

The Design of a Deep Learning-based Adaptive Multi-Channel Fusion Network for Diabetes Diagnosis

Peng Xia¹, Qi Qi¹, Yucong Duan¹, Ni Li^{2,3} and Xinying Wang⁴

¹ School of Computer Science and Technology, Hainan University, Haikou, China

² School of Mathematics and Statistics, Hainan Normal University, Haikou, China

³ Key Laboratory of Data Science and Smart Education, Hainan Normal University, Ministry of Education, Haikou, China

⁴ School of Management and Economics, North China University of Water Resources and Electric Power, Zhengzhou, China

Corresponding Author

Dr. Qi Qi

Hainan University

E-mail address: qqi@hainanu.edu.cn

Abstract. Accurate diagnosis of diabetes is crucial for effective health management of patients. Recent advances in machine learning have shown promising predictive results in diabetes diagnosis. In this paper, we developed an Adaptive Multi-Channel Fusion Network (AMCFN). Specifically, we have defined a feature enhancement module that combines an attention mechanism to dynamically assign relevant weights to input data, thereby enabling the model to focus on processing task-relevant outputs. Meanwhile, we design a multi-channel fusion network that utilizes different network structures to simultaneously extract various deep features, including temporal features and nonlinear features, from the input data. Extensive experiments were conducted on the Pima Indian Diabetes Dataset (PIDD) and the Early-Stage Diabetes Risk Prediction Dataset (ESDRPD). Our model achieved high predictive accuracies of 95.83% and 99.6%, respectively. These results outperformed existing baseline models in diabetes diagnosis. Ablation experiments emphasized the power of the feature enhancement module and the multi-channel fusion network. Finally, we analyzed the prediction process of AMCFN using SHapley Additive exPlanations (SHAP).

The analysis results show the importance ranking of each feature to the model output in different channels, and the importance ranking of each channel to the final diabetes diagnosis. This enhances the interpretability of AMCFN and validates the effectiveness of the multi-channel design. Our model demonstrates potential in diabetes diagnosis and is expected to increase end-user trust and confidence in early detection of diabetes.

Keywords: Diabetes diagnosis, Adaptive Multi-Channel Fusion Network, Feature enhancement, SHapley Additive exPlanations.

1 Introduction

Diabetes is a common long-term disease. Diabetes occurs when the pancreas fails to produce adequate insulin or when the body becomes resistant to insulin's effects. Early diabetes detection contributes to extending patients' lifespans, which enable them to lead healthier and happier lives. At the same time, it can also reduce costly medical expenses. The insulin hormone and glucose levels in the bloodstream commonly serve as diagnostic factors for diabetes. When our body digests food, glucose is released into the bloodstream and insulin's job is to instruct cells to convert glucose in the blood into energy [1]. If insulin is not utilized correctly, the cells are unable to absorb glucose and blood glucose levels rise uncontrollably [2]. In addition, age is another factor associated with diabetes. Diabetes-related problems such as high blood pressure are more common in older people, making diabetes more difficult to control.

With recent advances in artificial intelligence, patient data can be used to train and develop a precise classifier for early diabetes diagnosis [1][3]. Various machine learning methods, including Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machines (SVM), among others, have been applied in the diagnosis of diabetes [4][5]. Deep learning, which has achieved tremendous success in a variety of sectors during the last decade, can identify deep feature patterns with higher accuracy compared to traditional machine learning methods [6-9].

In the research on diabetes diagnosis, we utilize tabular data. Deep neural networks encounter several obstacles when applied to tabular data, including but not limited to the lack of locality, data sparsity, mixed feature types, and limited knowledge of the dataset structure [5]. Despite these challenges, there is still value in exploring novel

deep-learning architectures for tabular data. Deep learning can significantly improve prediction performance, achieve end-to-end learning, capture representative features, and is suitable for streaming data and multiple data types, while simplifying tedious feature engineering [10]. Motivated by the advantages of deep learning, we have designed an adaptive multi-channel fusion network to improve the prediction performance of diabetes, which combines an attention mechanism to capture and enhance relevant information of the input data.

Neural networks have found extensive application and validation in the field of medicine, demonstrating predictive performance superior to traditional statistical methods. However, neural networks often exhibit a "black box" nature, meaning that it is difficult to explain the internal mechanisms of the model and the basis for its predictive results. This is a drawback in the medical field because healthcare professionals need to understand how the model makes predictions based on patient features and provide relevant recommendations. At the same time, they also need to know which features significantly influence the prediction results. The aim of this study is to design a model for predicting diabetes and to apply the SHAP method for interpretability analysis of the model. By utilizing the SHAP method to rank the importance of each feature in different channels, it not only reveals the relevant features that affect the risk of diabetes, but also proves the effectiveness of the model design.

The present study's contributions can be briefly outlined in three aspects:

- (1) Feature Enhancement Module: An adaptive attention weight matrix is defined to enhance features in tabular data.
- (2) Multi-Channel Fusion Network: A three-channel neural network is designed, with each channel employing a different network architecture. Subsequently, the outputs from the three channels are fused together for final prediction. This design takes into account the nonlinearity, time, and other complex data information of the input data by observing the input data from multiple angles. The proposed method outperforms all baseline models in comparison experiments.
- (3) SHAP Evaluation: Using SHAP for explanatory analysis of AMCFN not only reveals the importance ranking of input features within each channel to the model output, as well as the importance ranking of different channels to the final prediction results, but also demonstrates the effectiveness of the design of AMCFN.

