

PPG Based Noninvasive Blood Glucose Monitoring Using Multi-View Attention and Cascaded BiLSTM Hierarchical Feature Fusion Approach

Mubashir Ali , *Student Member, IEEE*, Jingzhen Li , *Member, IEEE*, Bokun Fan, and Zedong Nie , *Senior Member, IEEE*

Abstract—Diabetes is a chronic disease with exponential growth and poses significant challenges to global healthcare. Regular blood glucose (BG) monitoring is key for avoiding diabetic complications. Traditional BG measurement techniques are invasive and minimally invasive, causing pain, discomfort, cost, and infection risks. To address these issues, we developed a noninvasive BG monitoring approach on photoplethysmography (PPG) signals using multi-view attention and cascaded BiLSTM hierarchical feature fusion approach. Firstly, we implemented a convolutional multi-view attention block to extract the temporal features through adaptive contextual information aggregation. Secondly, we built a cascaded BiLSTM network to efficiently extract the fine-grained features through bidirectional learning. Finally, we developed a hierarchical feature fusion with bilinear polling through cross-layer interaction to obtain higher-order features for BG monitoring. For validation, we conducted comprehensive experimentation on up to 6 days of PPG and BG data from 21 participants. The proposed approach showed competitive results compared to existing approaches by RMSE of 1.67 mmol/L and MARD of 17.88%. Additionally, the clinical accuracy using Clarke error grid (CEG) analysis showed 98.80% of BG values in

Zone A+B. Therefore, the proposed approach offers a favorable solution in diabetes management by noninvasively monitoring the BG levels.

Index Terms—Cascaded bilstm, hierarchical feature fusion noninvasive blood glucose monitoring, multi-view attention, photoplethysmography, ppg.

I. INTRODUCTION

DIABETES is a chronic disease that arises when the human body fails to produce or regulate insulin which is essential to manage blood glucose (BG) levels [1]. The prevalence of diabetes is very alarming and impacted nearly 536 million adults by 2021, and is projected to be 783.2 million adults by 2045 [2]. Consistent lower BG levels (hypoglycemia) and higher BG levels (hyperglycemia) lead to microvascular (kidney, eye, neuropathy) and macrovascular (brain, foot, cardiovascular) complications [3], [4]. These complications caused 6.7 million deaths in 2021 [5]. Diabetes cannot be cured but continuous BG monitoring plays a crucial role in efficient diabetes management by avoiding life-threatening complications. A finger prick test using commercial glucometers is a widely adopted method for BG measurement. This test is invasive, painful, costly, and has critical infection risks [6]. The recent development of continuous glucose monitoring (CGM) devices offers real-time BG monitoring with a minimally invasive nature [7]. The CGM sensor is inserted into the subcutaneous tissue of humans and works for up to 14 days for BG monitoring. The regular replacement of sensors causes pain, discomfort, and medical expenses making it a hard choice for many patients. Hence, there is an urgent need for an affordable and noninvasive BG monitoring method for effective diabetes management. Such a method is life-changing for people living in developing countries and unable to access healthcare services [8].

PPG is a noninvasive technology to monitor the change in blood volume through the microvascular bed of tissues during a cardiac cycle [9], [10]. It consists of a photo-detector (PD) and light-emitting diode (LED) that emits the light which is transmitted or reflected through the microvascular tissues and detected by PD. PPG is primarily associated with measuring blood flow dynamics, which are influenced by various physiological parameters. Several physiological principles link the BG

Received 6 April 2024; revised 20 July 2024; accepted 15 September 2024. Date of publication 18 September 2024; date of current version 4 July 2025. This work was supported in part by the National Key R&D Program of China under Grant 2022YFB3203702, in part by the National Natural Science Foundation of China under Grant 62173318, in part by the Science and Technology Service Network Plan of CAS - Huangpu Special Project under Grant STS-HP-202203, in part by the CAS Key Laboratory of Health Informatics, and in part by the University Chinese Academy of Sciences and Alliance of International Science Organization (ANSO) under Grant 2021A8017729010. (Corresponding author: Zedong Nie.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) of Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (SIAT-CAS) under Application Nos. SIAT-IRB-200315-H0461 and SIAT-IRB-200815-H0525, and performed in line with the Declaration of Helsinki.

Mubashir Ali is with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the University Chinese Academy of Science, Beijing 101408, China (e-mail: mubashir@siat.ac.cn).

Jingzhen Li, Bokun Fan, and Zedong Nie are with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: lijz@siat.ac.cn; xjy109213@siat.ac.cn; zd.nie@siat.ac.cn).

Digital Object Identifier 10.1109/JBHI.2024.3464098

with blood flow and vascular properties. BG levels can affect blood volume through osmotic effects, endothelial dysfunction, the autonomic nervous system, blood rheology, and vascular compliance [11], [12], [13], [14]. Recent studies reveal that these effects can be analyzed through PPG [15]. PPG is widely used in wearable devices for healthcare monitoring due to its noninvasive, easy integration, and cost-effective nature [16]. Nowadays, PPG sensors are available in smart watches, smart bands, rings, clothes, chairs, etc. So, the PPG-empowered BG monitoring method could revolutionize diabetes management by offering affordable, accessible, and noninvasive solutions.

Recently, PPG has emerged as a noninvasive solution for healthcare applications such as cardiovascular disease assessment, sleep monitoring, stress analysis, etc [17], [18]. There has been limited attention given to PPG for diabetes management. In [19], an end-to-end transformer-based framework has been developed for multitask learning from videos to reconstruct the temporal and frequency correlation for BG prediction and blood pressure (BP) classification. A casual net was introduced within the container to analyze the higher-order information during fusion for better semantics understanding. Then, pair-wise temporal frequency maps were generated to extract the features from PPG. They revealed that the normotensive signals yield more discriminative features than hypertensive signals at the same BG level and this trend limited the BG feature extraction. In [20], a deduction learning-based approach for personalized BG monitoring has been introduced over a limited training dataset to overcome the issue of outlier predictions. Compared with induction learning (IL), a customized convolutional neural network (CNN) architecture was implemented with deduction learning to generate the additional feature dependencies. The model was trained over multiple rounds of data and low-confidence outlier predictions were removed by implementing an auto-screening algorithm. They revealed that the deduction learning model generates higher non-uniform feature patterns which were validated over 30 diabetic patients. In [21], a PPG system-on-chip (SoC) has been developed for noninvasive BG monitoring by offering minimum power consumption. A module named glucose prediction processor (GPP) was implemented for noise removal and feature extraction from PPG to construct a non-linear medium Gaussian SVM. The system was validated over short-recorded PPG data from 200 participants. They focused on the hardware part and reduced the overall cost by 25% by ensuring the minimum power consumption. In [22], noninvasive BG monitoring has been achieved through PPG extracted from mobile phone camera videos. The signals were preprocessed using a fitting-based sliding window algorithm including noise removal and segmentation. Gaussian function was adopted to extract and analyze the time and frequency domain features. They tested the proposed method over the short PPG data of 80 subjects to categorize the BG values as normal, edge, and warning.

