Am J Transl Res 2025;17(5):3951-3960
www.ajtr.org /ISSN:1943-8141/AJTR0162986

Original Article

The multikinetic fusion feature of PPG was combined with MCNN_vision_transformer for diabetes detection

Ming-Xia Xiao, An-Yao Zhang, Ting-Ting Jin, Shi-Dong Fang

School of Electrical and Information Engineering, North Minzu University, No. 204 North Wenchang Street, Yinchuan 750021, Ningxia, China

Received December 30, 2024; Accepted May 11, 2025; Epub May 15, 2025; Published May 30, 2025

Abstract: Background: Diabetes is a chronic condition that significantly impacts the cardiovascular system and various other organs. Photoplethysmogram (PPG) signals have been shown to correlate with variations in vascular blood flow and the presence of atherosclerosis. To effectively explore the complex nonlinear relationship between PPG signals and diabetes, we propose an automatic detection model based on the fusion of PPG features. Methods: The proposed model consists of two main components: 1. Dynamic Fusion Feature Extraction: Short PPG signal window segments are processed using the SGR spatial encoding algorithm to extract dynamic fusion features. 2. Feature Representation Learning: Multi-scale convolutional layers (MCNN) are employed to learn feature representations, while the Vision Transformer (ViT) model is utilized to capture global contextual semantic features. Results: The model was trained and validated on a self-collected medical dataset. The experimental results demonstrate that the classification model, which integrates short time window information, significantly improves detection performance. Specifically, the multi-period sequence input model achieves an accuracy of 91.11%, with a Receiver Operating Characteristic (ROC) curve area of 0.9341, indicating strong diagnostic capability. Conclusion: This study is a retrospective case-control study that collected clinical data from three groups of people: those with normal glucose levels, those with poorly controlled diabetes, and those with well-controlled diabetes. The study aims to utilize deep learning algorithms for the early prevention and screening of diabetes.

Keyword: PPG, diabetes, deep learning, MCNN, transformer

Introduction

Diabetes mellitus is a chronic metabolic disorder characterized by abnormal glucose accumulation in the bloodstream. It can lead to severe cardiovascular damage and is associated with a wide range of complications [1]. Regular daily monitoring plays a crucial role in the prevention and early screening of diabetes. Although traditional invasive and minimally invasive testing methods provide reliable and accurate clinical evaluation and diagnosis, they are associated with patient discomfort and a risk of bloodborne infections, and they are limited by relatively high costs and the need for specialized personnel.

Photoplethysmogram (PPG) signal has been successfully integrated into commercial wearable devices and is widely used for heart rate and blood oxygen monitoring due to its low power consumption and high measurement efficiency [2]. The degree of hyperglycemia in diabetic patients can significantly influence blood viscosity and flow velocity [3]. For PPG signals, the characteristic waveform parameters encapsulate valuable physiological information, among which pulse wave velocity (PWV) is widely used to reflect the propagation time of the pressure wave [4]. In patients with diabetes, the width of the pulse wave at one-third of its peak height increases, while the amplitude of the dicrotic wave decreases [5].

Diabetic patients and healthy controls were classified using a Bayesian classifier based on bilateral finger PPG measurements in response to reactive hyperemia [6]. Based on Heart Rate Variability (HRV) analysis, two logistic regression models have been trained to predict more than five levels of diabetes risk [7]. Recently, a noninvasive diabetes detection system was

https://doi.org/10.62347/ZRMW1346

constructed using the mixed characteristics of 5SPPG signal segments [8]. A new index called the dynamic system vascular resistance index (DSVRI) is proposed; a pathologic feature associated with systemic vascular resistance [9]. These features indicate a strong correlation with diabetes. In recent years, Convolutional Neural Network (CNN) has shown the advantage of high efficiency and automatic image processing. Some scholars extract feature-related information to encode PPG signals to distinguish patients with type 2 diabetes [10]. Using deep neural network and PPG signals, it achieved a good success rate of 90.25%, and highlights the potential for future commercialization [11]. Additionally, a study demonstrated for the first time that smartphone-based PPG can be used for diabetes detection. This 18-layer CNN diabetes detection model, based on the original PPG signal, achieved a specificity of 65.4%, sensitivity of 75%, and average area under the curve (AUC) of 0.77 [12]. Further, some research focused on the two-dimensional image transformation form of PPG and diagnosed diabetes with the improved Visual Geometry Group Network (VGG-Net) model, which achieved an accuracy of 76.34% [13]. For the image coding of PPG signal and the use of CNN model with Multi-task fusion, the experimental results show that the best accuracy of Recursive map (RP) with the threshold ε 6000 reaches 90.6% [14]. Lu et al. used CNN to extract single-cycle and multi-cycle spatial characteristics and used Long Short-Term Memory (LSTM) to extract long-term related features for cardiovascular disease classification, achieving 80% accuracy [15]. Other researchers collected PPG signals from different subjects through self-made photoelectric sensors and used an artificial neural network integrated into the field programmable array (FPGA) to predict a blood glucose model, providing a new method for the auxiliary monitoring of diabetes [16].