2 Related Work

Many researchers [10-18] constructed classification models to predict diabetes using the PIDD. This dataset is widely regarded as suitable for conducting baseline tests on existing models. The utilization of ESDRPD is somewhat infrequent due to its recent release in the year 2020. Therefore, it is also a good option for validating our model. This section summarizes a few relevant studies using these two datasets, some of which [10-18] and [18-26] used PIDD and ESDRPD, respectively. Specifically summarized in Table 1.

Table 1. Performing classification model-related work using (*) Dataset 1 (PIDD) and (**) Dataset 2 (ESDRPD).

Proposed model and data used	Accuracy (%)	Refs.
Model: Ensemble using the soft voting method with Logistic Regression (LR), Naive Bayes (NB), and Random Forest (RF) classifiers *	79.04	[10]
Model: A stacking-based heterogeneous ensemble classifier (Level 0 models: MLP and SVM, and Level 1 classifier: Logistic Regression) *	78.20	[11]
Model: Random Forest (RF) integrated with chaotic multi-verse optimization (CMVO) for feature selection (OSMOTE-CMVO-RF) *	89.04	[12]
Model: Deep Neural Networks (DNN) fine-tuned with stacked autoencoders (SAE) using the backpropagation method (DNN-SAE) *	86.26	[13]
Model: The integration of the C4.5 algorithm with SMOTE and the Interquartile Range (IQR) algorithm (IQR-SMOTE-C4.5) *	89.50	[14]
Model: A Functional Fuzzy Wavelet Neural Network optimized using the Teaching-Learning-Based Optimization Algorithm. (TLBO- FFWNN) *	88.68±3.02	[15]
Model: ANN using Equilibrium Optimizer (ANN-EO) *	77.85 ± 0.06	[16]
Model: Spider Monkey Optimization-based Rule Miner (RM-SMO) *	89.87	[17]
Model: TabNet model optimized by Bayesian Optimization (TabNet-BO) * and **	92.20*, 99.40**	[18]
Model: Binary logistic regression (BLR) integrated with XGBoost for feature selection (XGBoost-BLR) **	98.10	[19]
Model: Extra Trees combined with Pearson Correlation feature selection (ET-PC) **	99.06	[20]
Model: Feed-Forward Neural Networks (FFNN) optimized using the CSA algorithm (CSA-FFNN) **	99.04	[21]
Model: Utilize Grey Wolf Optimization (GWO) and Adaptive Particle swarm	97.00	[22]

optimization (AP) to optimize MLP (APGWO-MLP) **		
Model: Artificial Neural Networks (ANN) combined with Information Gain (ANN-IG) **	98.10	[23]
Model: Integration between Association Rules and KNN (KNN-AR) **	97.36	[24]
Model: Stacking ensemble technique (CNN, SVM and CNN) **	98.71	[25]
Model: Stacking ensemble technique (LR, KNN, CART, SVM, AdaBoost and GDBT) **	97.00	[26]

The field of machine learning has witnessed significant progress in both the creation and implementation of techniques for the classification of diabetes, which from initial independent models to current deep learning models. Research presented in Table 1 shows that these models exhibit high classification accuracy in predicting diabetes. However, there still exists the potential for further improvement in diabetes prediction. We proposed the AMCFN, which integrates deep learning techniques such as attention mechanism and GRU to enhance diabetes prediction.

In the field of eXplainable Artificial Intelligence (XAI), the SHAP [27] has garnered significant attention as a valuable contribution to addressing model opacity and enhancing interpretability. The SHAP employs a game-theoretic approach aimed at determining Shapley values to measure the influence of each feature on instance predictions. Specifically, the SHAP reveals the decision-making process of a model by computing the contribution of each feature to the prediction outcome. For instance, Gu et al. [28] utilized the SHAP method to calculate the weights of different features (i.e., Shapley values) to explain both positive and negative outcomes of breast cancer recurrence. Additionally, researchers can utilize SHAP feature importance to explain the relationship between the features generated by the model and the outcomes. For example, Meena & Hasija [29] employed feature weights to rank and identify significant genes correlated with the progression of squamous cell carcinoma. The SHAP is highly trusted because it considers all possible input combinations for all potential predictions within a sample, thereby ensuring consistency and accuracy in explanations.

3 Methods

3.1 Data pre-processing

Despite diabetes being a prevalent and life-threatening condition, only a few databases are publicly available. Among these, the PIDD is the most commonly used, whereas the latest one is the ESDRPD. In this study, benchmark testing was conducted using PIDD and ESDRPD to compare the performance of the proposed model.

In this study, data description was conducted to better understand the features of the two datasets used, as illustrated in Table 2. Unlike the ESDRPD, the PIDD contains many outliers. In this study, box plots and interquartile range (IQR) were used to detect outliers in the PIDD, as shown in Figure 1. Before performing data cleaning, the division of the two datasets into training and testing sets was implemented to mitigate the risk of data leakage. The two datasets are randomly split in a 70:30 ratio, with 70% of the data used for model training and 30% for model testing. Hence, both the training and testing datasets underwent separate cleaning processes. This study removed duplicate values and employed median imputation to fill in missing values. Following this, outliers were replaced with the median value of each feature. For discrete variables, label encoding was applied, and finally, the dataset was normalized.