In [23], extreme gradient boost regression (XGBR) and support vector regression (SVR) models have been developed on Mel frequency cepstral coefficients (MFCC) features for noninvasive BG monitoring from wristband PPG. XGBR performed well because of its parallel nature, efficient tree pruning, and regularization. In [24], various machine learning (ML) algorithms have been implemented for noninvasive BG monitoring

on in-ear PPG collected from 4 subjects using an oral glucose tolerance test (OGTT). The ensemble regression tree achieved the best results using temporal and energy features calculated from the Teager Kaiser energy operator (TKEO). In [25], BG level regression from mobile phone videos has been evaluated using different machine ML models. A gaussian filter with asymmetric least square (ALS) was implemented to remove the noise and motion artifacts. Partial least square has shown the best results in comparison with SVM and RF. In [26], a multimodel approach using weight-based Choquet integral and spatiotemporal feature fusion has been developed for BG monitoring on ECG and PPG. First, they extracted temporal and statistical features from raw and discrete wavelet transform (DWT) signals up to seven levels. Then, optimal features were selected using univariate feature selection (UFS), L1-based feature selection, and recursive feature elimination (RFE) techniques. Secondly, they extracted the spatial morphological features using five layers of ResNets and compressed them using a deep neural network (DNN). Then a multimodel fusion technique was implemented on both features using weight-based choquet integral which shows good results for BG monitoring. In [27], another deep learning (DL) approach has been developed for noninvasive BG monitoring on a larger cohort of short PPG from healthy and on-medication subjects. They used HRV and morphological features to train CNN with various lengths of filters in micro and macro training. Initially, they found that the BG prediction accuracy dropped up to 60% in the A region of CEG, because of medication subjects and the higher physiological diversity of subjects. Further analysis of separate groups (healthy and on medication) along with HbA1c boosted the performance up to a significant level. In [28], a wearable device has been developed for healthcare applications. They proposed a central state sensing (CSS) algorithm with DNN for noninvasive BG monitoring.

In reviewing the literature, various limitations have been encountered in previous studies. Firstly, most studies employed temporal and statistical features to develop the BG monitoring algorithms. Sometimes manual feature engineering is unable to capture the hidden relationships of data. Secondly, numerous studies are based on smaller datasets with limited subjects, with PPG recordings of a few minutes or seconds. It is difficult to capture the actual BG trends and events in short recorded PPG signals. Thirdly, a few DL approaches have been introduced for BG monitoring. However, they often fail to capture the BG trends because of short signals having individual differences and missing patterns. PPG signals are inherently non-linear and non-stationary, which presents a significant challenge for conventional DL techniques in capturing consistent relationships in data. This complexity demands further research on BG monitoring using larger datasets and domain-specific DL approaches.

To address these issues, we developed a cascaded bidirectional long short-term memory (BiLSTM) network with a cross-layer hierarchical feature fusion approach on multi-view attention for noninvasive BG monitoring using PPG. The main contributions of our study are described as follows:

- i. We design a convolutional multi-view attention block that efficiently captures the consistent features by parallelly implementing channel and spatial attention in

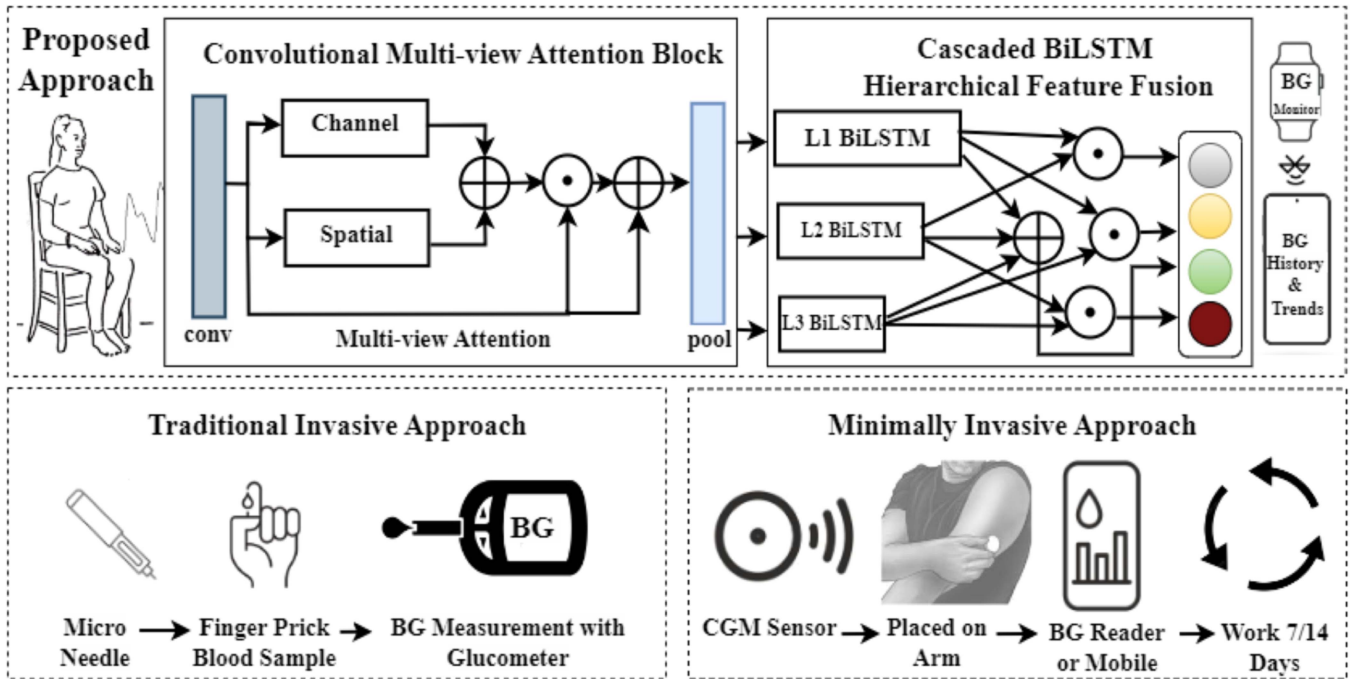


Fig. 1. Proposed noninvasive blood glucose monitoring approach vs invasive and minimally invasive methods.

separate branches, enabling precise spatial feature focus through adaptive contextual information aggregation.