In the aforementioned study, extensive data cleaning and preprocessing for denoising inevitably introduced significant bias. Simple deep learning models are often insufficient in comprehensively extracting the features of the signals. Therefore, we further explore the characteristics of blood flow fluctuations by spatially encoding short-term PPG signals. Using the multi-scale convolutional layers (MCNN) combined with the Vision Transformer (ViT) in the Multi-scale Convolutional Neural Network Vision Transformer (MCNN-ViT) architecture we delve deeper into the cardiovascular dynamics embedded in the signal, indirectly mapping the complex nonlinear relationship between PPG and diabetes pathology. This approach enables screening of the diabetic population, offering a novel strategy for early prevention and non-invasive continuous diagnosis of diabetes.

Materials and methods

Subjects

This study is a retrospective case-control analysis that gathers clinical data from diabetic patients at a hospital. The datasets were obtained from the affiliated hospital of Ningxia Medical University, with research focused on the elderly population in Ningxia. The gender distribution within each group was approximately balanced, with a ratio near 1:1. Furthermore, the data collection process adhered to stringent protocols, ensuring high scientific rigor. **Table 1** shows significant differences between the three groups for the Hemoglobin A1c (HbA1c), Blood Sugar AC, and Cholesterol parameters, which provides data support for the subsequent population grouping. The study was conducted according to the Declaration of Helsinki and approved by the biomedical research ethics committee of North Minzu University (No.2024-2).

Signal pretreatment

Ensemble Empirical Mode Decomposition - Hilbert-Huang Transform (EEMD-HHT) denoising algorithm for PPG signal: During PPG acquisition, interference from machine power frequency and baseline shifts caused by human respiration can occur [17]. The EEMD-HHT method is employed to denoise the signal, demonstrating superior performance for nonlinear and non-stationary signals compared to traditional Fourier transform and band-pass filters [18]. The experiment begins by applying a second-order band-pass filter to the collected signal for initial noise reduction, minimizing interference from noise and simple artificial distortion. A 6-second sliding window is then used to segment the original signal cycle. Due to the non-linear and non-smooth characteristics of the PPG signal, the EEMD-HHT algorithm is chosen to eliminate the effects of power frequency interference, low-frequency breathing noise, and baseline drift. Finally, the signal is normalized to the range [0, 1] to complete the preprocessing (**Figure 1**).

Table 1. Basic human physiological parameters of the participants in the three groups

Significant Parameters	Group 1 (n = 32)	Group 2 (n = 22)	Group 3 (n = 48)
Male/Female	16/16	11/11	25/23
Age (years)	56.38±7.42	65.18±10.55	62.56±11.40
BMI (kg/m^2)	25.07±3.43	26.77±2.99	26.98±6.14
HbA1c (%)	5.88±0.35	6.32±0.32**	8.47±1.56**
Blood Sugar AC (mg/dL)	97.63±10.09	120.91±25.93**	163.3±54.12**
Cholesterol (mg/dL)	211.13±34.14	170±36.42**	180.5±37.48

Note: Group 1: Healthy subjects; Group 2: Subjects with better diabetes control; Group 3: Subjects with poor diabetes control; n: number of people. Where * indicates P<0.05, statistical difference between groups, ** indicates P<0.001, significant statistical difference between groups. HbA1c: Hemoglobin A1c.

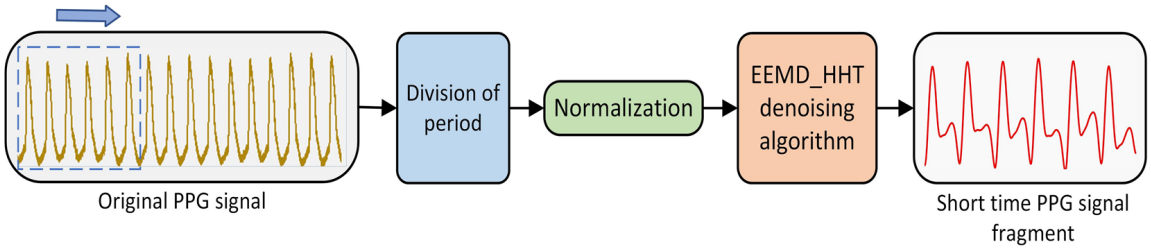

Figure 1. Block diagram of PPG signal pretreatment. Note: PPG: Photoplethysmogram, EEMD-HHT: Ensemble Empirical Mode Decomposition-Hilbert-Huang Transform.

Multi-kinetic fusion features: For PPG signals, the time-domain characteristics are highly similar. To extract relevant feature information and distinguish subtle differences between populations, we utilize the dynamic characteristics of the PPG signal to indirectly map the complex nonlinear relationship between PPG signals and diabetic pathology.

