Study on Regression Prediction of Blood Glucose Levels Based on RFA-CNN Model

Weipu Wu, Zhenhua Qu, Yuan Cao*

Abstract—**Diabetes mellitus, a leading global health threat, demands precise blood glucose prediction for effective clinical management, yet traditional models struggle to capture nonlinear dynamics. This study proposes the RFA-CNN model, a novel framework integrating Random Forest (RF) and Convolutional Neural Network (CNN), to address this challenge: RF identifies 31 key physiological features (e.g., age, hemoglobin) via an automatic search algorithm while CNN extracts local features from high-dimensional data, and the SHAP framework is integrated to quantify feature contributions and reveal nonlinear interactions, enabling both global importance ranking and individual sample attribution. Employing multiple imputations for data preprocessing, the model reduces RMSE by 1.2% compared to suboptimal methods, ensuring data integrity. Evaluated on 5,642 clinical samples, RFA-CNN achieves a test set MSE of 0.003689 (4.97% lower than the second-best MLP model), an MAE of 0.035084, and Clarke Error Grid results with 83.9% of predictions in clinically accurate Region A and over 99% in A+B regions, outperforming benchmarks like DNN and LSTM. SHAP analysis highlights age, high-density lipoprotein cholesterol, and albumin as critical influencers, underscoring the model's interpretability in uncovering complex physiological correlations. By synergizing RF's feature selection, CNN's local feature extraction, and SHAP's interpretability, RFA-CNN effectively captures nonlinear blood glucose patterns, offering robust, accurate predictions with superior generalization and clinical compliance to support personalized diabetes management, reducing decision risks, and enhancing therapeutic interventions, with future work focusing on multicenter validation and dynamic data integration.**

Index Terms—**Diabetes, Blood Glucose Prediction, CNN, Random Forest, Clarke Error Grid, SHAP**

I. Introduction

DIABETES is a chronic health condition in which the body is unable to properly process blood glucose, often due to insufficient insulin production (Type 1) or insulin resistance (Type 2) [2]. Alarmingly, this metabolic disorder ranks 7th among the top ten global causes of death, as highlighted by the 2016 Global Burden of Disease Study. On a global scale, diabetes imposes a substantial financial burden on healthcare systems, with 537 million (10.5%) adults aged 20–79 currently managing the disease [3]. It is genuinely concerning that the incidence rate of Diabetes Mellitus (DM) is projected to rise significantly. Specifically, it is expected to reach 643 million, accounting for 11.3% of the global population by 2030, and further increase to 783 million, which will make up 12.2% of the global population by 2045 [4]. Within this global trend, China faces a particularly pressing challenge. As the country with the world's largest diabetic population [5], China reported 140 million cases (10.6% prevalence) among 20–79-year-olds in 2021, with projections surpassing 174 million by 2045 [6]. Tragically, these numbers coincide with 1.39 million annual diabetes-related deaths in China, accounting for 20% of the global total [7].

Diabetes poses significant health threats not only through its high prevalence but also via its intricate complications. Chronic hyperglycemia progressively damages the systemic vascular system, leading to severe consequences such as cardiovascular diseases [8], diabetic retinopathy [1], nephropathy [9], and neuropathy [10]. According to the International Diabetes Federation, cardiovascular complications account for approximately 50% of diabetes-related deaths globally, with diabetic patients facing 2-4 times higher risks of coronary heart disease and stroke compared to the general population. In China, approximately 60% of type 2 diabetes patients have comorbid hypertension [11], further exacerbating cerebrovascular risks. Alarmingly, diabetes frequently coexists with other metabolic disorders like obesity and dyslipidemia, creating a vicious cycle that exponentially complicates disease management. This interconnected health crisis underscores the critical need for integrated strategies targeting both hyperglycemia and associated comorbidities.

Significant bottlenecks persist in current diabetes research and model development. In blood glucose prediction, dual limitations in methodology and data quality constrain the improvement of model performance. Conventional approaches, relying heavily on empirical formulas and linear regression models, struggle to capture the non-linear characteristics and dynamic fluctuations of blood glucose levels, suffering from notable prediction delays. While machine learning has transcended traditional frameworks, existing algorithms—confined mainly to early models like random forests or support vector machines—perform inadequately in processing high-dimensional time-series data, failing to fully explore the profound correlations between blood glucose and metabolic markers. Regarding data quality, contemporary studies commonly face the dual challenges of insufficient sample size and limited feature diversity. Most research incorporates only hundreds or even dozens of samples, leading to overfitting of individual characteristics and poor generalization to broader populations. Additionally, monitoring indicators primarily focus on basic parameters such as fasting blood glucose and HbA1c, lacking systematic analysis of critical pathological markers like insulin sensitivity and inflammatory cytokines. This results in models unable to comprehensively unravel the intricate pathological mech-

Manuscript received February 14, 2025; revised May 30, 2025. This work was supported in part by the Doctoral Research Initiation Fund of Shandong University of Technology under Grant No. 417037.

Weipu Wu is a graduate student in applied statistics at the School of Mathematics and Statistics, Shandong University of Technology, China (e-mail: 2863375150@qq.com).

Zhenhua Qu is a graduate student in statistics at the School of Mathematics and Statistics, Shandong University of Technology, China (e-mail: 18463064907@163.com).

Yuan Cao is an associate professor at the School of Mathematics and Statistics, Shandong University of Technology, China (Corresponding author, e-mail: yuancao@sdut.edu.cn).

anisms of diabetes, ultimately compromising both prediction accuracy and generalization capabilities.

This paper proposes a blood glucose concentration level prediction model named RFA-CNN (Random Forest Assisted Convolutional Neural Network) based on deep learning methods in response to the above issues. The main contributions are as follows:

(1)This paper innovatively proposes a blood glucose concentration prediction model, RFA-CNN. This model organically combines the CNN neural network model with the random forest feature selection method. In the face of massive non-time series physiological data, the optimal number of features that fit the data characteristics is determined through a built-in automatic search algorithm during the feature processing stage. This process fully utilizes CNN's powerful local feature extraction ability, enabling the model to accurately extract key features related to blood glucose concentration from high-dimensional physiological data. To enhance the interpretability of the model, we integrate the SHAP (Shapley Additive exPlanations) framework. Based on game-theoretic principles, SHAP can precisely quantify each feature's contribution to the model's prediction. Applying SHAP to the RFA-CNN model allows us to analyze how features like age, hemoglobin, and other physiological indicators impact blood glucose predictions. It offers a global view of feature importance and enables local interpretation for individual samples. This helps understand the complex non-linear relationships between features and the predicted blood glucose levels, providing more in-depth insights into the model's decision-making process. At the same time, the information crucial for blood glucose prediction is retained by taking advantage of the random forest's superiority in feature importance evaluation. The synergistic effect of the CNN, random forest, and SHAP significantly improves the model's blood glucose concentration prediction accuracy and endows the prediction results with higher robustness.

(2)In the data preprocessing stage, this study systematically compared five commonly used data imputation methods to explore the most suitable method for imputing missing values. Through the key evaluation index of Root Mean Squared Error (RMSE), the imputation effects of each method were quantitatively evaluated. Finally, the data imputation method with the best performance was determined. Compared with the second-ranked method, this optimal method achieved a 1.2% improvement in performance. This improvement not only enhances the integrity and accuracy of the data, effectively reduces the interference of missing data on subsequent analysis, and improves the quality of the input data for the model, but also makes the model more robust when processing data, significantly improving the reliability and effectiveness of the model's predictions.

(3)The 40 features of the demographic and physiological information used in this paper cover demographic and physiological information. These features are diverse in dimensions and comprehensive in content, encompassing multiple aspects of the research subjects' lives and health. The rich and comprehensive data effectively prevents the model from overfitting due to its simplicity. It creates favorable conditions for the model to learn complex and realistic data distributions, ensuring that it has excellent generalization performance and can maintain stable prediction and analysis performance in different scenarios.

(4)In this paper, the RFA-CNN model is compared and analyzed with the other five benchmark models using the Clarke Error Grid Diagram. The results show that the RFA-CNN model has significant performance advantages. Compared with the second-ranked model, its performance has been improved by 4.97%. In the Clarke Error Grid Diagram, the proportion of sample points in region A of the RFA-CNN model is the highest, and the total proportion of sample points in regions A and B exceeds 99%. This result fully meets the strict clinical requirements. Since regions A and B represent the clinically acceptable error range, the high proportion of the RFA-CNN model in these regions indicates that its prediction results are highly reliable. It can provide accurate blood glucose prediction data for clinical decision-making. While reducing the risks of clinical misdiagnosis and missed diagnosis, it dramatically enhances the timeliness and effectiveness of medical interventions, providing strong support and guarantee for the formulation of clinical diagnosis and treatment plans and optimizing health management strategies.