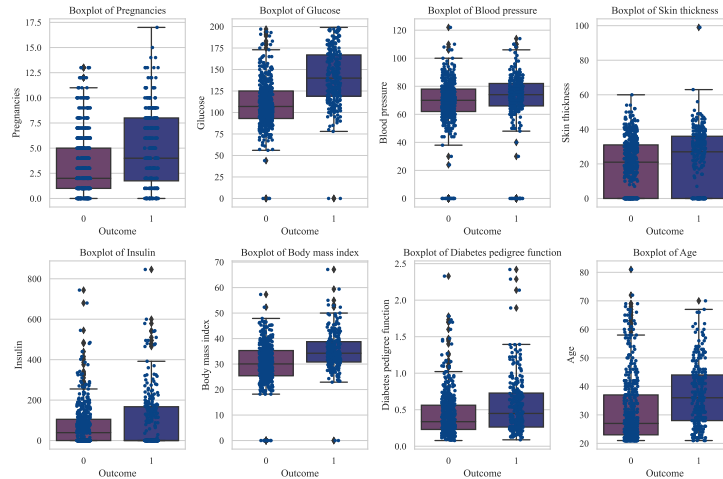


Fig. 1. The PIDD Outlier Detection with Box Plots and Interquartile Range Analysis.

Table 2. Description and original range of attributes for PIDD and ESDRPD.

Dataset 1-PIDD

Attributes	Description	Range	Type
Pregnancies	Number of pregnant	(0-17)	Numeric
Glucose	Oral Glucose Tolerance Test (OGTT) 2-hour plasma glucose concentration (mg/dl)	(0-199.0)	Numeric
Blood Pressure	Diastolic blood pressure (mm Hg)	(0-122.0)	Numeric
Skin Thickness	Triceps Skinfold Thickness (mm)	(0-99.0)	Numeric
Insulin	2-hour Serum Insulin (μ U/ml)	(0-846.0)	Numeric
BMI	Body mass index (kg/m^2)	(0-67.1)	Numeric
Diabetes Pedigree Function	Diabetes pedigree function	(0.078- 2.42)	Numeric
Age	Patient ages (years)	(21-81)	Numeric
Outcome	Diabetic:1, Non-diabetic:0	(0-1)	Binary
Dataset 2-ESDRPD			
Age	Patient ages (years)	(16-90)	Numeric
Gender	Female:0, Male:1	(0-1)	Binary
Polyuria	Polyuria:1, No-Polyuria:0	(0-1)	Binary
Polydipsia	Polydipsia:1, No-Polydipsia:0	(0-1)	Binary
Sudden weight loss	Sudden weight loss:1, No-Sudden weight loss:0	(0-1)	Binary
Weakness	Weakness:1, No-Weakness:0	(0-1)	Binary
Polyphagia	Polyphagia:1, No-Polyphagia:0	(0-1)	Binary
Genital thrush	Genital thrush:1, No-Genital thrush:0	(0-1)	Binary
Visual blurring	Visual blurring:1, No-Visual blurring:0	(0-1)	Binary
Itching	Itching:1, No-Itching:0	(0-1)	Binary
Irritability	Irritability:1, No-Irritability:0	(0-1)	Binary
Delayed healing	Delayed healing:1, No-Delayed healing:0	(0-1)	Binary
Partial paresis	Partial paresis:1, No-Partial paresis:0	(0-1)	Binary
Muscle stiffness	Muscle stiffness:1, No-Muscle stiffness:0	(0-1)	Binary
Alopecia	Alopecia:1, No Alopecia:0	(0-1)	Binary
Obesity	Obesity:1, No-Obesity:0	(0-1)	Binary
Class	Diabetic:1, Non-diabetic:0	(0-1)	Binary

3.2 Proposed Model Architecture

The innovation of the AMCFN lies in the design of its feature enhancement module and multi-channel fusion module. The feature enhancement module enables the model to focus on processing features relevant to the task of diabetes prediction. The multi-channel fusion module integrates the outputs from each channel to diagnose diabetes ultimately. Due to the diverse structural designs of each channel, the model can observe input data from various perspectives, effectively enhancing the model's performance. We conducted an analysis of the prediction process of AMCFN using the SHAP and visualized it, demonstrating the importance ranking of input features across different channels as well as the importance ranking of different channels for the final diagnosis of diabetes. The high-level architecture diagram for diabetes diagnosis is illustrated in Figure 2.