Spatial position encoding (SPE): Combining multiple spatial coding matrices enables a more comprehensive exploration of the effects of blood glucose levels. The use of SPE allows for the integration of spatial positioning and feature information, which enhances the ability of subsequent models to extract and distinguish key features. For PPG signals, the temporal information at different spatial locations reflects distinct characteristic fluctuation patterns and blood flow dynamics, which facilitates the exploration of both local and global temporal dependencies, thereby enhancing the overall feature representation and utility [16].

$$SPE_{ij} = \left\| \vec{x}_i - \vec{x}_j \right\| = \sqrt{(\vec{x}_i - \vec{x}_j)^T (\vec{x}_i - \vec{x}_j)} \quad (1)$$
$$i,j \in [0,m]\ SPE \in R^{m \times m}$$

Graham angle field (GAF): By calculating the differences and relative angles between time points, each time point is mapped to an angular and radial value in polar coordinates, which reveals the dynamic evolution patterns across time points and helps to explore the characteristic blood flow fluctuations within blood vessels encoded in the PPG signal.

$$\phi_i = \arccos(\vec{x}_i), r_i = \frac{i}{m}, i \in [0,m] \quad (2)$$

Where Φ_i is the angle vector and r_i is the radius. Two forms of GAF can be derived by calculating either the sum or the difference of angles between different time points:

$$GASF_{ij} = [\cos(\phi_i + \phi_j)] = \begin{bmatrix} \cos(\phi_1 + \phi_1) \ldots \cos(\phi_1 + \phi_j) \\ \cos(\phi_2 + \phi_1) \ldots \cos(\phi_2 + \phi_j) \\ \ldots \quad \ldots \quad \ldots \\ \cos(\phi_i + \phi_1) \ldots \cos(\phi_i + \phi_j) \end{bmatrix}$$
$$GADF_{ij} = [\sin(\phi_i - \phi_j)] = \begin{bmatrix} \sin(\phi_1 - \phi_1) \ldots \sin(\phi_1 - \phi_j) \\ \sin(\phi_2 - \phi_1) \ldots \sin(\phi_2 - \phi_j) \\ \ldots \quad \ldots \quad \ldots \\ \sin(\phi_i - \phi_1) \ldots \sin(\phi_i - \phi_j) \end{bmatrix} \quad (3)$$

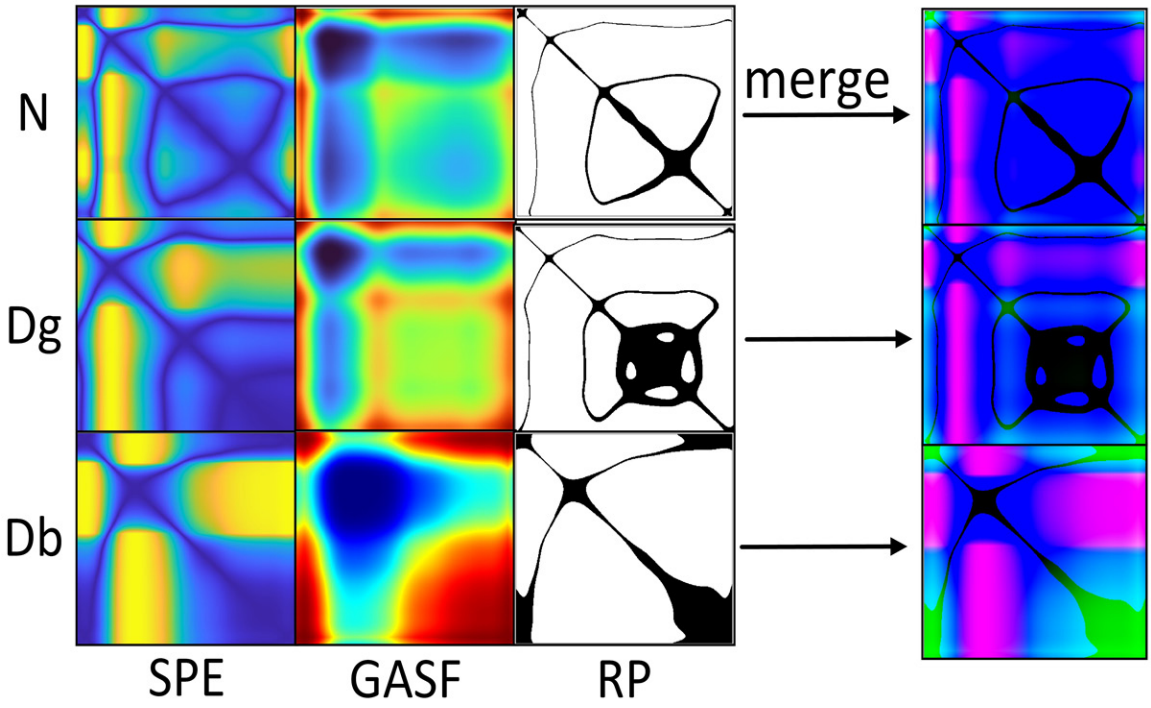

Figure 2. N: normal group, Dg: diabetes group with good glycemic control, Db: diabetes group with good glycemic control. The SPE, GASF, and RP (SGR) fusion algorithm: The spatial representations of these segments are encoded using three complementary techniques: Symbolic Permutation Entropy (SPE), Gramian Angular Summation Field (GASF), and Recurrence Plot (RP). These three encoded modalities are subsequently integrated into a unified three-channel fusion image to enhance feature representation for downstream analysis.