II. Research Overview

A. Prediction Methods Based on Machine Learning

BG Choi (2019) developed a 5-year T2DM prediction model using 8,454 EMRs from Korea University Guro Hospital, comparing logistic regression (LR), LDA, QDA, and KNN via 10-fold cross-validation. The LR model achieved the highest AUC (0.78), though no significant performance gaps emerged among algorithms [12]. Meanwhile, Kun Lv (2023) leveraged over 60,000 Chinese NFG individuals' data, defining diabetes by HbA1c 48 mmol/mol. Logistic regression outperformed other ML methods, yielding the DRING model (13 features) with validated AUCs of 0.964 and 0.899 on independent datasets [13]. Key risk factors included BMI, age, and sex, supported by a clinical web tool. In noninvasive glucose monitoring, Meng Qi (2023) proposed near-infrared spectroscopy models: LS-DBN-SVR (99.9% correlation) and deep siamese network + SVR (MSE=0.024, correlation +0.49), incorporating label-sensitive feature selection [14]. Mi Peng (2023) developed a PPG-based system using SVM/pulse rescreening, XGBoost feature selection, and transfer learning to optimize a PSO-BP model, achieving RMSE=0.902 mmol/L with 88.89% A-zone accuracy [15]. For type 1 diabetes management, Daphne N. Katsarou (2025) integrated XGBoost and SVR on 29 GlucoseML patients' data, reducing hypoglycemia misclassification rates to enhance insulin precision and minimize short-term glucose volatility. Collectively, these studies advance ML-driven diabetes prediction and management through EMR analysis, noninvasive monitoring, and clinical tool development [16].

B. Prediction Methods Based on Deep Learning

Lin Zhongyan (2024) utilized single-layer perceptron (SLP) on ophthalmology patient data from a Chinese hospital, identifying significant correlations between albumin levels, red blood cell counts, and comorbid diabetes/hypertension—with male patients demonstrating higher risks [17]. He Yibo (2024) developed a 1DCNN-LSTM-Attention hybrid model using near-infrared technology, incorporating a rime ice optimization algorithm (RIME) for

parameter tuning. The architecture achieved MAE=0.121, MSE=0.0186, and correlation coefficient=0.9823 by leveraging 1DCNN for local feature extraction, LSTM for temporal dependencies, and attention mechanisms to emphasize critical information. These results outperformed traditional models, offering a novel solution for high-precision, non-invasive glucose detection [18]. Tang Baoyu (2024) constructed the InsuNet ensemble neural network model using discrete insulin injection records. The system achieved real-time prediction without continuous monitoring by fusing 13 architectural models and validating with heterogeneous datasets (RMSE= 14.78mg/dL). Ablation studies clarified feature importance, and a visualization system was developed to enhance clinical interpretability [19]. Zhao Hang (2024) optimized temporal feature extraction in deep learning models (TEM-Former, Gluformer) through dilated convolutions, periodic enhanced embedding layers, and multi-scale residual convolutions. Combined with Huber Loss for robustness, the model improved 30-minute prediction accuracy (RMSE reduced by 7%–15%, MAE decreased by 4%–16%) and validated clinical safety via Clarke error grid analysis (safe prediction proportion of 98.74%–99.38%) [20]. Xiao Zeqiu (2025) proposed the CNN-LSTM-attention deep learning model (PBI-CLA) using time-series blood glucose and insulin dose data. The model demonstrated significant improvements over traditional methods, with RMSE reduced by 12.82%, MAE decreased by 10.24%, and MAPE decreased by 10.24%, establishing a reliable framework for real-time clinical glucose management [21].

III. Algorithm Introduction

A. Feature Selection Method: Random Forest

Random Forest is an ensemble learning algorithm with multiple decision trees [22]. When performing feature selection, Random Forest can evaluate the importance of each feature. Its basic idea is to determine the importance of a feature by comparing the degree of change in the Random Forest model's prediction accuracy after randomly shuffling that feature's values. The calculation steps and formula for feature importance are as follows: Calculate the out-of-bag data error of each tree T_i (i=1,2,,n, where$n$ is the number of trees) as

$$OOB^i_{\text{error}} = \frac{1}{N_{\text{OOB}}} \sum_{j \in OOB} I\left(y_j \neq \hat{y}_{ij}\right). \tag{1}$$

Here, $NOOB$ is the number of out-of-bag samples, $I(\cdot)$ is the indicator function, y_i is the actual label, and $\hat{y}_{i,j}$ is the predicted value of the $j-th$ out-of-bag sample by the tree T_i. For each feature f, shuffle its values in the out-of-bag data and then recalculate the out-of-bag data error, which is denoted as OOB^i_{error} . The formula for the importance score for the feature f is

$$\text{importance } f = \frac{1}{n} \sum_{i=1}^{n} \left(OOB^i_{\text{error },f} - OOB^i_{\text{error}}\right) \tag{2}$$

B. Regression Model: CNN Model

The Convolutional Neural Network (CNN) is a type of deep learning model widely used in fields such as image recognition and predictive classification [23]. Its core components include convolutional layers, activation function layers, pooling layers, and fully connected layers, which can automatically extract features from data.

(1)Convolutional Layer: The convolutional layer is one of the core components of a Convolutional Neural Network (CNN). It performs a convolution operation on the input data through convolutional kernels to extract the features within the data. The one-dimensional convolutional layer conducts a convolution operation on a one-dimensional input sequence with the aid of convolutional kernels, thereby extracting the local features within the sequence.

Suppose the input sequence is a one-dimensional signal with the size of $X \in R^{L \times C_{in}}$, where L is the length of the sequence and C_{in} is the number of input channels. The size of the convolutional kernel is, here $K \in R^{L \times C_{in} \times C_{out}}$ is the length of the convolutional kernel, and C_{out} is the number of output channels. The convolutional operation will slide the convolutional kernel over the input sequence and, at each position, calculate the sum of the products of the elements of the convolutional kernel and the corresponding elements of the input sequence. The calculation formula of the output feature map $Y \in R^{L_{out} \times C_{out}}$is as follows:

$$Y_{i,j} = \sum_{m=0}^{k-1} \sum_{c=0}^{C_m - 1} X_{i+m,c} K_{m,c,j} \tag{3}$$

Among them, $Y_{i,j}$ is the value of the output feature map at the $i-th$ position and the $j-th$ channel; $X_{i+m,c}$ is the value of the input sequence at the $(i+m)-th$ position and the $c-th$ channel; $K_{m,c,j}$ is the value of the convolutional kernel at the $m-th$ position, the $c-th$ input channel, and the $j-th$ output channel; b_j is the bias of the $j-th$ output channel. The length L_{out} of the output sequence can be calculated by the following formula:

$$L_{out} = \left[\frac{L - k + 2 \times \text{ padding}}{\text{stride}}\right] + 1 \tag{4}$$

Here, padding is the size of the padding, and stride is the step size by which the convolutional kernel slides.

(2)Activation Function Layer: The activation function layer is a processing layer located after the convolutional layer. It performs a non-linear transformation on the linear results output by the convolutional layer, thereby enhancing the expressive ability of the CNN and enabling the network to learn complex data features. The activation function layer in the RFA-CNN model of this paper uses the SiLU function, and the mathematical expression of the SiLU function is:

$$f(x) = x \times \sigma(x) \tag{5}$$

Among them, $\sigma(x) = \frac{1}{1+e^{-x}}$ is the Sigmoid function. The SiLU function combines the characteristics of the linear and Sigmoid functions. It exhibits linear behavior when the input value is large and nonlinear behavior when the input value is small.