- ii. We introduce a cascaded BiLSTM network with varied hidden sizes, which effectively explores the broader search space and dependencies via bidirectional learning to extract the fine-grained features.
- iii. Finally, we developed a hierarchical feature fusion with bilinear pooling through cross-layer interaction for non-invasive BG monitoring. In comparison with state-of-the-art works, the proposed approach showed a brilliant performance with root mean square error (RMSE) of 1.67 mmol/L and mean absolute relative difference (MARD) of 17.88%.

The rest of the paper is organized as follows. Section II is about the proposed methodology which describes the overall framework of the proposed approach. Section III explains the experimental setup and data acquisition method while the results are presented in Section IV. The discussion and conclusion of the study are reported in Sections V and VI, respectively.

II. PROPOSED METHODOLOGY

A. Algorithm Framework

Using the idea of multi-view attention, bilinear pooling, cross-layer interaction, and cascading on BiLSTM, we introduced a cascaded BiLSTM hierarchical feature fusion approach for noninvasive BG monitoring from PPG signal [29], [30], [31], [32]. The code of our approach (CBHFF) is available at (<https://github.com/drmubashirali/CBHFF>). Fig. 1 presents the proposed noninvasive BG monitoring approach in comparison with traditional invasive and minimally invasive methods. The traditional invasive method typically employs a glucometer to

measure BG levels through blood samples obtained via finger pricking. Despite its widespread use, this method causes pain, discomfort, and allergic reactions without continuous monitoring. In contrast, the minimally invasive approach provides continuous BG monitoring through continuous glucose monitors. However, this method incurs higher costs due to the need for periodic sensor replacement. The proposed approach introduces a noninvasive BG monitoring technique using PPG signals. The proposed approach is based on a convolutional multi-view attention block and a cascaded BiLSTM hierarchical feature fusion block, as explained below.

1) *Convolutional Multi-view Attention Block*: Initially, we created a convolutional multi-view attention block, which extracts the temporal features ranging from basic to complex by preserving the relationships. This block employs a separate branch channel and spatial attention in a parallel manner. The adaptive contextual aggregation of information is applied to attention maps to highlight the most pertinent features. It ensures that the model adapts to dynamic feature importance, thereby enhancing the accuracy of feature extraction for BG monitoring.

2) *Cascaded BiLSTM Hierarchical Feature Fusion*: In this block, a BiLSTM network is introduced in a cascaded manner with varied hidden sizes to learn bidirectionally. Subsequently, hierarchical feature fusion is implemented on the cascaded BiLSTM network with bilinear pooling through cross-layer interaction to generate higher-order features for BG monitoring from PPG data.

The structural form of the convolutional multi-view attention block is presented in Algorithm 1, where f_{PPG} and f'_{PPG} denote the input and output feature maps, respectively. Both Channel_Gate and Spatial_Gate compute the channel attention and spatial attention maps, respectively, which construct the final attention map, $f_{m'_{PPG}}$. Algorithm 2, presents the structural

Algorithm 1: Convolutional Multi-view Attention Block.

```

1: Input:  $f_{PPG}$ 
2: Output:  $f_{PPG}''$ 
3: MULTI-VIEW_ATTENTION( $f_{PPG}$ )
4: function CHANNEL_GATE( $f_{PPG}$ )
5:    $f_c \leftarrow \text{AvgPool}(f_{PPG})$ 
6:   for  $i \leftarrow 1$  to  $N$  do
7:      $f'_c \leftarrow \text{FC}(f_c)$ 
8:      $f''_c \leftarrow \text{BN}(f'_c)$ 
9:      $f'''_c \leftarrow \text{ACT}(f''_c)$ 
10:  end for
11:   $f_{CA} \leftarrow \text{reshape}(f'''_c, C)$ 
12:  return  $f_{CA}$ 
13: end Function
14: function SPATIAL_GATE( $f_{PPG}$ )
15:   $f_s \leftarrow \text{Conv}(f_{PPG}, C/R)$ 
16:   $f'_s \leftarrow \text{BN}(f_s)$ 
17:   $f''_s \leftarrow \text{ACT}(f'_s)$ 
18:  for  $i \leftarrow 1$  to  $D$  do
19:     $f d_s \leftarrow \text{Conv}(f''_s, C/R, d)$ 
20:     $f d'_s \leftarrow \text{BN}(f d_s)$ 
21:     $f d''_s \leftarrow \text{ACT}(f d'_s)$ 
22:  end for
23:   $f_{SA} \leftarrow \text{Conv}(f d''_s, S)$ 
24:  return  $f_{SA}$ 
25: end Function
26:  $f_{CA} \leftarrow \text{CHANNEL\_GATE}(f_{PPG})$ 
27:  $f_{SA} \leftarrow \text{SPATIAL\_GATE}(f_{PPG})$ 
28:  $f_m \leftarrow f_{CA} \oplus f_{SA}$ 
29:  $f_{m'} \leftarrow \sigma(f_m)$ 
30:  $f_{m'_{PPG}} \leftarrow f_{PPG} \otimes f_{m'}$ 
31:  $f_{PPG}'' \leftarrow f_{PPG} \oplus f_{m'_{PPG}}$ 
    
```