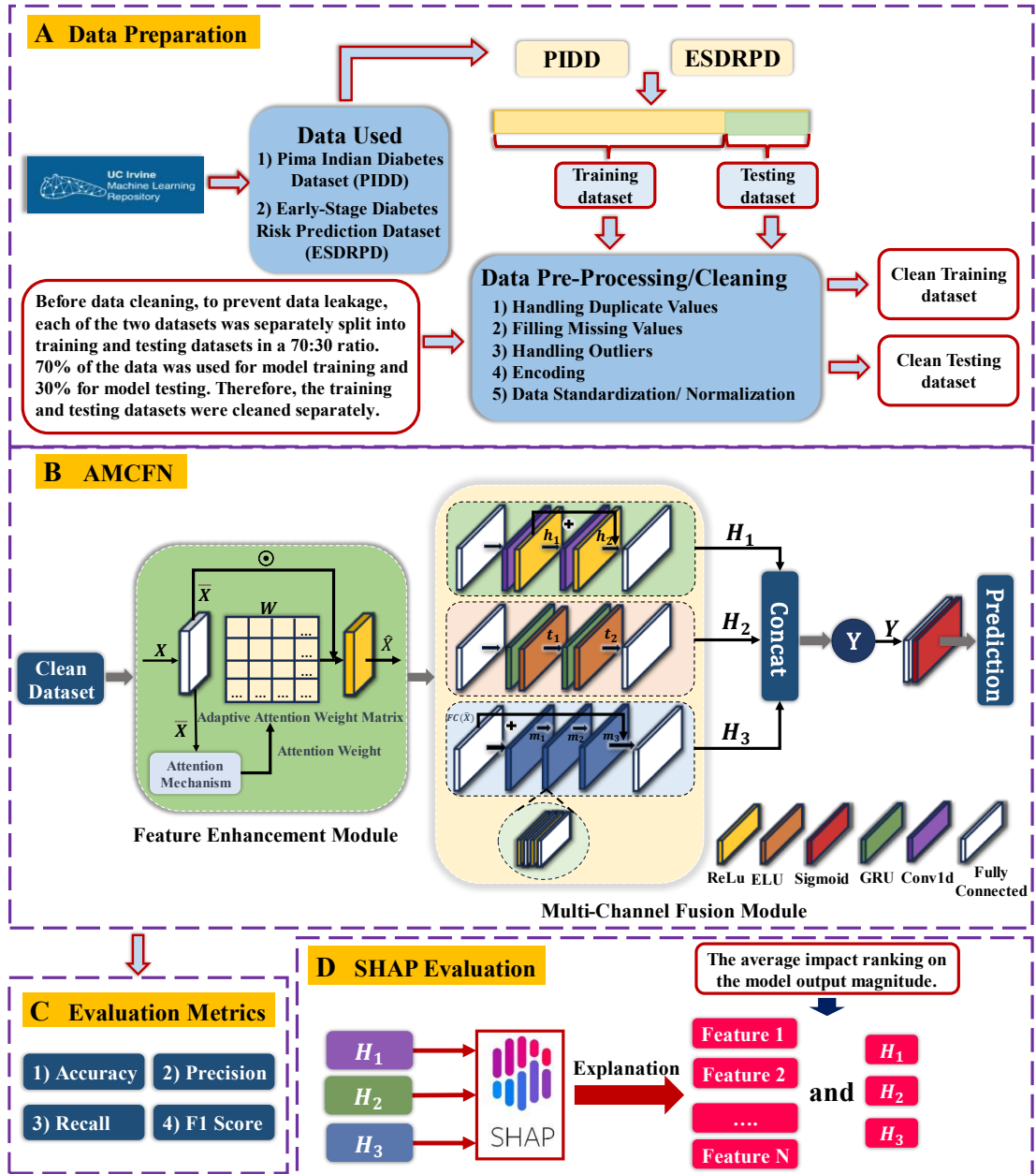


Fig. 2. High-level architecture diagram of the diabetes diagnosis model.

Feature Enhancement Module

In the diabetes dataset, not all features have an equal impact on the diagnosis of diabetes. Therefore, in this study, we introduce a feature enhancement module aimed at dynamically improving the input data. This enables the model to focus on processing task-relevant features, thereby improving the accuracy of diabetes diagnosis.

The input to the feature enhancement module is the pre-processed diabetes feature matrix X . Firstly, the feature matrix X is input into a fully connected layer with a large number of neurons. This layer maps the features into a new feature space, effectively increasing the dimensionality of the input data, denoted as \bar{X} . Secondly, we defined an adaptive attention weight matrix W , which can automatically adjust based on changes in the output dimensions of different fully connected layers. Filling the matrix W with the weights obtained using the dot-product attention mechanism [30]. Finally, the enhancement of the input data is achieved by multiplying the feature matrix \bar{X} with the weight matrix W . The result is denoted as \hat{X} :

$$\hat{X} = ReLu(W \odot \bar{X}) = ReLu(W \odot FC(X)). \quad (1)$$

where $ReLu(.)$ represents the ReLu activation function, $FC(.)$ represents the fully connected layer, \odot represents the Hadamard product.

Multi-Channel Fusion Module

After the feature enhancement module, we constructed a multi-channel network architecture, where each channel adopts a different network structure. This design aims to comprehensively analyze input data from different perspectives, such as nonlinearity and time. Specifically, we designed three channels, each of which initially includes a fully connected layer to process \hat{X} and adapt it to the input dimensions of each channel. Finally, each channel undergoes a fully connected layer to ensure that their output dimensions are the same for the ultimate multi-channel fusion.

The first channel consists of two fully connected layers and two one-dimensional convolutional layers. Firstly, after the fully connected layer, different local features in the data are learned through the first convolutional layer, and the result is recorded as h_1 . Secondly, through skip connections, the model can establish connections between convolutional layers, which helps gradient propagation and effectively learns the characteristics of the data. Finally, a fully connected layer is used to control the output dimension of the first channel, which can be written as:

$$\begin{aligned}
h_1 &= \text{Relu}\left(\text{Conv}\left(\text{FC}(\mathcal{X})\right)\right), \\
H_1 &= \text{FC}(h_2 + h_1) = \text{FC}(\text{Relu}(\text{Conv}(h_1)) + h_1).
\end{aligned} \tag{2}$$

where $\text{Conv}(\cdot)$ represents the one-dimensional convolutional layer.

In the second channel, the Gated Recurrent Unit (GRU) is utilized to capture the temporal dependence between the target and feature variables. The GRU is a recurrent neural network that utilizes GRU cell units. The hidden state S^t of each time step is controlled by the hidden state S^{t-1} of the previous time step and the current input X^t . The information flow in the GRU model is controlled by the GRU cell unit, and its structure is shown in Figure 3. The GRU unit incorporates two primary gates: the reset gate and the update gate, indicated by dashed boxes. On the left is the reset gate, which is responsible for resetting S^{t-1} , and on the right is the update gate, which is used to update the previous state S^{t-1} and the current hidden state \tilde{S}^t . The detailed mechanisms working on the two gates can be represented as follows:

$$\begin{aligned}
z^t &= \sigma(W_z X^t + U_z S^{t-1} + b_z), \\
r^t &= \sigma(W_r X^t + U_r S^{t-1} + b_r).
\end{aligned} \tag{3}$$

where z^t and r^t represents the update gate and reset gate, $\sigma(\cdot)$ represents the sigmoid activation function, trainable parameters: $W_z, W_r, U_z, U_r, b_z, b_r$.