(3)Pooling Layer: The pooling layer is crucial in a Convolutional Neural Network (CNN) and is primarily used for downsampling the output of the convolutional layer. It reduces the data dimension and computational cost by performing specific operations (such as taking the maximum or average value) in local regions while enhancing the robustness of features. In this paper, three max-pooling layers

are adopted. The max-pooling layer is a type of downsampling operation layer used to reduce the dimension of the feature sequence output by the convolutional layer in a one-dimensional CNN. It achieves downsampling by selecting the maximum value within a local window of a specified length. Let the input feature sequence be $X \in R^{L\times C}$, where L is the sequence length and C is the number of channels. The size of the pooling window is p, and the stride is s. The output feature sequence is $Y \in R^{L_{out}\times C}$.Among them,$L_{out} = \left[\frac{L-p}{s}+1\right]$.For each channel c $\in \{1, 2, \cdots, \mathrm{C}\}$, the value of the output sequence at position $i \quad (i = 0, 1, \cdots, L_{out}-1)$, denoted as $Y_{i,c}$ is: $Y_{i,c} = \max_{j=0}^{p-1} X_{i\times s+j,c}$, That is, for each channel, the maximum value is selected from the p elements within each pooling window as the value of the corresponding position in the output sequence.

C. Missing Value Handling Method: Multiple Imputation

The multiple imputation method is a data preprocessing technique. It deals with the missing values in the dataset through multiple strategies, comprehensively using methods such as mean imputation, regression imputation, and multiple imputation to improve the integrity of the data. The multiple imputation method plays a crucial role in the data preprocessing stage. It can flexibly select various imputation methods according to different data missing situations and data characteristics, effectively reducing the adverse effects of data missing on subsequent data analysis and model construction.

D. Evaluation Indicators:MSE,MAE and EGA

This paper aims to predict blood glucose concentration levels in diabetic patients, which is inherently a multivariate regression task. Given the high-performance requirements for blood glucose concentration prediction and numerous outliers in the data samples, this paper selects Mean Squared Error (MSE) and Mean Absolute Error (MAE) as the primary metrics for evaluating model performance. The advantage of MSE lies in its squaring of errors, which sensitively reflects the degree of deviation between predicted and actual values. MAE's strength is in directly calculating the average of the absolute errors, offering relative robustness to outliers, and providing a more intuitive reflection of the actual average size of prediction errors. The combination of both metrics better reflects the overall performance of the model.

Additionally, the Clarke Error Grid Analysis (EGA) [29] is commonly used in the medical field to assess the accuracy of blood glucose measurement prediction models. The Clarke Error Grid sets the horizontal axis to the reference actual values and the vertical axis to the model's predicted values. Divided by multiple line segments, it is categorized into five regions. Region A is considered clinically accurate; Region B has a small error and is within clinically acceptable limits; Region C has a larger error; and Regions D and E has a significant error between the predicted and true values, which is not guidance for practical clinical application. Therefore, this paper also adopts the Clarke Error Grid as a metric for determining the accuracy of experimental results.

The calculation formula of the Mean Squared Error (MSE) is shown as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

The calculation formula of the Mean Absolute Error (MAE) is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

The true value is y_i , the predicted value is $\hat{y}_i$, and the number of samples is n .The smaller the values of Mean Squared Error (MSE) and Mean Absolute Error (MAE) are, the smaller the actual error of the model is, and the more accurate the prediction results will be.

E. Benchmark Model

In order to verify the prediction effect of the RFA-CNN model, this paper selects five representative models for comparative experiments:

DNN [25](Deep Neural Network): A computational model that achieves hierarchical feature learning through multiple layers of non-linear transformations, excelling at automatically extracting abstract representations from raw data and widely applied in fields such as image classification, speech recognition, and natural language processing.

ANN [26](Artificial Neural Network): A computational model constructed by imitating the information-processing mechanism of the human nervous system. It has powerful adaptive and self-learning characteristics, can model complex non-linear relationships, and can be widely applied in many fields, such as predictive analysis, pattern recognition, and data clustering.

MLP [27](Multilayer Perceptron): A network model based on a feed-forward architecture consisting of an input layer, multiple hidden layers, and an output layer. It processes neuron inputs through non-linear activation functions and can approximate any complex continuous function. It can solve various machine-learning tasks such as classification and regression.

LSTM [24](Long Short-Term Memory network): A recurrent neural network architecture that dynamically manages long-term and short-term information through input, forget, and output gates. It can effectively capture long-range dependencies in sequential data and is widely applied in natural language processing, speech recognition, and time series prediction tasks.

GRU [28](Gated Recurrent Unit): A simplified variant of recurrent neural networks that control the flow and forgetting of information through an update gate and a reset gate. It effectively alleviates the issues of gradient vanishing or exploding, efficiently learns long-range dependencies in sequential data, and is commonly used in tasks such as machine translation, speech recognition, and text generation.

F. SHAP Interpreter

SHAP (Shapley Additive exPlanations) is a unified model interpretation framework based on the principle of Shapley values in game theory. It aims to quantify the marginal contribution of each feature in a machine-learning model to the

final prediction result. Its core idea is to regard the prediction result as a payoff distribution problem in a multi-player cooperative game. A fair distribution of feature contributions is achieved by calculating the average marginal contribution value of a feature in all possible combinations. Unlike traditional local interpretation methods, SHAP supports simple algorithms such as linear models and tree-based models and can also effectively interpret complex black-box models like deep neural networks. It can capture both the interaction effects and non-linear relationships among features. This framework constructs a feature-dependent expected value decomposition, transforming the prediction results of complex models into an interpretable additive form. It provides a global feature importance ranking and enables local attribution analysis for individual samples. It has broad application value in medical diagnosis and financial risk control, which require highly reliable decision-making support. The balance between its mathematical rigor and computational efficiency makes it one of the most mainstream model interpretation tools currently. SHAP plots are a set of visualization-based interpretation tools derived from this framework, including various types such as summary plots, dependence plots, and force plots. By superimposing the feature value distribution and SHAP value scatter points, the summary plot intuitively shows the average impact direction and intensity of features on the prediction results and reveals the correlations among features simultaneously.

IV. Model Establishment

A. Data Source

The research data for this paper is sourced from the Alibaba Cloud AI Precision Medicine Competition project. The dataset comprises basic information, physical indicators, and blood glucose levels of the sample population, totaling 5642 samples. The feature variables include basic demographics of the sample population, such as gender, physical examination date, and age, as well as various physical indicators like hemoglobin, platelet count, white blood cell count, creatinine, total cholesterol, etc. The data label is the blood glucose level of the sample population.

B. Data preprocessing

Irrelevant variables in the original dataset, such as personal identifiers and physical examination dates, were removed. Features with missing values exceeding 70% of the total dataset were excluded, while the comparative analysis of data imputation methods will be detailed in the subsequent Data "Imputation Methods" section. Thirty-four original clinical indicators, including age, gender, alanine aminotransferase (ALT), etc., were selected as independent variables, with blood glucose levels as the dependent variable. To mitigate heteroscedasticity caused by varying measurement units across medical indicators, all variables were normalized to a [0,1] range using min-max scaling. The dataset was then partitioned into training and test sets at a 4:1 ratio. Data augmentation was applied to the training set by introducing random noise to expand the sample size. The processed data were subsequently fed into the model for analysis.

C. Comparison of Data Imputation Methods

This paper systematically evaluates five missing value imputation methods—mean imputation, random forest imputation, spline interpolation, KNN imputation, and multiple imputation—across six predictive models, with a total of 19,059 missing values imputed during experiments. The comparative results of glucose concentration prediction using different data completion approaches are visualized in Fig. 1.

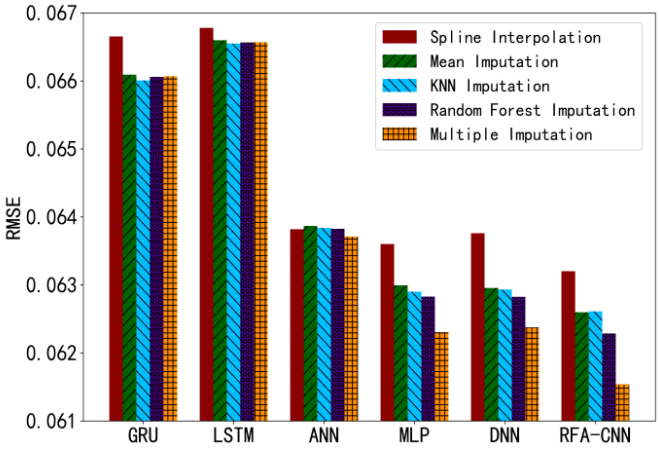

Fig. 1. Comparison Diagram of RMSE for Missing Value Imputation Methods

Specifically, the proposed RFA-CNN model achieved the lowest test set RMSE when combined with multiple imputations, leading to its adoption as the optimal method. Multiple imputation preserves original data uncertainty and statistical properties by generating and aggregating results from multiple plausible datasets. It outperforms single imputation methods by comprehensively capturing complex dependencies while avoiding information loss (mean imputation) or local bias (KNN imputation). In the RFA-CNN framework, Bayesian-based multiple imputation mitigates overfitting risks and enhances adaptability to high-dimensional nonlinear features through iterative refinement of imputed values.