Algorithm 2: Cascaded BiLSTM Hierarchical Feature Fusion.

```

1: Input:  $f_{PPG}''$ 
2: Output:  $BGV$ 
3:  $f_{m_A} \leftarrow \text{BiLSTM\_A}(f_{PPG}'')$ 
4:  $f_{m_B} \leftarrow \text{BiLSTM\_B}(f_{PPG}'')$ 
5:  $f_{m_C} \leftarrow \text{BiLSTM\_C}(f_{PPG}'')$ 
6:  $f_{b_1} \leftarrow \text{CONCAT}(f_{m_A}, f_{m_B}, f_{m_C})$ 
7:  $f_{m_{AB}} \leftarrow f_{m_A} \otimes f_{m_B}$ 
8:  $f_{m_{AC}} \leftarrow f_{m_A} \otimes f_{m_C}$ 
9:  $f_{m_{BC}} \leftarrow f_{m_B} \otimes f_{m_C}$ 
10:  $f_{b_2} \leftarrow \text{CONCAT}(f_{m_{AB}}, f_{m_{AC}}, f_{m_{BC}})$ 
11:  $HOF \leftarrow \text{CONCAT}(f_{b_1}, f_{b_2})$ 
12:  $BGV \leftarrow \text{PREDICT}(HOF)$ 
    
```

form of the cascaded BiLSTM hierarchical feature fusion block. It takes the attention feature map f_{PPG}'' as input and predicts the BG values from higher-order features HOF , which is obtained through hierarchical feature fusion from cascaded BiLSTM.

B. PPG Preprocessing

Previous studies have observed that BG monitoring based on PPG signals provides relatively low accuracy due to the easy

influence of motion and noise artifacts. Further, the short PPG signal fails to monitor the glycemic events. Selecting an appropriate window size in the PPG is crucial for efficient information gain from the signal. Choosing a window size encompassing multiple beats/peaks is essential for effective BG monitoring. It mitigates the impact of the short and poor PPG waveform. Considering practical applications, the duration of physiological signal measurements has a notable effect on user acceptability. According to several studies, a waveform length of approximately 15 seconds is acceptable to most users. Therefore, a 15-second window has been selected for this work and validated through a multi-window size selection. The preprocessing encompasses three primary steps: PPG filtering, signal segmentation, and segment normalization. In the first step, a high-pass filter with a cutoff frequency of 0.5 Hz and a low-pass filter with a cutoff frequency of 10 Hz was applied to remove baseline wander and noise. In the second step, the continuous PPG signals were split into 5-minute intervals, with each interval further segmented into a 15-second window, aligning with the reference BG value. After that, z-score normalization was applied to segments to normalize the high-amplitude data points.

The initial PPG dataset (PD) is defined as a collection of PPG signals $X(i, j)$ and their corresponding BG values $Y(i, j)$ for subject i and day j , which can be expressed as (1):

$$PD = \{[X(i, j), Y(i, j)] \mid i \in \{1, \dots, n\} \text{ and } j \in \{1, \dots, m\}\} \quad (1)$$

Where n and m represent the total number of subjects and days, respectively. The filtered PPG signal is obtained by applying a bandpass filter, which can be expressed as (2):

$$X_f(i, j) = \text{lpf}(\text{hpf}(X(i, j), f_{HP}), f_{LP}) \quad (2)$$

Where hpf refers to a high-pass filter and $hpf(X(i, j), f_{HP})$ applies a high-pass filter with a cutoff frequency of f_{HP} to remove baseline wander. The lpf denotes a low-pass filter and $lpf(X(i, j), f_{LP})$ applies a low-pass filter with a cutoff frequency of f_{LP} to remove high-frequency noise. The filtered PPG signal is then divided into 5-minute intervals, each interval further segmented into 15-second windows, which can be expressed as (3):

$$S_k(i, j) = \text{Segment}(X_f(i, j), T_{\text{start}}, T_{\text{end}}) \quad (3)$$

Where $S_k(i, j)$ denotes the k^{th} segment of the filtered PPG signal for subject i on the day j while k is the segment index within the 5-minute interval. T_{start} and T_{end} define the start and end times of the 15-second window within each interval. Each segment $S_k(i, j)$ is normalized using z-score normalization to account the high amplitude of motion, which can be expressed as (4):

$$S_k^{\text{norm}}(i, j) = \frac{S_k(i, j) - \mu_{S_k(i, j)}}{\sigma_{S_k(i, j)}} \quad (4)$$

Where $S_k^{\text{norm}}(i, j)$ is the normalized segment which is computed through mean $\mu_{S_k(i, j)}$ and standard deviation $\sigma_{S_k(i, j)}$ of the $S_k(i, j)$. The 1st graph in Fig. 2 depicts the raw PPG recording for subject i for day j , with recording time shown in minutes on the x-axis and signal amplitude on the y-axis. Higher amplitude fluctuations in the raw PPG signal show visible motion artifacts. The 2nd graph presents a 15-second filtered segment k , with

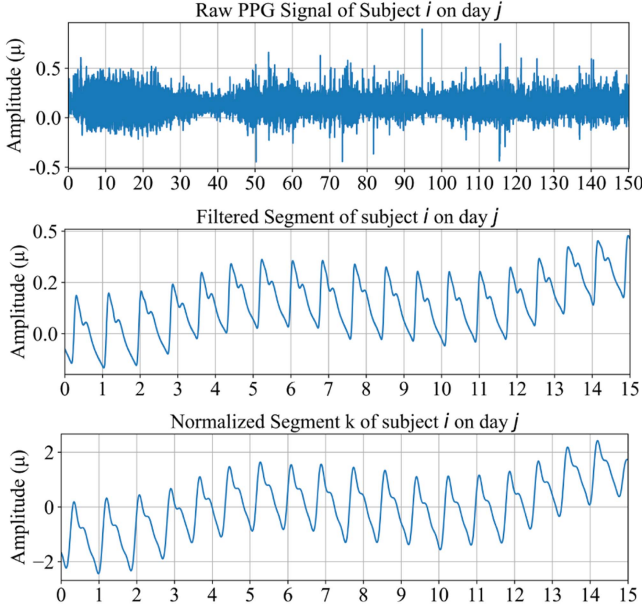


Fig. 2. Visualization of raw PPG signal in minutes, while filtered and normalized segments in seconds.

time on the x-axis and unnormalized amplitude on the y-axis. The 3rd graph shows the same 15-second segment k after normalization. The normalized segment has been scaled to a consistent amplitude for further processing. After preprocessing, each PPG signal $X(i, j)$ is transformed into a series of normalized segments $S_k(i, j)$. These segments are paired with their corresponding BG values $Y(i, j)$ resulting in a new dataset PD_{norm} , which can be expressed as (5):

$$PD_{norm} = \{[S_k^{norm}(i, j), Y(i, j)] \mid k \in \{1, \dots, K\}, i \in \{1, \dots, n\}, j \in \{1, \dots, m\}\} \quad (5)$$

Where K is the total number of segments and each $S_k^{norm}(i, j)$ represent the k^{th} normalized segment of the subject i on day j .