Recursive map (RP): RP analyzes the intrinsic temporal structure of the signal by capturing its periodicity, chaotic complexity, and nonlinear non-stationarity. By extracting and representing the underlying hemodynamic characteristics of the PPG signal, RP enhances both the resolution and versatility of the model, thereby improving its capability to perceive and interpret the potential physiological information embedded within the signal [19].

$$RP_{ij} = \Phi(\lambda - \| \vec{x}_i - \vec{x}_j \|), i,j \in [0,m] \quad (4)$$

$$\Phi(\bullet) = \begin{cases} 1, (\lambda - \| \vec{x}_i - \vec{x}_j \|) \geq 0 \\ 0, (\lambda - \| \vec{x}_i - \vec{x}_j \|) \leq 0 \end{cases} \quad (5)$$

The threshold λ take 0.1 (normalized peak is 1, taking 10% of the peak), $\Phi(\bullet)$ as a step function.

The SPE, GASF, and RP (SGR) fusion algorithm is utilized to combine the generated multi-modal two-dimensional images into a three-channel image, similar to an 'RGB' format. This fusion technique enables complementary information exchange across modalities, improves the acquisition and preservation of comprehensive feature information from multiple dimensions, and effectively captures the subtle fluctuation characteristics embedded in the signal (**Figure 2**).

The MCNN-ViT classification model

Due to the complexity and diversity of multi-dynamic image datasets, a robust model capable of capturing both intricate local features and global contextual relationships in medical images is required. This study integrates multi-scale convolution with ViT to develop a novel hybrid deep learning model, which aims to achieve effective local feature extraction while simultaneously modeling long-range global dependencies.

The constructed model MCNN-ViT consists of three important parts: MCNN, dual patch partition module, and ViT. The specific description is as follows (**Figure 3**).

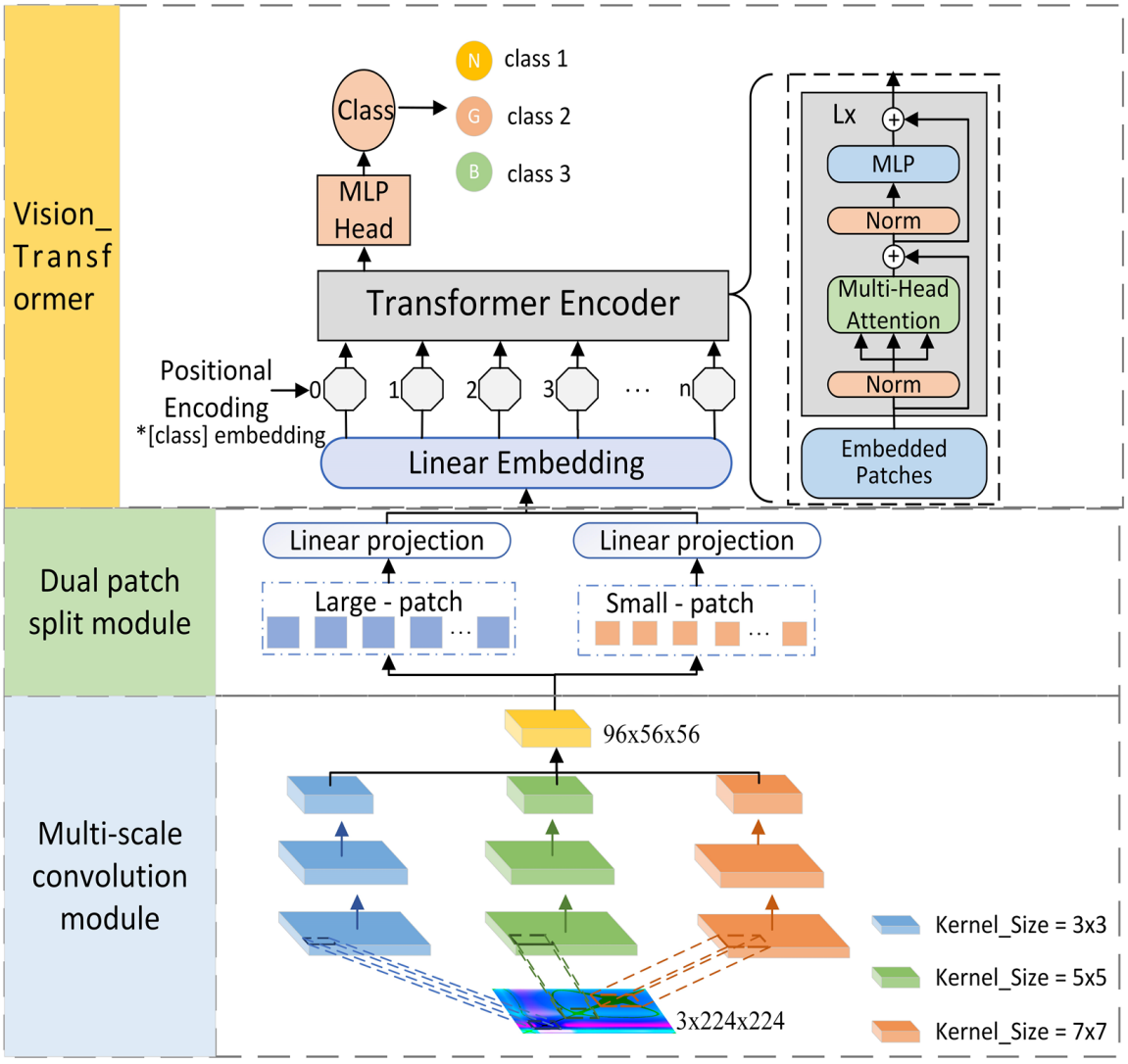

Figure 3. The Multi-Scale Convolutional Neural Network-Vision Transformer (MCNN-ViT). Model architecture. Simple process: The image after fusion of multiple spatial encoding methods is passed through three different convolutional kernels to generate different feature maps. These feature maps are then fused for linear projection. By extracting two types of global information and utilizing the ViT encoder, multiple feature extractions are performed, ultimately enabling the screening of diabetic populations.