The output of the GRU unit can be described as follows:

$$\begin{aligned}
\tilde{S}^t &= \tanh(W_h X^t + U_h (r^t \odot S^{t-1}) + b_h), \\
S^t &= z^t \odot \tilde{S}^t + (1 - z^t) \odot S^{t-1}.
\end{aligned} \tag{4}$$

where W_h, U_h, b_h are trainable parameters.

In the second channel, GRU is stacked twice, and the final output is acquired through a fully connected layer, which can be written as:

$$\begin{aligned}
t_1 &= \text{ELU}\left(\text{gru}\left(\text{FC}(\mathcal{X})\right)\right), \\
H_2 &= \text{FC}(t_2) = \text{FC}(\text{ELU}(\text{gru}(t_1))).
\end{aligned} \tag{5}$$

where $\text{gru}(\cdot)$ represents the GRU layer, ELU represents the ELU activation function.

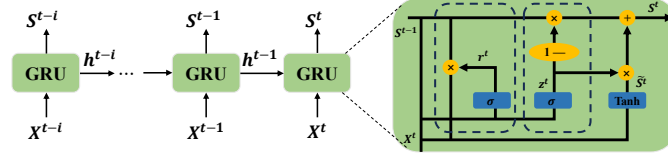


Fig. 3. The architecture of the GRU.

The third channel aims to deepen the structure of the network layers, preserving and reusing the feature information learned at previous levels to enhance the model's feature learning capabilities. The deep blue layer in Figure 2 is referred to as the "mapping" layer. Each "mapping" layer consists of two ReLu activation functions and two fully connected layers. Firstly, feature transformation is performed by the first fully connected layer, which modifies the feature dimensions. Subsequently, deep feature representations are learned through three "mapping" layers. Finally, the current feature is added to the previously stored features through skip connections. This approach enables information retention and transfer, facilitating deeper learning of feature representations and ultimately improving the overall network performance, which can be written as:

$$\begin{aligned}
 m_2 &= \text{map}(m_1) = \text{map}\left(\text{map}\left(\text{FC}(\hat{X})\right)\right), \\
 H_3 &= \text{FC}\left(m_3 + \text{FC}(\hat{X})\right) = \text{FC}\left(\text{map}(m_2) + \text{FC}(\hat{X})\right).
 \end{aligned} \tag{6}$$

where $\text{map}(\cdot)$ represents the "mapping" layer.

We aggregate the outputs of the three channels H_1, H_2 and H_3 . Finally, applying the fully connected layer and the sigmoid activation function, the binary prediction output is obtained, which can be written as:

$$\text{Prediction} = \sigma\left(\text{FC}(\text{concat}(H_1, H_2, H_3))\right). \tag{7}$$

SHAP Method

To enhance the interpretability of our model, this study adopts the SHAP method to identify the feature discrepancies that each channel attends to, as well as how much each channel influences the final diagnosis of diabetes. The SHAP method utilizes principles from cooperative game theory to allocate importance scores to each input feature for a given prediction outcome. Game theory consists of a set of principles in

which players involved in a game hold a range of strategies and are motivated by various forms of rewards. Shapley values are utilized to elucidate the individual contributions of each participant within the game. To elucidate the model, strategies symbolize the generation of outcomes, players represent features, and rewards signify the quality of the outcomes obtained. By treating the input data as a collection of individual participants, the SHAP approximates Shapley values through Shapley kernels, thereby providing reasonable and reliable explanatory weights for each feature and the model's prediction outcomes. This effectively enhances the interpretability and applicability of the model. The calculation method for SHAP values is based on the following formula:

$$\phi_i(f) = \sum_{S \subseteq \{1, \dots, M\} \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f(x_S \cup \{i\}) - f(x_S)]. \quad (8)$$

where $\phi_i(f)$ represents the contribution of feature i to the model f (i.e., the SHAP value.), M is the total number of features, x is the input sample, S is a set of features, x_S is the input sample with only the features in the set S , $f(x_S)$ is the model prediction without the features in set S , and $f(x_S \cup \{i\})$ is the model prediction with feature i included.

4 Results and Discussion

4.1 Model Performance

In this part, we demonstrate the performance difference between the proposed model AMCFN and the baseline model through extensive experiments, as shown in Table 3 and Table 4. In this study, accuracy, precision, recall, and F1 score were used as evaluation metrics.

Combining Tables 3 and 4, our proposed model outperforms the baseline model in overall performance. Specifically, on the PIDD, the AMCFN accuracy is 95.83%, precision is 94.40%, recall is 95.74%, and F1 score is 95.07%. The AMCFN outperformed the baseline model by 3.63%, 3.74%, 6.34%, and 5.57%, respectively. On the ESDRPD, the AMCFN accuracy is 99.60%, precision is 100%, recall is 99.32%, and F1 score is 99.51%. The accuracy, recall, and F1 score are respectively higher than the baseline model by 0.2%, 0.32%, and 0.21%. We conducted ablation experiments on each of the two datasets as shown in Table 5. The results indicate that the

introduction of the feature enhancement module enables the model to better process inputs relevant to the current task, thereby improving overall performance. Due to the distinct designs of each channel, considering different feature information, the fusion of the three channels significantly enhances the overall performance of the model, reaffirming the rationality and excellent performance of the AMCFN design.