C. Convolutional Multi-View Attention Block

We developed a convolutional multi-view attention block which is based on the convolutional layer, channel attention and spatial attention branch, batch normalization layer, and maximum pooling layer for preliminary feature extraction. The normalized segments have been passed through the convolutional layer to generate the local feature map with a set of learnable filters. It can be expressed as (6):

$$F_{conv} = W_{conv} * X + b_{conv} \quad (6)$$

Where F_{conv} represents the convolved output which is generated through a convolution operation on input data X with a number of convolutional filters W_{conv} . The b_{conv} represents the bias term that allows the filter to adjust its output independently of the weighted sum, providing additional flexibility to learning. After that, the input feature map $F_{input} = F_{conv}$ forwarded to separate branch channel attention and spatial attention to generate the M_{cha_att} and M_{spa_att} attention maps as illustrated in Fig. 3. In channel attention, we exploit within-channel feature

dependencies by dynamically recalibrating and emphasizing key features to highlight the feature-specific response. Initially, the average-pooled features were computed as (7):

$$F_{avg_pool}^c = \frac{1}{N} \sum_{n=0}^{N-1} F_{input}^c[n] \quad (7)$$

Where $F_{avg_pool}^c$ is the average of the features, F_{input}^c is the feature map, and N is the number of elements in channel. Subsequently, the pooled features were projected into a lower-dimensional space with several fully connected layers to give attention to the most pertinent features as (8):

$$M_{cha_att}(F_{input}) = \text{ReLU} \left(\begin{bmatrix} W_{c,0} \cdot F_{avg_pool} + b_{c,0} \\ W_{c,1} \cdot G_c + b_{c,1} \end{bmatrix} \right) \quad (8)$$

Where $W_{c,0}$ and $W_{c,1}$ are the weight matrices of the fully connected layers and G_c represents the processed gated features at each stage.

The contextual information is pivotal to finding the most important spatial locations from data-receptive fields. In comparison with standard convolutions, we used the dilated convolutions to increase the receptive fields for constructing a more efficient spatial map. The process of spatial attention branch can be expressed as (9):

$$M_{spa_att}(F_{input}) = \text{Conv} \left(\text{DilatedConv} \left(\begin{bmatrix} W_{s,i} \cdot F_{input} + b_{s,i} \\ \text{ReLU}(G_s) \end{bmatrix} \right) \right) \quad (9)$$

Where dilated convolutional operation extracts the most pertinent spatial features from a large receptive field. After the last dilated convolution, we apply a convolution to get a single spatial attention map. We acquired the final attention map by applying the sigmoid function on combined M_{cha_att} and M_{spa_att} , which can be expressed as (10):

$$M_{attention} = \sigma(M_{cha_att} + M_{spa_att}) \quad (10)$$

Finally, the refined output features were acquired by adding an input feature map F_{input} with the Hadamard product of attention map $M_{attention}$ and input feature map F_{input} , which can be expressed as (11):

$$F_{output} = F_{input} + (F_{input} \odot M_{attention}) \quad (11)$$

D. Cascaded BiLSTM Hierarchical Feature Fusion

A BiLSTM network is an advanced variant of recurrent neural network (RNN) and showed effectiveness for biomedical signal processing [33], [34], [35]. Unlike standard LSTM networks, BiLSTMs process data bidirectionally, capturing information from both past and future states, thus offering a comprehensive understanding of time-dependent features in biomedical signals. Each LSTM unit within a BiLSTM comprises a cell state and three gates: input, forget, and output, which collectively regulate the flow of information. The cell state carries relevant information throughout the sequence processing and mitigates the vanishing gradient problem in traditional RNNs. This bidirectional analysis and gating mechanism make BiLSTMs suited for handling time-series data in biomedical signals.

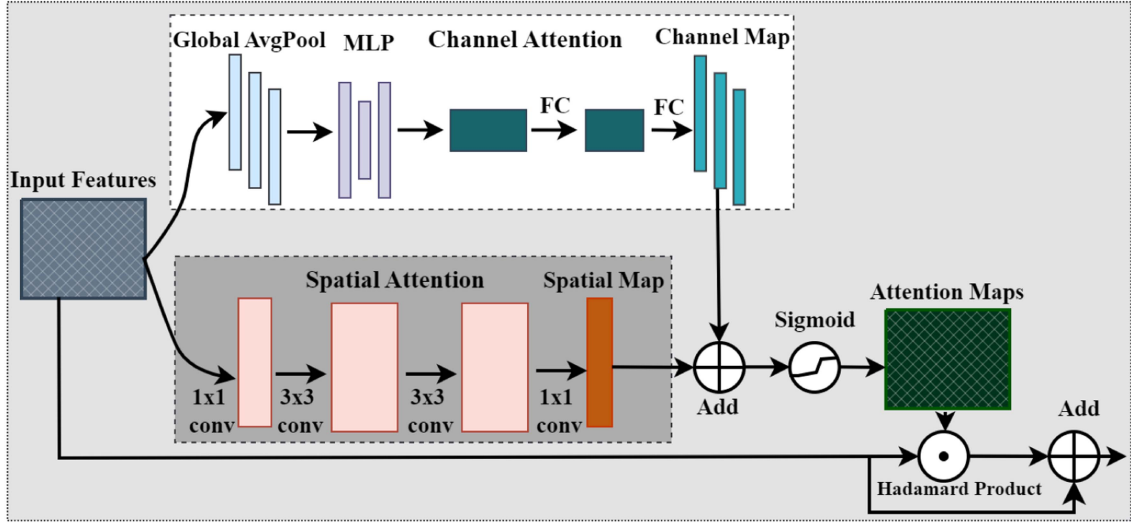


Fig. 3. Diagram of multi-view attention module (separate branch implementation of channel attention and spatial attention).