D. Parameter Setting And Model Architecture

Hyperparameter optimization is a crucial aspect of machine learning and deep learning methods, directly affecting the model's performance and effectiveness. Commonly used hyperparameter optimization methods include grid search, random search, Bayesian optimization, etc. Grid search for hyperparameters can systematically traverse hyperparameter combinations, search for better configurations, and improve the model's performance. This paper employs the grid search method with five-fold cross-validation to tune the hyperparameters of the RFA-CNN model, thereby minimizing the model's prediction error. In the experiment, the Mean Squared Error (MSE) index is used to evaluate the hyperparameters' prediction effect, and the RFA-CNN model's optimal hyperparameters are obtained, as shown in TABLE I.

After using the grid search with five-fold cross-validation for hyperparameter tuning, the researchers defined the CNN-based neural network model CNNModel based on these

TABLE I
HYPERPARAMETER SETTING TABLE

Hyperparameters	Optimal Values	Parameter Ranges
Conv_layers_number	3	[1,2,3,4,5]
kernel_size	3	[3,5,7]
Pool_layers_number	3	[1,2,3]
Pooling_kernel_size	2	[1,2,3]
Dropout Rate	0.3	[0.1,0.2,0.3]
Learning Rate	0.0001	[0.001,0.0001]
Weight Decay	0.001	[0.001,0.0001]
num_epochs	500	[300,500,1000]
Optimizer	Adam	[Adam,SGD]

TABLE II
PARAMETER QUANTITIES PER LAYER OF RFA - CNN MODEL

Layer Name	Output Dimension	Parameters
Conv1d - 1	[-1, 64, 31]	256
BatchNorm1d - 2	[-1, 64, 31]	128
SiLU - 3	[-1, 64, 31]	0
MaxPool1d - 4	[-1, 64, 15]	0
Conv1d - 5	[-1, 128, 15]	24704
BatchNorm1d - 6	[-1, 128, 15]	256
SiLU - 7	[-1, 128, 15]	0
MaxPool1d - 8	[-1, 128, 7]	0
Conv1d - 9	[-1, 256, 7]	98560
BatchNorm1d - 10	[-1, 256, 7]	512
SiLU - 11	[-1, 256, 7]	0
MaxPool1d - 12	[-1, 256, 3]	0
Linear - 13	[-1, 512]	393728
SiLU - 14	[-1, 512]	0
Dropout - 15	[-1, 512]	0
Linear - 16	[-1, 256]	131328
SiLU - 17	[-1, 256]	0
Linear - 18	[-1, 1]	257

parameters. This enables the model's structural and parameter configuration to better conform to the characteristics of the data, allowing an appropriate number of convolutional kernels to extract features more effectively. The architecture of the RFA-CNN model is shown in the following figure. Subsequently, a regressor CNNTorchRegressor compatible with sci-kit-learn was customized. The model was trained using the Mean Squared Error (MSE) loss function and the Adam optimizer in the fit method. Moreover, the learning rate was dynamically adjusted through a learning rate scheduler, and an early stopping strategy was adopted to prevent overfitting.

As can be seen from TABLE II, the model structure starts with an input layer that receives data from 31 variables. The first convolutional layer, Conv1, has a convolutional kernel parameter of 1×64×3, where 1 represents the number of input channels, 64 represents the number of output channels, and 3 represents the convolutional kernel size. Immediately following are a BatchNorm1d layer for batch normalization, a SiLU activation function layer for introducing non-linearity, and a MaxPool1d layer with a pooling kernel size of 2 for downsampling. The second convolutional module is composed of Conv2, with a convolutional kernel parameter of 64×128×3 (where 64 is the number of input channels, 128 is the number of output channels, and 3 is the size of the convolutional kernel). Subsequently, a BatchNorm1d layer, a SiLU layer, and a MaxPool1d layer (with a pooling kernel size of 2) are connected sequentially, just like in the previous module. The Conv3 in the third convolutional module has a convolutional kernel parameter of 128×256×3 (where 128 is the number of input channels, 256 is the number of output channels, and 3 is the size of the convolutional kernel). It is also followed by the corresponding BatchNorm1d layer, SiLU layer, and MaxPool1d layer (with a pooling kernel size of 2). After the convolutional and pooling operations, the data enters the fully connected layer part. The Linear1 layer has an input dimension of 256 and an output dimension of 512. Then, there is a SiLU activation function layer and a Dropout layer to prevent overfitting. The subsequent Linear2 layer inputs 512 dimensions and outputs 256 dimensions, followed by another SiLU layer. Finally, the Linear3 layer inputs 256 dimensions and outputs one dimension, generating the final output result.

E. Variable Selection

This study conducts feature selection to reduce data dimensionality, prevent overfitting, and improve the computational efficiency, generalization ability, and interpretability of the model. The aim is to identify the number of features corresponding to the lowest Mean Squared Error (MSE) on the test set and then apply this optimal number of features to the optimal model. Given that the proposed RFA-CNN model has three pooling layers and considering that an insufficient number of features may lead to the omission of crucial information, feature selection starts when the number of features is eight. This study employs the Random Forest (RF) algorithm for feature selection. Specifically, it trains the model using RandomForestRegressor, obtains the importance scores of each feature through the feature_importances_ attribute, and finally uses the np. argsort function to rank the features in descending order of importance. Subsequently, the top eight most important features are fed into the model for calculation, initiating an automated search for the optimal number of features.

As shown in Fig. 2, the top 15 important features include Triglyceride, Age, Uric Acid, *Aspartate Aminotransferase, *Alanine Aminotransferase, *Alkaline Phosphatase, *Gamma-Glutamyl Transferase, Urea, Mean Corpuscular Volume, White Blood Cell Count, High - Density Lipoprotein Cholesterol, Total Cholesterol, Red Blood Cell Count, Red Cell Distribution Width, and Albumin. Triglyceride has the highest importance score, indicating its dominant role in influencing the model's output, which may imply a strong intrinsic correlation with the target variable. Age follows closely, suggesting its non-negotiable impact on the model's prediction, perhaps reflecting age-related patterns in the data. Uric Acid also holds a relatively prominent position. Various aminotransferases and Urea highlight that these biochemical indicators are vital for the model, likely being closely associated with the underlying mechanisms or patterns the model aims to capture.

The method for determining the screening threshold (i.e., the optimal number of features) is as follows: iterate through