The CNN can mimic the human visual system and effectively recognize patterns and structures in scenes [20]. The diagonally symmetric space encodes critical features present in all matrices derived from the time series of pulse waveforms, including the SPE, GASF, and RP representations. Observations indicate that the resulting images exhibit three distinct regions along the diagonal. However, conventional single-scale networks are typically designed to capture features at specific scales, limiting their ability to effectively capture dynamic changes within the data and the cross-scale dependencies embedded in these features [21]. Therefore, three convolutional kernels were applied to convolve the upper-left, center, and lower-right regions of the image. The three-channel image was subsequently transformed into 8-channel, 16-channel, and 32-channel feature maps through three convolutional layers. The step size for each kernel was set to 1, 2, and 3, respectively. Max pooling, batch normalization, and the ReLU activation function were employed, and the features extracted from the three scales were fused to obtain the preliminary feature representations.

Table 2. Comparison of MCNN-ViT model indexes based on two inputs

Datasets	Input sequence	ACC (%)	SEN (%)	PRE (%)	F1-Score (%)	AUC
Validation	One	90.56	90.62	90.55	90.29	0.9296
	Three	93.89	93.90	93.89	93.89	0.9543
Test	One	88.89	88.89	88.89	88.89	0.9167
	Three	91.11	91.21	91.11	91.16	0.9341

Note: MCNN-ViT: Multi-Scale Convolutional Neural Network-Vision Transformer. Four evaluation metrics: Accuracy (ACC), Sensitive (SEN), Precision (PRE), Area Under the Curve (AUC).

To further capture the potential relationships between features, the fused features were divided into 8×8 lag patches and 4×4 small patches as input tokens. These tokens were used to capture the two global aspects of the fused features, thereby enhancing the model's ability to perceive global features. Subsequently, linear fusion was applied to embed the features into the ViT. The fused feature information follows a multi-scale and multi-level architecture, reinforcing the correlations between features. The experiment employs the ViT model to analyze the relationships between regions of the image using the multi-head attention mechanism, enabling the model to understand a broader context beyond the local features. It uses the multi-head attention mechanism to obtain different input projections, to deal with different concerns, and get multiple groups of attention results, and then uses the results for splicing and linear projection to get the final output, experimental design long attention mechanism head for 4. The embedded dimension size is 128, and finally, through the forward transmission network, it gets classification output.

Training parameter dataset

The training and testing datasets were split into an 8:2 ratio, with 1/4 of the training dataset aside for validation. Adam was chosen as the optimizer for the stochastic gradient descent algorithm, with key hyperparameters set as follows: learning rate = 0.001, betas = (0.9, 0.999), and epsilon = 1e-8. To prevent overfitting, a normalization layer and a dropout layer were applied. The cross-entropy loss function was used to train the model and enhance its accuracy. The experiment was conducted using python 3.10 and pyorch.

Experiments and results

Comparison of MCNN-ViT on two periodic datasets

In the following study, the experimental results of single-cycle and multi-cycle short time windows are explored. In pulse fluctuations, the oscillation of a single cycle may be influenced by the preceding cycle, resulting in a strong correlation with the subsequent adjacent cycle. In contrast, processing multi-cycle signal waveforms may produce more stable and complete results. To evaluate multi-cycle signal fragments, the current cycle segment is combined with the adjacent segments before and after it, forming a new fragment for further analysis and discussion: $y_n = Conbined\ (X_{n-1}, X_n, X_{n+1})$.

The experiment converts the input single-period and multi-period sequences into SGR dynamic images. The above result reflects that, when using sequence modeling with multiple cycles, which may contain more kinetic feature information, its model has higher generalization and better classification performance (**Table 2** and **Figure 4**). Single cycle signal fluctuation may be delayed to before and after the two cycles. The delay characteristic fluctuation information also plays a role in improving the classification performance of the model (**Figure 4A**, **4B**).

Comparing with other models on two datasets

Table 3 presents a comparison of four classification metrics for commonly used models based on two datasets, exploring the specific performance of each model. This further validates the advantage of the multi-period dataset and highlights the overall performance of the proposed model through the evaluation of the four metrics.