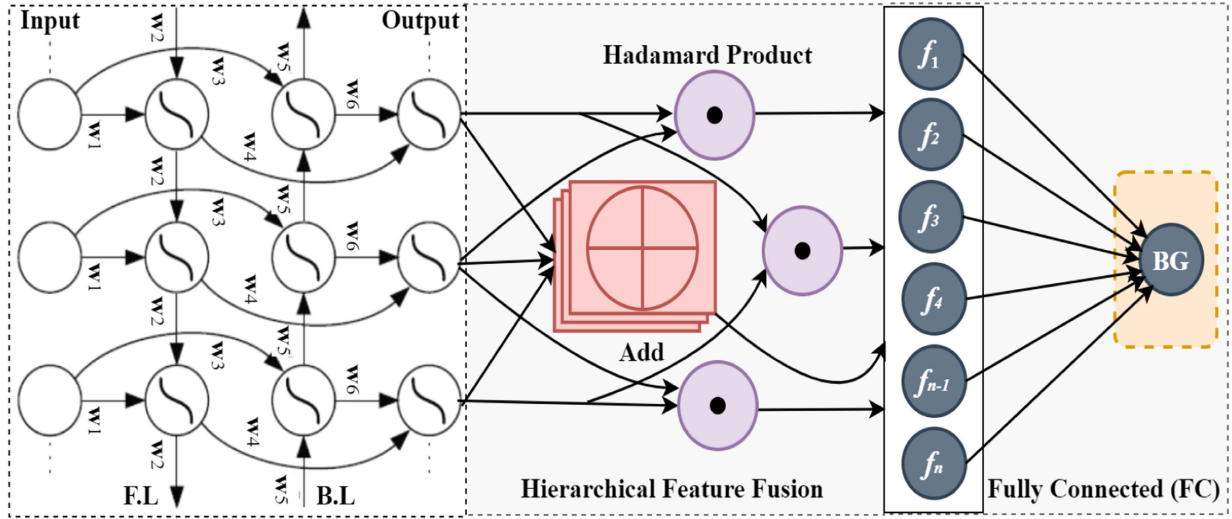


Fig. 4. Diagram of cascaded BiLSTM network with hierarchical feature fusion at cross-layer interaction.

So, we created a cascaded BiLSTM network with varied hidden sizes which processes the feature map to extract the long-range time-dependent features from both forward (future) and backward (past) directions. Through this bidirectional processing, the network garners a comprehensive understanding of the temporal context, which is crucial for accurately interpreting the PPG signals. The cascaded BiLSTM network operates at multiple scales of feature complexity. The smallest hidden size aims to extract the basic temporal dependencies and fundamental patterns, while the larger sizes are directed at more abstract and complex temporal patterns. This multiscale approach allows the network to simultaneously manage model capacity and complexity, ensuring nuanced feature learning. The layered structure, moving from smaller to larger hidden sizes, facilitates a hierarchical feature extraction paradigm where higher-level BiLSTM layers build upon the refined outputs of their predecessors, capturing a richer and more diverse set of temporal

features. The illustration of the cascaded BiLSTM hierarchical feature fusion block is presented in Fig. 4. The feature map from the attention block is passed through the cascaded BiLSTM network in which each BiLSTM unit consists of two sets of gates and cell states: one for forward processing and one for backward processing. For simplicity, we describe the operations of one direction, keeping in mind that a similar set of operations occurs in the reverse direction, and the final outputs are combined. The forget gate in the BiLSTM decides about information to be discarded from the cell state and can be expressed as (12):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (12)$$

Where σ denotes a sigmoid function that outputs values between 0 and 1 for decision. The h_{t-1} , x_t , W_f , and b_f represent the previous hidden state, current input, weight matrix for the forget gate, and forget gate bias, respectively. If f_t is close to 0, it means that the cell state should forget information, whereas a

value close to 1 indicates retention. The input gate determines how much of the new information should be stored in the cell state, and can be expressed as (13).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (13)$$

The input gate uses the σ function to determine which values in the state to update. The cell state is a memory line that runs through the entire chain, with some minor linear interactions, and can be expressed as (14):

$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (14)$$

Where \tanh is a hyperbolic tangent function to produce output values between -1 and 1, thus ensuring that the states are normalized. The cell state is also known as the memory of the network. Its update can be expressed as (15):

$$C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t \quad (15)$$

Where C_t denotes the memory cell update. First, it forgets the decided memories, and then it adds the new candidate values, scaled by how much decided to update each state value. The output gate controls how much of the cell state is used in the output, ensuring that irrelevant information is filtered out, and can be expressed as (16):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (16)$$

The output gate decides the next hidden state h_t which can be expressed as (17):

$$h_t = o_t \times \tanh(C_t) \quad (17)$$

Where the cell state C_t is pushed through \tanh to normalize it and then scaled by the output gate for hidden state decision. In the bidirectional case, the forward pass hidden states \vec{h}_t and the backward pass hidden states \overleftarrow{h}_t are computed with their respective weights and biases. The final hidden state for each time step in the BiLSTM is the concatenation of the forward and backward hidden states, which can be expressed as (18):

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (18)$$

The beauty of LSTMs lies in their gates, which carefully regulate the flow of information to keep only what's necessary, allowing them to remember or forget patterns. In BiLSTM, this ability is augmented as it can access future context in its backward pass, making it adept at understanding more complex, bidirectional contexts in sequences. Bilinear pooling has proven effective in fine-grained feature learning. So, we implemented bilinear pooling with cross-layer interaction in a cascaded BiLSTM network. Instead of simple sequential bilinear pooling, we exploit interactions from multiple layers. Suppose, the descriptors fm_A , fm_B , and fm_C denote the feature maps acquired from the cascaded BiLSTM network. The first, larger feature block was obtained by combining these individual feature maps, which can be expressed as (19):

$$\mathbf{fb}_1 = [fm_A; fm_B; fm_C] \quad (19)$$

Where \mathbf{fb}_1 denotes the 1st feature block. Subsequently, bilinear pooling was applied on feature maps by exploiting cross-layer

interaction, and the second feature block was acquired, which can be expressed as (20):

$$\mathbf{fb}_2 = \begin{bmatrix} fm_A \odot fm_B \\ fm_A \odot fm_C \\ fm_B \odot fm_C \end{bmatrix} \quad (20)$$

Where \mathbf{fb}_2 specifies the 2nd feature block, acquired through the Hadamard product between different cross-layer interactions. The final higher-order feature block was then constructed by combining \mathbf{fb}_1 and \mathbf{fb}_2 , which can be expressed as (21):

$$\mathbf{hof} = [\mathbf{fb}_1; \mathbf{fb}_2] \quad (21)$$

Where \mathbf{hof} denotes the higher-order feature vector used to predict the BG values. In a nutshell, the combination of BiLSTM and cascading allows the model to explore a broad search space of temporal dependencies and spatial patterns in the PPG signals. Moreover, bilinear pooling at cross-layer interaction with hierarchical feature fusion helps to construct a rich feature vector for robust BG monitoring.