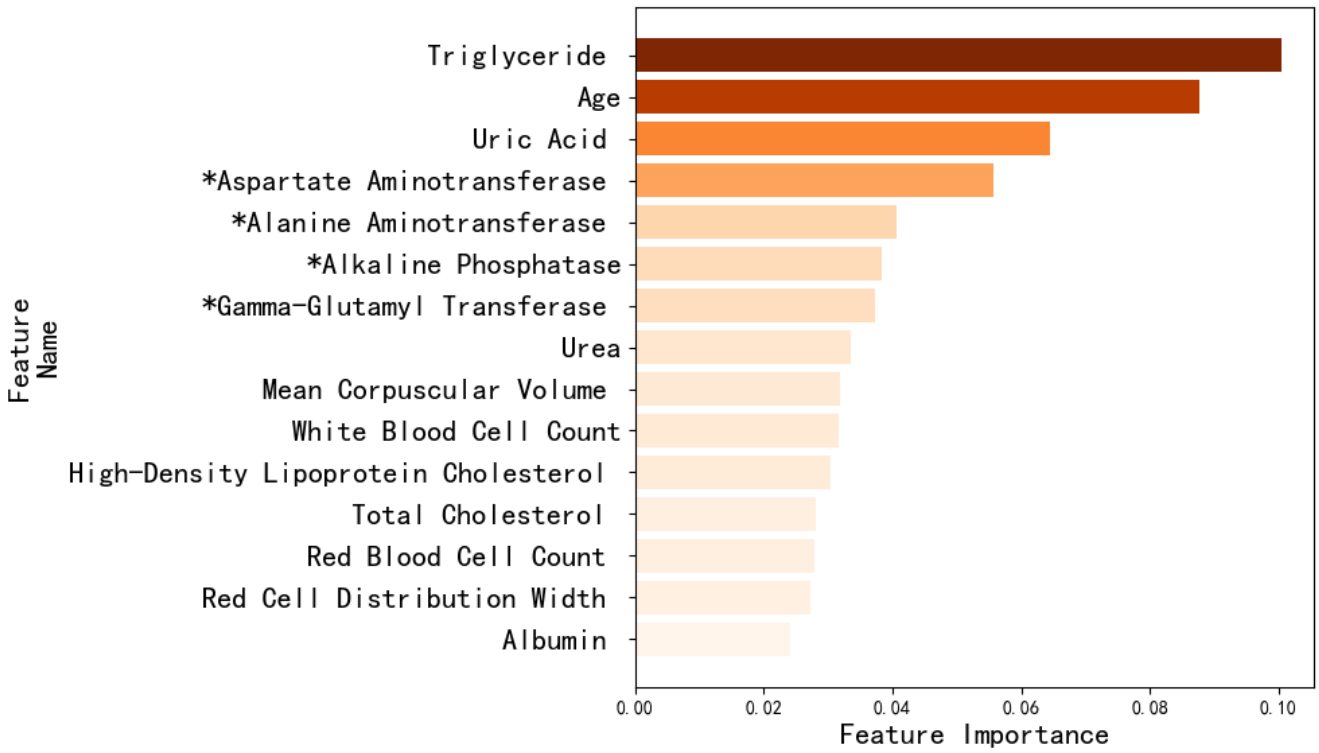

Fig. 2. Top 15 Features by Importance

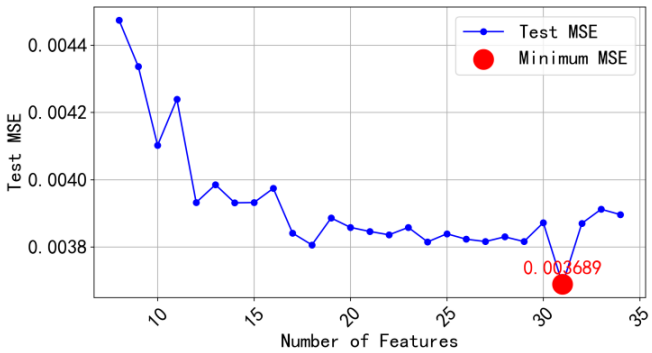

Fig. 3. The Result Graph of Feature Selection

different numbers of features, from eight to the total number of features. Each time, select the corresponding number of important features to train the model and make predictions on the test set. Calculate the Mean Squared Error (MSE) of the test set and take the number of features corresponding to the minimum MSE as the optimal number.

Fig. 3. indicates that starting from eight features, the MSE gradually decreases and reaches its lowest point when the number of selected features is 31. Subsequently, the MSE begins to increase. Therefore, the optimal number of features is 31. If feature selection is not performed and all variables are directly input into the model for calculation, the test set's Mean Squared Error (MSE) is 0.003787. Therefore, after feature selection, the MSE of the model is reduced by 2.59%.

V. Model Performance In-Depth Analysis

A. Insights from Prediction Error Scatter Plot

Fig. 4. visually depicts the distribution of prediction errors from the RFA-CNN model. The horizontal axis represents the sample point numbering after data augmentation, while the vertical axis indicates the prediction error values. Most prediction errors are clustered near zero, demonstrating a high degree of consistency between predicted and actual values, reflecting the model's high accuracy and reliability in output. Furthermore, most errors fall within the range of [-0.1, 0.1], highlighting the precision of the model's predictions. Nevertheless, some outliers are present in the dataset, particularly noticeable within certain intervals along the horizontal axis. These outliers represent cases where prediction errors deviate significantly from the expected range, potentially attributable to noise introduced during data augmentation by researchers. However, their limited quantity has only a marginal impact on the model's overall performance.

B. Learning Curve Analysis

The learning curve depicted in Fig. 5. illustrates the convergence behavior of the RFA-CNN model during its training phase. The x-axis represents the number of epochs, while the y-axis indicates the loss value measured by mean squared error (MSE). Initially, the training loss (yellow line) and test loss (purple line) exhibit a sharp decline trend within the first few epochs. This rapid reduction suggests that the model quickly learns from the training data and adjusts its parameters to minimize errors. As training progresses beyond approximately 50 epochs, both curves tend to stabilize,

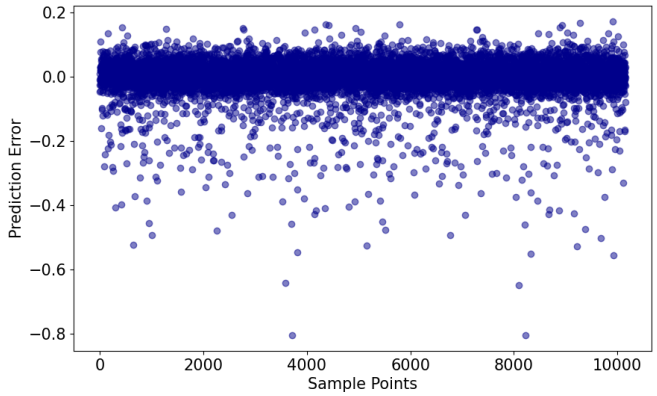

Fig. 4. Prediction Error Scatter Plot of RFA - CNN Model

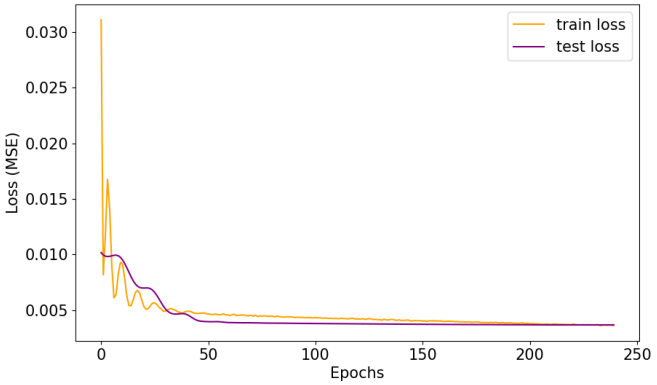

Fig. 5. Learning Curve of RFA - CNN Model

although occasional minor fluctuations or rebounds may occur. This stability implies that the model has reached a relatively mature stage where further training might not yield significant improvements and could even introduce noise due to diminishing returns. In summary, this learning curve demonstrates favorable performance characteristics of the RFA-CNN model throughout the training process, including efficient error reduction across all training phases, minimal overfitting, and ultimate stabilization, indicating optimal parameter settings. These attributes collectively underscore the effectiveness of the chosen architecture and hyperparameter tuning processes applied for specific tasks.

C. Model Ablation Analysis

This paper adopts a module-level ablation experiment method to verify the contribution of key modules to the prediction performance of the RFA-CNN model. The components of the convolutional block, which consists of a convolutional layer (Conv), a batch normalization layer (BN), a SiLU activation function, a pooling layer (Pool), and the Dropout regularization layer, are removed separately for comparative analysis.

The results presented in TABLE III above show that when the convolutional blocks are ablated step-by-step, the model's performance deteriorates notably. Removing Block 1 leads to around a 1.82% drop in performance. When Blocks 1 and 2 are ablated, the performance decline is approximately 4.74%. If Blocks 1, 2, and 3 are further ablated, the MSE of the test set skyrockets to 0.005156, and the performance drops by a significant 40.85%. This shows the core benefits

TABLE III
MODEL ABLATION ANALYSIS TABLE

Models	Evaluation Indicators	Indicator Values
RFA - CNN	Training MSE	0.003547
	Test MSE	0.003689
RFA - CNN Model w/o Block 1	Training MSE	0.004052
	Test MSE	0.003756
RFA - CNN Model w/o Blocks 1 & 2	Training MSE	0.004422
	Test MSE	0.003864
RFA - CNN Model w/o Blocks 1 & 2 & 3	Training MSE	0.005904
	Test MSE	0.005156
RFA - CNN Model w/o Dropout Layers	Training MSE	0.003650
	Test MSE	0.003701

of the convolutional block. It can efficiently extract important features through the joint work of multiple components. The convolutional layer is responsible for feature extraction, the batch normalization layer stabilizes the training process, the SiLU activation function boosts non-linear expression, and the pooling layer compresses the feature dimensions. This ensures the integrity and stability of the model's feature learning.