To control for variability between models, all the aforementioned models use CNN as the initial encoder, which is closely aligned with the structure of the baseline model. The results indicate that when the multi-period dataset is used as input, the four evaluation metrics for each model show improvement. An examination of

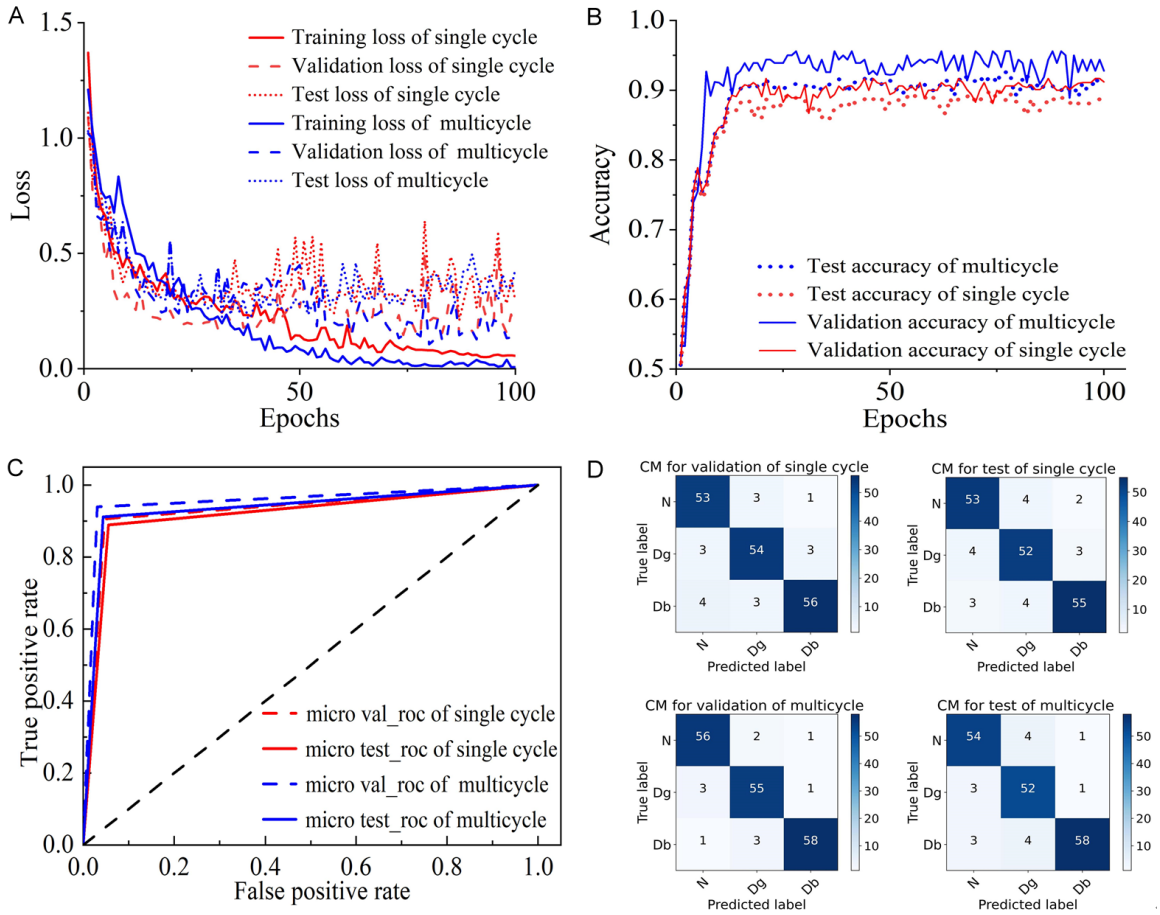

Figure 4. (A and B) Training, validation, and test loss values and accuracy curves for two inputs, (C and D) Receiver Operating Characteristic (ROC) curves and confusion matrices for validation and testing of two inputs.

the test datasets with two different periodic forms revealed that the overall performance of the Google Network (GoogLeNet), Densely Connected Convolutional Network-121 (DenseNet-121), and Visual Geometry Group 16 (VGG16) models were suboptimal, with accuracy ranging from 74.44% to 84.07%. In contrast, the Residual Network-18 (ResNet-18) model, with its residual structure, exhibited a slight improvement, achieving an accuracy of 84.81%. Compared to the other four commonly used classification models, MCNN-ViT demonstrated significant improvements across all evaluation metrics. On the multi-period dataset, all four model evaluation metrics reached values above 91%.

Model ablation experiments

Based on these results, further ablation studies were conducted on the multi-period test dataset of PPG signals to validate the importance of the multi-channel dynamic features and the proposed network architecture. With the successive fusion of input encoded images, the performance of the MCNN-ViT model showed improvement, with the accuracy gradually increasing. The first fusion resulted in a 2.22% improvement, while the second fusion achieved an increase of 3.89%. When the three dynamic features were fused as fixed inputs, it was observed that as the model network architecture continued to improve, the performance metrics of the ViT, CNN-ViT, and MCNN-ViT models progressively increased. Ultimately, all four evaluation metrics of the MCNN-ViT model surpassed 91% (**Table 4**).