E. Performance Evaluation Metrics

For a comprehensive evaluation of the proposed approach for BG monitoring, this study adopts a dual-assessment strategy focusing on both numerical and clinical accuracy. Specifically, the numerical accuracy is assessed using Root Mean Square Error (RMSE) and Mean Absolute Relative Difference (MARD) [36]. RMSE indicates the magnitude of error between the predicted and the actual BG values. The smaller the RMSE, the closer the predicted values are to the actual values. Mathematically, it can be expressed as (22):

$$RMSE_{BG} = \sqrt{\frac{1}{N} \sum_{i=1}^N (BG_{ref,i} - BG_{pred,i})^2} \quad (22)$$

Where BG_{ref} is the actual BG value, BG_{pred} is the predicted BG value, and N is the total number of observations. MARD provides an average of the relative errors between predicted and actual BG values, as a percentage. It can be expressed as (23):

$$MARD_{BG} = \frac{100}{N} \sum_{i=1}^N \frac{|BG_{ref,i} - BG_{pred,i}|}{BG_{ref,i}} \quad (23)$$

The Clarke error grid (CEG) analysis is a prominent technique used for evaluating the clinical accuracy of blood glucose monitoring systems [37]. It visualizes how much predicted BG measurements deviate from the actual BG values and lead to potential clinical implications. CEG is divided into five zones (A to E) based on the clinical implications of errors. The zone A predictions are clinically accurate. Zone B predictions are not as accurate as zone A but are still clinically acceptable. Zone C predictions are incorrect and lead to unnecessary treatment while zone D predictions fail to detect hypoglycemia or hyperglycemia. The predictions that fall in zone E are the most dangerous because of almost opposite to the reference BG values. So, larger predictions in zone A and zone b presents a clinically better and acceptable accuracy.

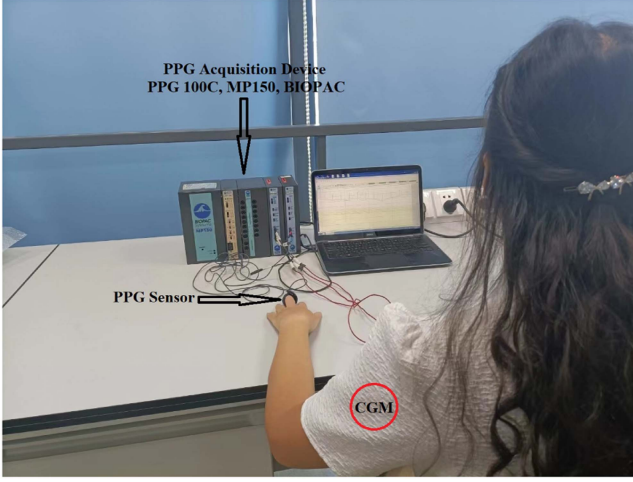


Fig. 5. Experimental setup for BG and PPG data acquisition.

III. EXPERIMENTAL SETUP

A. Study Population

In our study, we collected data from 21 participants at the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (SIAT-CAS) from 2020 to 2021. This group included 17 normal participants and 4 with diabetes. On average, participants were 36.9 years old with a standard deviation of 12.1 years, and they had a body mass index of 23.3 with a standard deviation of 8.5 kg/m². All participants were carefully selected without any heart-related complications or other diseases. Our study protocols were approved by the SIAT-CAS Institutional Review Board, with the approval numbers SIAT-IRB-200315-H0461 and SIAT-IRB-200815-H0525. All procedures adhered to the principles of the Declaration of Helsinki.

B. PPG and BG Acquisition

We conducted the experiment under a controlled environment to minimize variables like movement, environmental factors, and individual physical conditions to reduce the noise. Before the experiment, participants were instructed to fast for at least 8 hours, avoid medications and any intense activities like playing football. On the day of the test, participants were asked to sit for 5 minutes before the experiment started and to remain seated throughout its duration. The experimental setup illustrated in Fig. 5, involved a ring-type PPG sensor placed on the left hand's middle finger for continuous signal acquisition. We used a PPG 100C device (BIOPAC, Goleta, USA) with a high sampling frequency of 1 kHz. For BG values, we employed Food and Drug Administration (FDA) approved continuous glucose monitor (CGM) G6 from Dexcom [38]. It records and saves BG readings every 5 minutes. The entire experiment, spanning roughly 150 minutes, started at 9:00 AM and concluded by 11:30 AM for each participant. Experiment was divided into two stages: the fasting and the dietary stages. In the fasting stage, from 9:00 AM to 9:30 AM, PPG signals were recorded without food intake. In the dietary stage, participants consumed 100 g of bread and 250mL of water at 9:30 AM. Continuous PPG data was collected until the experiment ended at 11:30 am.

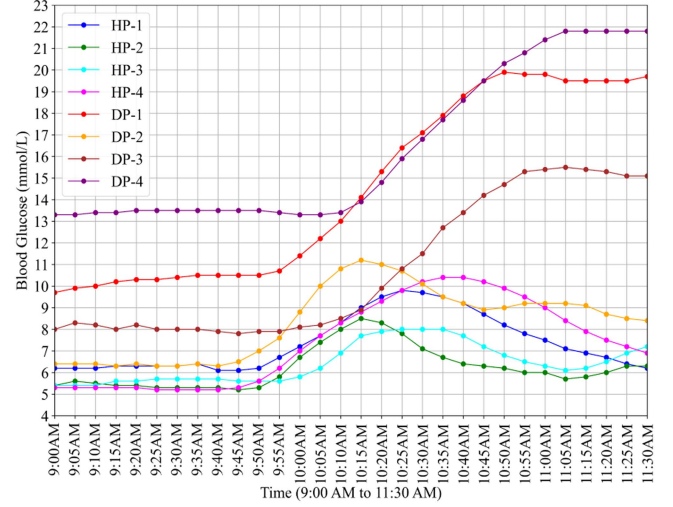


Fig. 6. BG data of eight different participants.