Table 3. PIDD- Comparison of AMCFN with methods in other literature

Dataset 1-PIDD				
Method [Refs.]	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
AMCFN [Current study]	95.83	94.40	95.74	95.07
Soft voting (LR, NB and RF) [10]	79.04	73.48	71.45	80.60
Stacking (MLP, SVM and LR) [11]	78.20	72.20	—	59.40
OSMOTE-CMVO-RF [12]	89.04	88.17	—	89.00
DNN-SAE [13]	86.26	90.66	87.92	89.27
IQR-SMOTE-C4.5 [14]	89.50	90.00	89.40	89.50
TLBO- FFWNN [15]	88.68±3.02	—	—	—
ANN-EO [16]	77.85 ± 0.06	74.61 ± 0.11	—	—
RM-SMO [17]	89.87	—	—	—
TabNet-BO [18]	92.20	89.50	87.20	88.30

Table 4. ESDRPD- Comparison of AMCFN with methods in other literature

Dataset 2-ESDRPD				
Method [Refs.]	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
AMCFN [Current study]	99.60	100.00	99.32	99.51
TabNet-BO [18]	99.40	100.00	98.60	99.30
XGBoost-BLR [19]	98.10	—	—	—
ET-PC [20]	99.06	98.77	—	98.53
CSA-FFNN [21]	99.04	—	—	—
APGWO-MLP [22]	97.00	97.00	99.00	98.00
ANN-IG [23]	98.10	98.40	—	98.40
KNN-AR [24]	97.36	98.22	—	—

Stacking (CNN, SVM and CNN) [25]	98.71	—	—	—
Stacking (LR, KNN, CART, SVM, AdaBoost and GDBT) [26]	97.00	—	—	—

Table 5. Ablation experiments

Ablation experiment	Accuracy	Precision	Recall	F1 Score
PIDD				
AMCFN	0.9583	0.9440	0.9574	0.9507
No Feature Enhancement Module	0.9167	0.8973	0.9291	0.9129
H_2+H_3	0.9233	0.9402	0.8936	0.9163
H_1+H_3	0.9300	0.9348	0.9149	0.9247
H_1+H_2	0.9266	0.9343	0.9078	0.9208
ESDRPD				
AMCFN	0.9960	1.0000	0.9932	0.9951
No Feature Enhancement Module	0.9744	0.9688	0.9894	0.9789
H_2+H_3	0.9935	0.9937	0.9935	0.9936
H_1+H_3	0.9872	1.0000	0.9804	0.9901
H_1+H_2	0.9808	0.9896	0.9794	0.9845

4.2 SHAP Evaluation

Figures 4 and 5 display the average impact ranking of various features and channels on the magnitude of model output, calculated using the SHAP method. The ranking of features is determined based on their importance (i.e., SHAP value) to the model output in different channels. In the PIDD, when considering all three channels, the top three ranked features are “Insulin”, “Glucose”, and “Age”, while the bottom ranked two features are “Blood pressure” and “Diabetes pedigree function”, as shown in Figure 4.

The second channel contains a GRU structure sensitive to time series, which results in more attention being focused on the “Age” feature, as depicted in (b) of Figure 4. In the final diagnosis of the model, the impact of the third channel is the greatest, while the impact of the second channel is the smallest, as shown in (d) of Figure 4. In the process of model prediction, the features of “Insulin”, “Glucose”, and “Age” have a more significant impact on the predictive outcomes.

For the ESDRPD, we conducted the same experiment, integrating the feature importance rankings from three channels. The top ranked three features are “Polydipsia”, “Age”, and “Itching”, while the last ranked feature is “Obesity”, as shown in Figure 5. Similarly, the second channel pays more attention to “Age”, as shown in (b) of Figure 5. In the final diagnosis of the model, the third channel has the greatest impact on the model output, while the first channel has the smallest impact on the model output, as shown in (d) of Figure 5.

In summary, the SHAP method reveals differences in the importance ranking of features in each channel, and the impact of each channel on the final diagnosis is also different. This further enhances understanding of the model's prediction process and improves the model's interpretability. It provides medical professionals with evidence of the importance of each feature, helping doctors better understand the model's prediction process.