The ablation experiment on the Dropout layer reveals that when this layer is removed, the test MSE rises to 0.003701, and the performance decreases by about 0.325%. This indicates that the Dropout layer inhibits overfitting by randomly deactivating neurons. It improves the model's generalization ability on the test data and keeps the training stable. Although its small effect is still a vital component for ensuring the model's performance.

D. Model Interpretation with SHAP

Researchers employed the SHAP interpreter for model interpretation to further explain the RFA-CNN model in our study. They selected the first 100 data points from the training set as background data to form a SHAP interpreter and chose the top ten variables regarding SHAP values among these 100 data points for explanation. According to Fig. 6. below, this SHAP summary plot vividly demonstrates the impact of various features on the model output. Each dot in the figure represents a sample. Its position on the horizontal axis indicates the SHAP value (the impact on the model output), and the color reflects the feature value (red for high and blue for low). For example, "Age" shows that as its feature value increases (red dots), the SHAP value tends to be positive, indicating that the older the age, the higher the blood glucose level tends to be. Similarly, "High-density Lipoprotein Cholesterol" and "Albumin" also exhibit a correlation between higher feature values (red dots) and positive SHAP values, meaning they significantly contribute to increasing the model output. Notably, for "Age," "High-Density Lipoprotein Cholesterol," and "Albumin," most of the dots are located on the left side of the axis, which suggests that the blood glucose levels of most people are relatively low. In addition, features such as "Neutrophil Percentage" show a more complex distribution with mixed colors, indicating diverse impacts on the model output.

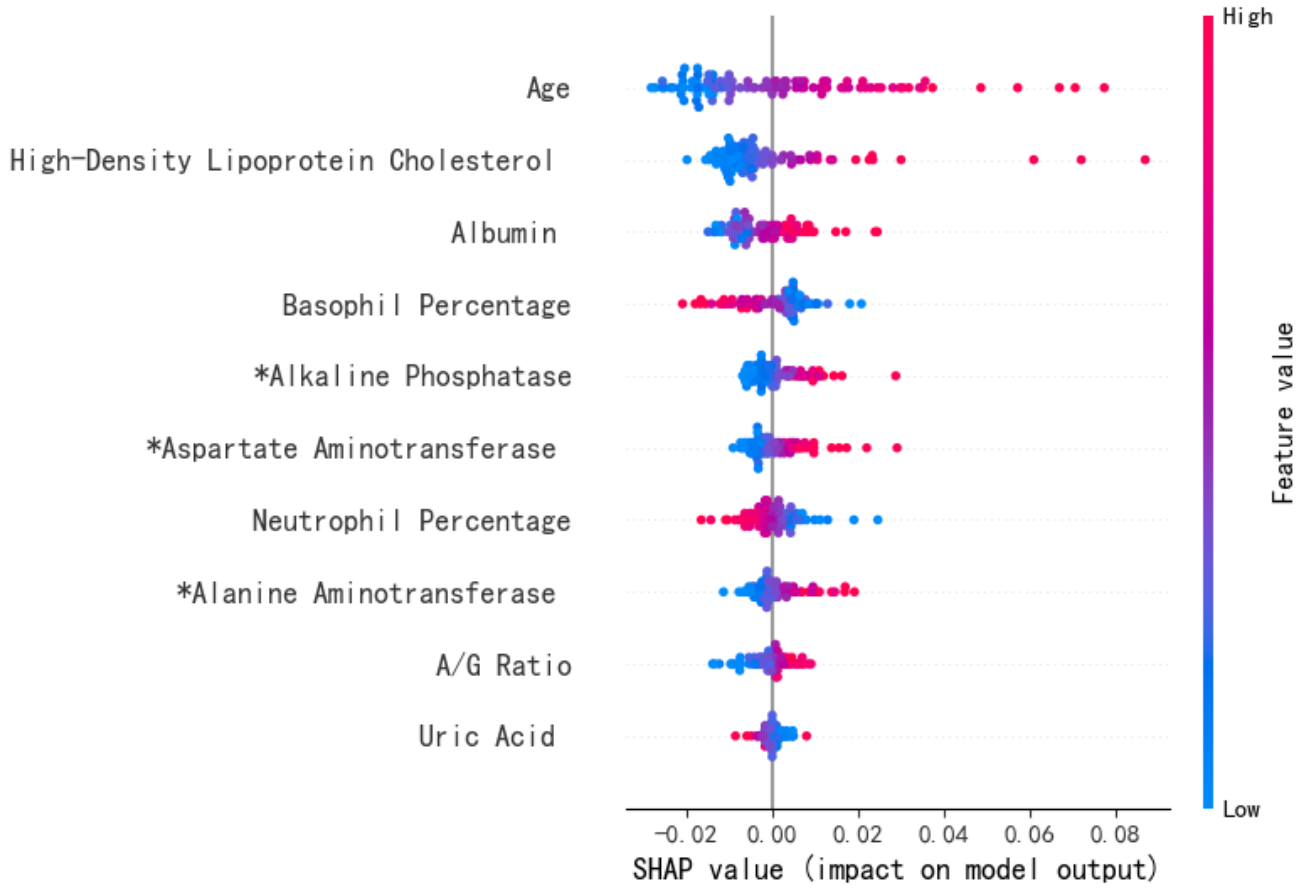

Fig. 6. SHAP Value Summary Plot for Key Features

VI. Comparative Experiment Analysis

A. Model-wise Comparison of MSE, MAE and MSE*

In this study, blood glucose concentration predictions were conducted using the RFA-CNN, DNN, ANN, MLP, LSTM, and GRU models. For these six models, the significance test results of pairwise comparisons between the RFA-CNN model and each benchmark model based on independent sample t-tests are shown in Table IV below, and the indicators of Mean Squared Error (MSE), Mean Absolute Error (MAE), and MSE* (the difference between the training set MSE and the test set MSE) for different prediction methods are presented in Table V below.

TABLE IV
Pairwise Comparison of Models: p-values for MSE and MAE

Model 1	Model 2	p-value-MSE	p-value-MAE
RFA-CNN	DNN	1.70×10^{-9}	2.55×10^{-8}
RFA-CNN	ANN	2.08×10^{-9}	5.12×10^{-8}
RFA-CNN	GRU	3.49×10^{-12}	1.23×10^{-11}
RFA-CNN	LSTM	7.77×10^{-5}	3.93×10^{-5}
RFA-CNN	MLP	4.03×10^{-7}	6.04×10^{-6}

TABLE V
Comparison Table of Six Model Indicator

Models	Evaluation Indicators	Indicator Values
RFA - CNN	MSE	0.003689
	MAE	0.035084
	MSE*	0.000142
DNN	MSE	0.003891
	MAE	0.035121
	MSE*	0.000452
ANN	MSE	0.004059
	MAE	0.034705
	MSE*	0.000662
MLP	MSE	0.003882
	MAE	0.035867
	MSE*	0.000341
LSTM	MSE	0.004429
	MAE	0.03527
	MSE*	0.000809
GRU	MSE	0.004357
	MAE	0.037123
	MSE*	0.000752

The two tables clearly show that the p-values of the RFA-CNN model compared with the other five models on both MSE and MAE metrics are all less than 0.001, indicating that their prediction performances have extremely significant statistical differences. This suggests that the improvement in blood glucose prediction accuracy of the RFA-CNN model is not accidental but a substantial enhancement brought by the model structure design.

The Mean Squared Error (MSE) of the RFA-CNN model is significantly lower than that of the other five models.

Notably, the MSE of RFA-CNN is 4.97% lower than that of the MLP model (the second-best in MSE), indicating its higher overall prediction accuracy. Since MSE is sensitive to squared error terms, the lower MSE value of the RFA-CNN model demonstrates its higher efficiency in handling the deviation between predicted and actual values, especially in reducing the impact of large-error samples and enhancing the overall data fitting effect.