Discussion

The results, as shown in **Table 2** and **Figure 4A**, **4B**, reveal that the introduction of the multi-

Table 3. Performance comparison of different models on two datasets

Input dataset	Single cycle				Multi-cycle			
Model index	ACC (%)	SEN (%)	PRE (%)	F1_Score	ACC (%)	SEN (%)	PRE (%)	F1_Score
Densenet121	74.44	80.30	74.44	77.26	84.07	84.23	84.07	84.15
Googlenet	75.93	81.91	75.93	78.81	81.86	83.82	81.85	82.82
ResNet-18	78.15	83.28	78.15	80.63	84.81	84.18	82.59	83.38
VGG16	77.04	82.03	77.04	79.46	83.33	83.13	83.33	83.23
MCNN-ViT	88.89	88.89	88.89	88.89	91.11	91.21	91.11	91.16

Note: The Densely Connected Convolutional Network-121 (Densenet121), Google Network (GoogLeNet), Visual Geometry Group 16 (VGG16), the Residual Network-18 (ResNet-18) are commonly used deep learning models. Ours: Multi-Scale Convolutional Neural Network-Vision Transformer (MCNN-ViT) model. Four evaluation metrics: Accuracy (ACC), Sensitive (SEN), Precision (PRE) Area Under the Curve (AUC).

Table 4. Ablation experiments for components with different inputs and MCNN-ViT

Dynamic feature fusion			Model	ACC (%)	SEN (%)	PRE (%)	F1_Score (%)
SPE	GASF	RP					
√			MCNN-ViT	85.00	85.24	85.00	85.12
√	√		MCNN-ViT	87.22	87.48	87.22	87.35
√	√	√	ViT	88.43	89.10	88.43	88.76
√	√	√	CNN-ViT	89.44	90.22	89.44	89.83
√	√	√	MCNN-ViT	91.11	91.21	91.11	91.16

Note: Three spatial encoding methods: Spatial position encoding (SPE), Gramian Angular Summation Field (GASF), Recursive map (RP). Models: Vision Transformer (ViT) model, Convolutional Neural Network - Vision Transformer model, Multi-Scale Convolutional Neural Network-Vision Transformer (MCNN-ViT) model. Four evaluation metrics: Accuracy (ACC), Sensitive (SEN), Precision (PRE), Area Under the Curve (AUC).

period dataset accelerates the model's convergence. On both the validation set and the test set, the iteration curves gradually become more stable. This indicates that the multi-period dataset helps reduce feature loss, further enhancing the capture of blood flow fluctuation information. In **Table 3**, the comparison of models across two datasets verifies that the proposed model does not exhibit any special dependency on the dataset. Googlenet121, Densenet121, and VGG16 model, when applied to complex images from different regions, suffer from excessive stacking of convolutional layers, which deepens network degradation and leads to the accumulation of excessive loss. Although ResNet_18 alleviates this phenomenon to some extent, its expressive power is limited, making it difficult to effectively capture high-level patterns and complex relationships within the images.

In contrast, the MCNN-ViT model demonstrates a unique advantage in spatially fused encoded images, combining the strengths of MCNN and ViT. It effectively captures both local details and the fluctuation features of global variables, thereby enhancing the model generalization ability. **Table 4** shows that the continuous fusion of three spatial encodings (SPE, GASF, RP) progressively captures the fluctuations and dynamic evolution patterns of the time-series signals, improving the model's resolution and generalization ability. The integration of dynamic characteristics provides a more comprehensive representation of human blood flow fluctuations, enhancing the model's generalization ability and classification accuracy. Compared to the standalone ViT, MCNN-ViT, which includes CNN as the initial encoder, extracts low-level features that deepen the model's understanding of the image and allow it to capture more local details. The two components complement each other, enabling the extraction of more diverse and richer features. The MCNN used in this study leverages three different convolution kernels to extract semantic information from relevant regions. The fused features comprehensively reflect the internal complexity of the image. Additionally, the dual-token mechanism for obtaining global information enables a better understanding of the semantic concepts of different image blocks. Combined with

the multi-head attention mechanism of ViT, this improves the model sensitivity to complex medical images, ultimately enhancing its overall performance.

Compared to previous work, the proposed MCNN-ViT model achieves superior performance in processing PPG signals and screening diabetic populations. Its novelty lies in three aspects. First, the SGR fusion image encoding method introduced in this study helps reduce significant errors in feature parameter selection and dependency on the time domain of PPG signals. Unlike single spatial encoding, the continuous fusion and increase in multi-space encoding information enable the representation of more comprehensive human blood flow dynamics. This is closely related to the arterial sclerosis phenomenon in human blood vessels and the onset characteristics of diabetes. Second, experimental results fully demonstrate that the multi-cycle form, compared to single-cycle data, captures missing information in periodic signals and more comprehensively obtains complete human blood flow cycle fluctuation characteristics. Finally, in contrast to previous approaches using simple networks with accumulated convolution layers, the proposed model utilizes multi-scale convolutions to initially extract high-level semantic features and latent information from different regions of the image. By employing two-scale tokens to capture two types of global information from the fused features, the model enhances its ability to perceive global features. The introduction of ViT further facilitates the complementary capture of both local details and global dependencies, improving the model's generalization ability and robustness.

The limitation of this study lies in the complexity and parameter count of the MCNN and ViT architecture, which may introduce slight biases when balancing the fusion of information from different scales. Additionally, the hospital dataset lacks diversity, leading to a certain degree of overfitting in the model. Future research could expand the dataset and incorporate lightweight modules to accelerate model convergence and improve prediction efficiency.

