during gait in people with multiple sclerosis minimally disabled. *Biomechanics* **2**(3), 331–341 (2022).
41. Wei, H. C. et al. Percussion entropy analysis of synchronized ECG and PPG signals as a prognostic indicator for future peripheral neuropathy in type 2 diabetic subjects. *Diagnostics* **10**(1), 32 (2020).
42. Dibbern, K. N. et al. Scoping review of machine learning techniques in marker-based clinical gait analysis. *Bioengineering* **12**(12), 591 (2025).

Acknowledgements

Data processing was supported by the Ningxia Technology Innovative Team of Advanced Intelligent Perception and Control and the Key Laboratory of Intelligent Perception Control at North Minzu University.

Author contributions

X.J.T data acquisition; F.W, S.D.F, G.J.D Data analysis; F.W, M.X.X proposed the methodology; All authors contributed to the analysis and manuscript preparation; M.X.X supervised the entire research.

Funding

This research was supported by Ningxia National Science Foundation of China (2024AAC03153) the Youth Nurturing Program of North Minzu University (2023QNPY27), and the Graduate Student Innovation Project of NUM (CYX25198).

Declarations

Competing interests

The authors declare that there is no potential conflict of interest.

Ethical approval

This study was approved by the Biomedical Research Ethics Committee of Northern University for Nationalities (Approval number: 2024-2). Patients provided written informed consent to participate in this study.

Additional information

Correspondence and requests for materials should be addressed to M.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026