Regarding the Mean Absolute Error (MAE) metric, although the MAE of RFA-CNN is slightly higher than that of the ANN model, it is still lower than those of the other four models. This indicates that the ANN model has certain advantages in reducing the absolute error of specific samples. At the same time, RFA-CNN exhibits more stable performance in balancing the absolute errors across the entire sample set, providing more balanced and stable prediction results than most comparative models. Additionally, RFA-CNN has the lowest MSE* value (the difference between the training set MSE and the test set MSE), fully demonstrating the minor error difference between the training and test sets, lower overfitting risk, and superior generalization capability to unknown data, thus maintaining stable prediction performance across different data distributions. By integrating CNN at appropriate positions to capture the nonlinear characteristics of blood glucose data and combining random forests to enhance feature robustness, the RFA-CNN model further improves its prediction accuracy and reliability.

B. Model - wise Comparison of Clarke Error Grid Plot

To evaluate the performance of the models on the training set, we presented the results of the test set in the form of a Clarke Error Grid plot. The table compares the probabilities of six models falling into Zone A and Zone B. Specifically, the probability of the RFA-CNN model falling into Zone A is 83.87%, and the probability of it falling into Zone B is 15.41%. The RFA-CNN model has the highest probability of falling into Zone A, surpassing the ANN model (ranked second) and the LSTM model (ranked third). The GRU model has the lowest probability of falling into Zone A compared to the other models. The detailed data are shown in Table VI.

Fig. 7. illustrates that the prediction data points of the RFA-CNN model are densely clustered in Clinical Error Grid Area A, with a remarkable proportion of 83.9%, substantially higher than that of the other five models. In particular, the combined proportion of sample points in Area A and Area B for the RFA-CNN model exceeds 99%, with a relatively high percentage in Area A. This indicates the excellent precision and reliability of the model [31]. This concentration signifies that the deviation of the predicted glucose concentrations from the reference concentrations is negligible, with the overwhelming majority of outcomes falling within the clinically entirely acceptable range. In diabetes management, a high percentage of data points in Area A is crucial as it implies that the model's predictions are reliable and safe for clinical decision-making, reducing the risk of incorrect treatment actions.

In contrast, while the DNN, ANN, and MLP models have many data points in Area A, their greater scatter and tendency to diffuse towards Area B suggest less precision in predictions. The LSTM and GRU models, on the other hand, do not exhibit sufficient clustering in Area A, with some predictions significantly underestimating the actual values, leading to a higher incidence of sample points moving into Area B, which can be as high as 16.4% and 18.8% respectively. The high concentration of data points in Area A by the RFA-CNN model is therefore of significant importance in diabetes management, as it ensures that glucose predictions are aligned closely with actual values, thereby enhancing patient safety and the effectiveness of therapeutic interventions.

TABLE VI
COMPARISON OF PROBABILITY OF FALLING INTO DIFFERENT ZONES FOR SIX MODELS

Model	Zone	Probability of Falling into this Zone
RFA - CNN	A	83.87%
	B	15.41%
DNN	A	81.84%
	B	17.44%
ANN	A	83.52%
	B	15.76%
GRU	A	80.42%
	B	18.86%
LSTM	A	82.99%
	B	16.29%
MLP	A	81.48%
	B	17.80%

VII. CONCLUSION

Through the application of the RFA-CNN model for blood glucose value prediction, this paper arrives at the following four conclusions:

(1)The model demonstrates remarkable performance advantages. The RFA-CNN model proposed in this study innovatively combines Random Forest feature selection and SHAP interpretability, addressing prediction accuracy and clinical transparency. By quantifying feature contributions via SHAP, the framework reveals the nonlinear impacts of age, HDL-C, and albumin on blood glucose levels, making it a robust tool for data-driven clinical decision support. It exhibits outstanding performance in blood glucose concentration prediction compared to five other benchmark models. The Mean Squared Error (MSE) of its test set reaches 0.003689, which is a 4.97% reduction compared to the second-ranked Multi-Layer Perceptron (MLP) model. Moreover, its test set's Mean Absolute Error (MAE) is 0.035084, remaining at a relatively low level. In the Clarke Error Grid analysis, as high as 83.9% of the predicted values fall within the clinically accurate Region A, and the proportion of Region A+B exceeds 99%, fully meeting the strict clinical application standards. This indicates that the model can accurately capture the nonlinear patterns of blood glucose fluctuations, providing a reliable basis for clinical decision-making.

(2)The data preprocessing yields favorable results. The multiple imputation method enhances data quality: During

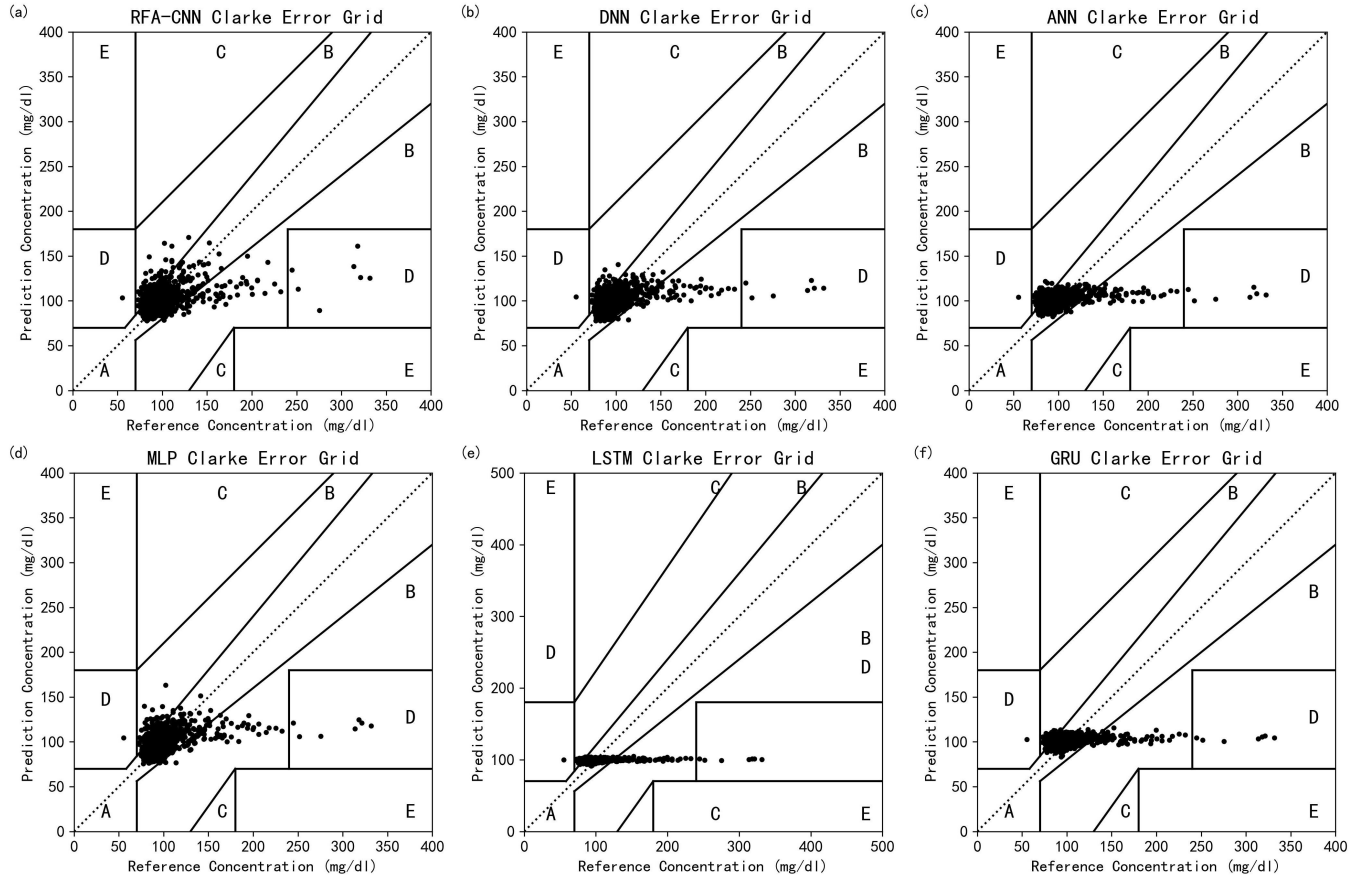

Fig. 7. Comparison of Clarke Error Grid Diagrams of Six Models (a) Clarke error grid for RFA-CNN model. (b) Clarke error grid for DNN model. (c) Clarke error grid for ANN model. (d) Clarke error grid for MLP model. (e) Clarke error grid for LSTM model. (f) Clarke error grid for GRU model.

the data preprocessing stage, through a systematic comparison of five data imputation methods, it is found that the multiple imputation method performs the best. Compared with the suboptimal K-Nearest Neighbors (KNN) imputation method, it reduces the Root Mean Squared Error (RMSE) by 1.2%. This method retains the uncertainty and statistical characteristics of the original data by generating and aggregating the results of multiple reasonable datasets, effectively reducing the interference of missing values on the model, improving the integrity and accuracy of the data, and laying a solid foundation for subsequent analysis.