Conclusion

This study utilizes the spatial fluctuation characteristics of PPG signals to construct a diabetes screening model based on MCNN and ViT. The model demonstrates excellent performance in terms of accuracy and generalization ability, providing strong clinical support for the early screening and prevention of diabetes. It also shows significant potential for non-invasive real-time diagnostics in daily practice.

Acknowledgements

The data processing was supported by the Ningxia Advanced Intelligent Sensing and Control Technology Innovation Team and the Key Laboratory of Intelligent Sensing and Control of Beiminzu University. This research was supported by Ningxia National Science Foundation of China (2024AAC03153), the National Natural Science Foundation of China (No. 62361001), the Youth Nurturing Program of NMU (2023QNPY27), and the Graduate Student Innovation Project of North Minzu University (YCX23136).

Disclosure of conflict of interest

None.

Address correspondence to: Ming-Xia Xiao, School of Electrical and Information Engineering, North Minzu University, No. 204 North Wenchang Street, Yinchuan 750021, Ningxia, China. E-mail: xiao_mx@nmu.edu.cn

References

[1] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, Stein C, Basit A, Chan JCN, Mbanya JC, Pavkov ME, Ramachandaran A, Wild SH, James S, Herman WH, Zhang P, Bommer C, Kuo S, Boyko EJ and Magliano DJ. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022; 183: 109119.

[2] Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007; 28: R1-39.

[3] Monte-Moreno E. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. Artif Intell Med 2011; 53: 127-138.

[4] Woolam GI, Schnur PL, Vallbona C and Hoff HE. The pulse wave velocity as an early indicator of atherosclerosis in diabetic subjects. Circulation 1962; 25: 533-539.

[5] Hao Y, Cheng F, Pham M, Rein H, Patel D, Fang Y, Feng Y, Yan J, Song X, Yan H and Wang Y. A

noninvasive, economical, and instant-result method to diagnose and monitor type 2 diabetes using pulse wave: case-control study. JMIR Mhealth Uhealth 2019; 7: e11959.
[6] Keikhosravi A, Aghajani H and Zahedi E. Discrimination of bilateral finger photoplethysmogram responses to reactive hyperemia in diabetic and healthy subjects using a differential vascular model framework. Physiol Meas 2013; 34: 513-25.
[7] Chu J, Yang WT, Hsieh TH and Yang FL. One-minute finger pulsation measurement for diabetes rapid screening with 1.3% to 13% false-negative prediction rate. Biomed Stat Inform 2021; 6: 8.
[8] Prabha A, Yadav J, Rani A and Singh V. Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier. Comput Biol Med 2021; 136: 104664.
[9] Gupta S, Singh A, Sharma A and Tripathy RK. DSVRI: a PPG-based novel feature for early diagnosis of type-II diabetes mellitus. IEEE Sens Lett 2022; 6: 1-4.
[10] Ouyang C, Gan Z, Zhen J, Guan Y, Zhu X and Zhou P. Inter-patient classification with encoded peripheral pulse series and multi-task fusion cnn: application in type 2 diabetes. IEEE J Biomed Health Inform 2021; 25: 3130-3140.
[11] Deng H, Zhang L, Xie Y and Mo S. Research on estimation of blood glucose based on PPG and deep neural networks. IOP Conf Ser Earth Environ Sci 2021; 693: 012046.
[12] Avram R, Tison G, Kuhar P, Marcus G, Pletcher M, Olgin JE and Aschbacher K. Predicting diabetes from photoplethysmography using deep learning. J Am Coll Cardiol 2019; 73: 16.
[13] Srinivasan VB and Foroozan F. Deep learning based non-invasive diabetes predictor using Photoplethysmography signals. EUSIPCO 2021; 29: 1256-1260.
[14] Ouyang C, Gan Z, Zhen J, Guan Y, Zhu X and Zhou P. Inter-patient classification with encoded peripheral pulse series and multi-task fusion CNN: application in type 2 diabetes. IEEE J Biomed Health Inform 2021; 25: 3130-3140.
[15] Lu P, Liu C, Mao X, Zhao Y, Wang H, Zhang H and Guo L. Few-shot pulse wave contour classification based on multi-scale feature extraction. Sci Rep 2021; 11: 3762.
[16] Vaswani A. Attention is all you need. Adv Neural Inf Process Syst 2017; 30: I.
[17] Kumar A, Tomar H, Mehla VK, Komaragiri R and Kumar M. Stationary wavelet transform based ECG signal denoising method. ISA Trans 2021; 114: 251-262.
[18] Arslan, Özkan and Mustafa Karhan. Effect of Hilbert-Huang transform on classification of PCG signals using machine learning. J King Saud Univ Comput Inf Sci 2022; 34: 9915-9925.
[19] Hubel DH and Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. J Physiol 1959; 148: 574-591.
[20] Wang Z and Oates T. Imaging time-series to improve classification and imputation. ArXiv [Preprint] 2015; 1506: 00327.
[21] Guo J, Han K, Wu H, Tang Y, Chen X, Wang Y and Xu C. Cmt: convolutional neural networks meet vision transformers. Proc IEEE/CVF Conf Comput Vis Pattern Recognit 2022; 12175-12185.