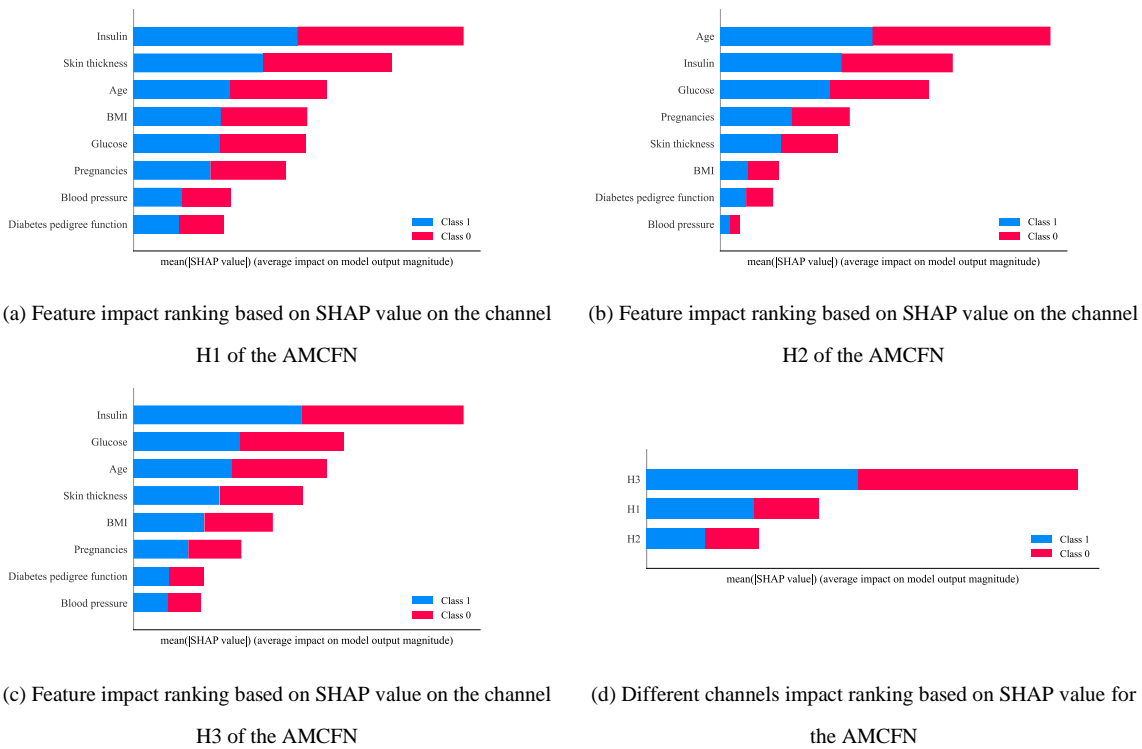
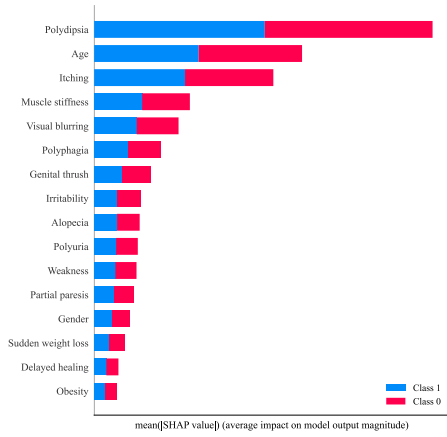
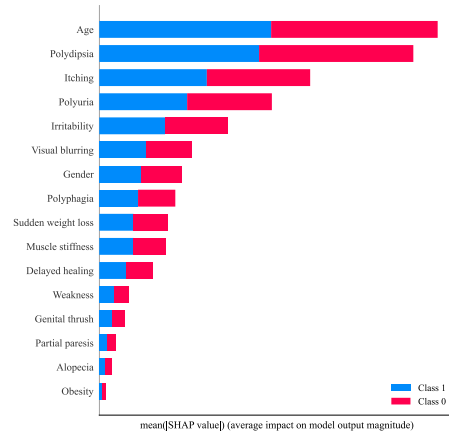


Fig. 4. Impact ranking of features and channels based on SHAP values for the AMCFN in the PIDD.

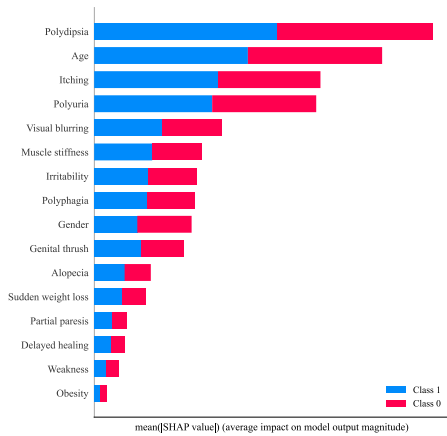


(a) Feature impact ranking based on SHAP value on the channel



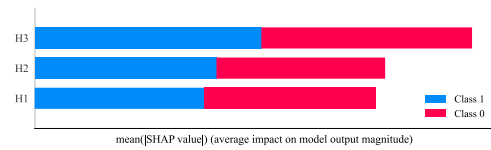
(b) Feature impact ranking based on SHAP value on the channel

H1 of the AMCFN



(c) Feature impact ranking based on SHAP value on the channel

H2 of the AMCFN



(d) Different channels impact ranking based on SHAP value for the AMCFN

Fig. 5. Impact ranking of features and channels based on SHAP values for the AMCFN in the ESDRPD.

5 Conclusion

In this paper, we introduced an adaptive multi-channel neural network for the diagnosis of diabetes. The model consists of the feature enhancement module and the multi-channel neural network module, enabling it to focus on processing input relevant to the

current task and observe data from multiple perspectives using different network architectures. We conducted experiments on two commonly used diabetes baseline datasets, and the results show that our proposed model outperforms the existing baseline models in the diagnosis of diabetes. The ablation experiments indicate that the adaptive attention weight matrix has significant advantages in feature enhancement, and multi-channel neural networks can effectively focus on input features from different perspectives. Finally, we analyzed the prediction process of AMCFN using the SHAP method. The analysis results show the ranking of the importance of each feature in different channels for the model output, and the ranking of the importance of each channel for the final diagnosis of diabetes. Provides a basis for healthcare professionals to help them understand the importance of various characteristics for diabetes prediction and be able to provide relevant medical advice to patients based on this information. Through the SHAP method, it enhanced the interpretability of the AMCFN in the process of diabetes prediction, and demonstrated the effectiveness of the multi-channel design.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant No. 11861030, the Hainan Provincial Natural Science Foundation of China under Grant No. 621RC511, and the Education Department of Hainan Province with project number Hnky2023-4.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Data availability. The data used in the study comes from the UCI public database. ([Home - UCI Machine Learning Repository](#))

References

- 1 J. J. Khanam and S. Y. Foo: "A comparison of machine learning algorithms for diabetes prediction." *ICT Express*, vol. 7, no. 4, pp. 432–439 (Dec. 2021).
- 2 J. Yadav, A. Rani, V. Singh, and B. M. Murari: "Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy." *Biomedical Signal Processing and Control*, vol. 18, pp. 214–227 (Apr. 2015).