(3)The feature selection achieves excellent outcomes. Feature selection by Random Forest optimizes the model: By leveraging the Random Forest algorithm for feature selection, a combination of 31 optimal features is determined. Compared with the input of all features, the MSE is reduced by 2.59%. This process reduces redundant information and retains key physiological indicators such as age and hemoglobin, enhancing the model's generalization ability and enabling the model to more efficiently extract useful information from high-dimensional data.

(4)The model architecture is rationally designed. The RFA-CNN model organically combines the Random Forest (RF) and the Convolutional Neural Network (CNN), giving full play to both advantages. The optimal hyperparameters of the model are obtained through grid search with five-fold cross-validation, resulting in better model performance. The local patterns are extracted through a three-layer convolutional and pooling structure. By integrating with the feature importance evaluation of the Random Forest, the prediction accuracy and robustness are significantly improved. Ablation experiments show that removing any convolutional block or dropout layer will substantially decline model performance, verifying the crucial roles of convolutional blocks and dropout layers in the model architecture.

REFERENCES

[1] Y. Zhang, Y. Xu, and J. Zhai, "Diabetic Retinopathy Image Segmentation Method Based on Fusion DenseNet and U-Net Network," *Engineering Letters*, vol. 33, no. 2, pp. 418-128, 2025.

[2] M. Khalifa and M. Albadawy, "Artificial intelligence for diabetes: enhancing prevention, diagnosis, and effective management," *Computer Methods and Programs in Biomedicine Update*, vol. 5, pp. 100-141, 2024.

[3] Md J. Hossain, Md Al-Mamun, and Md R. Islam, "Diabetes mellitus, the fastest growing global public health concern: Early detection should be focused," *Health Science Reports*, vol. 7, no. 3, pp. e2004, 2024.

[4] D. Magliano and EJ. Boyko, *IDF diabetes atlas. 10th edition*, International Diabetes Federation, 2021.

[5] W. Zheng, S. Fan, G. Wu, et al., "Network Meta-analysis of the Efficacy and Safety of Classical Prescriptions in the Treatment of Diabetic Nephropathy," *World Chinese Medicine*, vol. 19, no. 19, pp. 2942-2957, 2024.

[6] Y. PENG, N. XIE, M. JING, et al., "Development of a Blood Glucose Benchmark Report Based on the Information Glucose Monitoring System," *Chinese General Practice*, vol. 24, no. 33, pp. 4255-4260, 2021.

[7] H. Sun, P. Saeedi, S. Karuranga, et al., "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes research and clinical practice*, vol. 183, pp. 109-119, 2022.

[8] J. Fu, P. Yang, X. Xiao, et al., "Correlation Analysis between Platelet/Lymphocyte Ratio and the Risk of Cardiovascular Disease Mortality in Elderly Diabetic Patients," *Modern Preventive Medicine*, vol. 52, no. 05, pp. 949-954, 2025.
[9] X. Chen and L. Sun, "Several Issues Needing Attention in the Diagnosis and Treatment of Diabetic Kidney Disease," *Chinese Journal of Practical Internal Medicine*, vol. 45, no. 03, pp. 199-203, 209, 2025.
[10] L. Zeng, "Pay Attention to the Prevention of Hypoglycemia and Its Recurrent Episodes," *Chinese Journal of Diabetes*, vol. 31, no. 11, pp. 877-880, 2023.
[11] Chinese Hypertension Prevention and Treatment Guidelines Revision Committee, Hypertension League (China), Hypertension Branch of China International Exchange and Promotion Association for Medical and Healthcare, et al., "Chinese Guidelines for the Prevention and Treatment of Hypertension (2024 Revised Edition)," *Chinese Journal of Hypertension (Chinese and English)*, vol. 32, no. 07, pp. 603-700, 2024.
[12] B. G. Choi, et al., "Machine learning for the prediction of new-onset diabetes mellitus during 5-year follow-up in non-diabetic patients with cardiovascular risks," *Yonsei medical journal*, vol. 60, no. 2, pp. 191-199, 2019.
[13] K. Lv, et al., "Detection of diabetic patients in people with normal fasting glucose using machine learning," *BMC medicine*, vol. 21, no. 1, pp. 342, 2023.
[14] Q. Meng, "Research on Non - invasive Blood Glucose Concentration Prediction Algorithm Based on Deep Learning," [D], Changchun University of Science and Technology, 2023.
[15] P. Mi, "Research on Non-invasive Blood Glucose Detection System Based on PPG," [D], Chongqing University, 2023.
[16] N. D. Katsarou, I. E. Georga, A. M. Christou, et al., "Optimizing hypoglycaemia prediction in type 1 diabetes with Ensemble Machine Learning modeling," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, pp. 46-46, 2025.
[17] Z. Lin, Y. Zhu, Y. Zhang, et al., "Risk Assessment of Chronic Diseases in Ophthalmic Patients Based on a Single - Layer Perceptron Prediction Model: Taking Diabetes and Hypertension as Examples," *Chinese Journal of Health Statistics*, vol. 41, no. 06, pp. 889-892, 2024.
[18] Y. He, H. Jin, C. Zhou, et al., "Research on Non - invasive Blood Glucose Prediction Model Based on RIME and 1DCNN - LSTM - Attention," *Modern Electronics Technique*, vol. 47, no. 18, pp. 83-88, 2024.
[19] B. Tang, "Research and Implementation of an Insulin Efficacy Prediction Algorithm Based on an Ensemble Neural Network," [D], Beijing University of Posts and Telecommunications, 2024.
[20] H. Zhao, "Research on Blood Glucose Prediction for Diabetes Based on Informer," [D], Jilin University, 2024.
[21] Z. Xiao, Y. Li, X. Wang, "Blood Glucose Concentration Prediction for Diabetic Patients Based on the PBI-CLA Model," [J/OL], *Computer Engineering*, pp. 1 - 11 [Accessed 2025-03-20].
[22] Z. Shao, Y. Leng, D. Ge, et al., "Radar Active Jamming Recognition Based on Wavelet Packet Decomposition and Random Forest Feature Selection," *Modern Radar*, vol. 46, no. 12, pp. 45-52, 2024.
[23] Y. Huang, et al., "An Improved Hybrid CNN-LSTM-Attention Model with Kepler Optimization Algorithm for Wind Speed Prediction," *Engineering Letters*, vol. 32, no. 10, pp. 1957-1965, 2024.
[24] F. Zhu, et al., "Design of Digital FIR Filter Based on Long Short-Term Memory Neural Network," *Engineering Letters*, vol. 32, no. 10, pp. 1993-2001, 2024.
[25] Y. Wang, et al., "Adascale: Dynamic context - aware dnn scaling via automated adaptation loop on mobile devices," *IEEE Internet of Things Journal*, 2025.
[26] W. Li, et al., "Optimisation of PCM passive cooling efficiency on lithium-ion batteries based on coupled CFD and ANN techniques," *Applied Thermal Engineering*, vol. 259, pp. 124874, 2025.
[27] T. Bashir, et al., "Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models," *Renewable Energy*, vol. 239, pp. 122055, 2025.
[28] U. Kosarkar and G. Sakarkar, "Design an efficient VARMA LSTM GRU model for identification of deep-fake images via dynamic window-based spatio-temporal analysis," *Multimedia Tools and Applications*, vol. 84, no. 7, pp. 3841-3857, 2025.
[29] D. N. Katsarou, et al., "Optimizing hypoglycaemia prediction in type 1 diabetes with Ensemble Machine Learning modeling," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, pp. 46, 2025.
[30] Y. Takefuji, "Beyond XGBoost and SHAP: Unveiling true feature importance," *Journal of Hazardous Materials*, pp. 137382, 2025.
[31] International Organization for Standardization, "ISO 15197:2013 In - vitro Diagnostic Test Systems-Requirements for Blood-Glucose Monitoring Systems for Self - Testing in Managing Diabetes Mellitus," 2013.