- 3 C.-L. Huang, M.-C. Chen, and C.-J. Wang: "Credit scoring with a data mining approach based on support vector machines." *Expert Systems with Applications*, vol. 33, no. 4, pp. 847–856 (Nov. 2007).
- 4 I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda: "Machine Learning and Data Mining Methods in Diabetes Research." *Computational and Structural Biotechnology Journal*, vol. 15, pp. 104–116 (2017).
- 5 A. Alharbi and M. Alghahtani: "Using Genetic Algorithm and ELM Neural Networks for Feature Extraction and Classification of Type 2-Diabetes Mellitus." *Applied Artificial Intelligence*, vol. 33, no. 4, pp. 311–328 (Mar. 2019).
- 6 Shwartz-Ziv, R., Armon, A.: "Tabular data: Deep learning is not all you need." *Information Fusion*, vol. 81, pp. 84–90 (May 2022).
- 7 I. Goodfellow, Y. Bengio, A. Courville: "Deep Learning." MIT Press (2016).
- 8 D. Ravi et al.: "Deep Learning for Health Informatics." *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21 (Jan. 2017).
- 9 S. O. Arik and T. Pfister: "TabNet: Attentive Interpretable Tabular Learning." In: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 6679–6687 (2021).
- 10 S. Kumari, D. Kumar, and M. Mittal: "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier." *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40–46 (Jun. 2021).
- 11 S. K. Kalagotla, S. V. Gangashetty, and K. Giridhar: "A novel stacking technique for prediction of diabetes." *Computers in Biology and Medicine*, vol. 135, p. 104554 (Aug. 2021).
- 12 S. Sreejith, H. Khanna Nehemiah, and A. Kannan: "Clinical data classification using an enhanced SMOTE and chaotic evolutionary feature selection." *Computers in Biology and Medicine*, vol. 126, p. 103991 (Nov. 2020).
- 13 K. Kannadasan, D. R. Edla, and V. Kuppili: "Type 2 diabetes data classification using stacked autoencoders in deep neural networks." *Clinical Epidemiology and Global Health*, vol. 7, no. 4, pp. 530–535 (Dec. 2019).
- 14 N. Nnamoko and I. Korkontzelos: "Efficient treatment of outliers and class imbalance for diabetes prediction." *Artificial Intelligence in Medicine*, vol. 104, p. 101815 (Apr. 2020).
- 15 J. S. Majeed Alneamy, Z. A. Hameed Alnaish, S. Z. Mohd Hashim, and R. A. Hamed Alnaish: "Utilizing hybrid functional fuzzy wavelet neural networks with a teaching learning-based optimization algorithm for medical disease diagnosis." *Computers in Biology and Medicine*, vol. 112, p. 103348 (Sep. 2019).

- 16 Si, T., Bagchi, J., Miranda, P. B. C.: Artificial Neural Network training using metaheuristics for medical data classification: An experimental study. *Expert Systems with Applications*, vol. 193, p. 116423 (May 2022).
- 17 Cheruku, R., Edla, D. R., Kuppili, V.: SM-RuleMiner: Spider monkey based rule miner using novel fitness function for diabetes classification. *Computers in Biology and Medicine*, vol. 81, pp. 79–92 (Feb. 2017).
- 18 Joseph, L. P., Joseph, E. A., Prasad, R.: Explainable diabetes classification using hybrid Bayesian-optimized TabNet architecture. *Computers in Biology and Medicine*, vol. 151, p. 106178 (Dec. 2022).
- 19 Y. Wu et al.: "Novel binary logistic regression model based on feature transformation of XGBoost for type 2 Diabetes Mellitus prediction in healthcare systems." *Future Generation Computer Systems*, vol. 129, pp. 1–12 (Apr. 2022).
- 20 Hasan, S. M. Mahedy, Rabbi, Md. F., Champa, A. I., Zaman, Md. A.: A Machine Learning-Based Model for Early Stage Detection of Diabetes. In: 2020 23rd International Conference on Computer and Information Technology (ICIT), DHAKA, Bangladesh: IEEE, pp. 1–6 (Dec. 2020).
- 21 Yasar, A.: Data Classification of Early-Stage Diabetes Risk Prediction Datasets and Analysis of Algorithm Performance Using Feature Extraction Methods and Machine Learning Techniques. *ijisae* 9(4), 273–281 (Dec. 2021).
- 22 Le, T. M., Vo, T. M., Pham, T. N., et al.: A novel wrapper-based feature selection for early diabetes prediction enhanced with a metaheuristic. *IEEE Access* 9, 7869–7884 (2020).
- 23 Chaves, L., Marques, G.: Data mining techniques for early diagnosis of diabetes: A comparative study. *Applied Sciences (Switzerland)* 11(5), 1–12 (Mar. 2021).
- 24 Khafaga, D. S., Alharbi, A. H., Mohamed, I., Hosny, K. M.: An Integrated Classification and Association Rule Technique for Early-Stage Diabetes Risk Prediction. *Healthcare*, vol. 10, no. 10, p. 2070 (Oct. 2022).
- 25 Tan, Y., Chen, H., Zhang, J., Tang, R., Liu, P.: Early Risk Prediction of Diabetes Based on GA-Stacking. *Applied Sciences*, vol. 12, no. 2, p. 632 (Jan. 2022).
- 26 Saxena, S., Mohapatra, D., Padhee, S., Sahoo, G.K.: Machine learning algorithms for diabetes detection: a comparative evaluation of performance of algorithms. *Evol. Intell.* 1–17 (2021).
- 27 Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I.: From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 56–67 (2020).

- 28 Gu, D., Su, K., Zhao, H.: A case-based ensemble learning system for explainable breast cancer recurrence prediction. *Artif. Intell. Med.* 107, 101858 (2020).
- 29 Meena, J., Hasija, Y.: Application of explainable artificial intelligence in the identification of Squamous Cell Carcinoma biomarkers. *Comput. Biol. Med.* 146, 105505 (2022).
- 30 Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).