C. Data Processing

In this study, we collected and analyzed PPG signals along with their corresponding BG values spanning 103 days from 21 participants. The breakdown of data acquisition is as follows: 14 participants contributed data from 5 days each, 3 participants from 6 days each, 3 participants from 4 days each, and 1 participant for 3 days. This varied data collection across various participants adds diversity to our dataset. Fig. 6 shows the BG trends of 8 participants including four healthy and 4 diabetic patients. HP and DP refer to healthy and diabetic participants, respectively.

Given the uniqueness of physiological signals across individuals and even for the same individual across different days, the 103-day dataset offers a fair representation of physiological variations in the population. To the best of our knowledge, we have collected the largest long-recorded PPG dataset for BG analysis, as long-range recordings are essential to accurately capture the glycemic events. The majority of related studies have relied on short-duration PPG recordings for noninvasive BG monitoring, which are often insufficient for predicting glycemic events accurately. To robustly evaluate our proposed approach for BG monitoring, we employed 10-fold cross-validation across the entire dataset to examine the performance in different folds.

IV. EXPERIMENTAL RESULTS

A. Ablation Studies and BG Prediction Results

We performed several ablation studies to assess the performance of the proposed algorithm. The methods and results of different ablation studies are presented in Table I. Method A employs a convolutional multi-view attention block as the base model in our ablation studies. Method A consists of four stages of convolutional layers with 16, 32, 64, and 128 filters, each with a kernel size of 3. Each convolutional layer is followed by multi-view attention to extract and highlight the most pertinent features. The A method showed the RMSE, MARD, Zone A, and Zone A+B of CEG analysis as 1.95 mmol/L, 19.74%, 57.12%, and 96.26%, respectively. Method B extends the base

TABLE I
EXPERIMENTAL SETUP AND RESULTS OF MEAN OF 10-FOLD CROSS-VALIDATION OF DIFFERENT ABLATION STUDIES

Method	Algorithm Structure and Description	RMSE (mmol/L)	MARD (%)	CEG A (%)	CEG A+B (%)
A	Convolutional Multi-view Attention Block	1.95	19.74	57.12	96.26
B	Convolutional Multi-view Attention Block with BiLSTM	1.86	19.22	57.88	97.07
C	Convolutional Multi-view Attention Block with Cascaded BiLSTM Hierarchical Feature Fusion	1.67	17.88	61.40	98.80

TABLE II
10-FOLD CROSS-VALIDATION OF CBHFF

Fold	RMSE (mmol/L)	MARD (%)	CEG A (%)	CEG A+B (%)
1	1.73	18.13	60.42	98.50
2	1.66	17.96	61.05	98.84
3	1.65	18.82	58.68	99.02
4	1.67	17.82	61.87	98.62
5	1.72	18.69	59.91	98.79
6	1.67	17.79	61.99	98.84
7	1.68	17.23	62.59	98.76
8	1.66	17.28	62.38	98.94
9	1.61	17.80	61.54	99.08
10	1.70	17.28	63.60	98.62
Average	1.67	17.88	61.40	98.80

architecture of method A by integrating a BiLSTM layer to enhance the temporal feature analysis of PPG signals. With a hidden size of 4, the BiLSTM effectively manages the variability of PPG signals, utilizing both past and future data contexts to enhance predictive accuracy. This addition improves the model's capability to discern important temporal patterns for noninvasive BG monitoring. Specifically, the RMSE improved from 1.95 mmol/L to 1.86 mmol/L, and the MARD dropped from 19.74% to 19.22%. Furthermore, the percentage of BG values in zone A increased from 57.12% to 57.88% and in zone A+B from 96.26% to 97.07%. Method C presents the proposed model and integrates a cascaded BiLSTM network with hierarchical feature fusion into the base method A. Method C employs three BiLSTM layers with decreasing hidden sizes (4, 2, and 1). Each layer is designed to capture and process distinct temporal dependencies at different spectrums. The cross-layer interaction is exploited on the output features of these layers. Specifically, the operation of concatenation and pairwise products is taken on cross-layer features. This process allows the exploration of complex feature interactions to construct higher-order features. Method C gives the most optimal results as an RMSE of 1.67 mmol/L, MARD of 17.88%, zone A 61.40%, and A+B 98.80%.

For a better assessment of the proposed approach, we performed 10-fold cross-validation and the results are elaborated in Table II. Fold 9 shows the minimum RMSE of 1.61 mmol/L while fold 1 shows the maximum RMSE of 1.73. Fold 7 and fold 3 show the minimum and maximum MARD of 17.23% and 18.82%, respectively. The minimum RMSE and MARD indicate the most optimal results in terms of the lowest magnitude of error and the most consistent relative accuracy of BG predictions. The best zone A predictions were achieved in fold 10, reaching 63.60%, while the combined zone A+B predictions reached 99.08% in fold 9. The highest zone A and combined Zone A+B results reflect strong clinical relevance for BG prediction. Interestingly, all the BG prediction values fell into the clinically

acceptable regions with a higher ratio in zone A. Around 60% points fell in zone A and 98% points in zone A+B in all folds. The average MARD, RMSE, zone A, and zone A+B were 17.88%, 1.67 mmol/L, 61.40% and 98.80%, respectively.

We conducted a correlation study of the proposed model to assess whether the predicted BG values follow actual BG levels. We employed Pearson and Spearman correlation analyses for a comprehensive evaluation. The Pearson correlation coefficient measures the linear relationship between the predicted and actual BG values. It provides a value between -1 and $+1$, where 1 indicates a perfect positive linear relationship, and -1 indicates a perfect negative linear relationship. The Pearson correlation coefficient of 0.75 indicates a strong positive linear relationship between the predicted and actual BG levels, suggesting that as the predicted values increase, the actual values also tend to increase. The P-value ($P < 0.001$) statistically confirms the high unlikelihood of this correlation occurring by chance, validating the predictive accuracy in terms of linear relationships. Spearman's rank correlation coefficient assesses the monotonic relationship between predicted and actual BG values. It evaluates the rank order of the predicted BG with the rank order of the actual BG without assuming linearity. Spearman's rank correlation coefficient of 0.74 indicates a strong positive monotonic relationship between the predicted and actual BG levels. Further, the P-value ($P < 0.001$) rejects the null hypothesis of no monotonic correlation.

B. Comparison With Previous Studies

Table III presents the comparative analysis of the prediction performance of the proposed approach with different methods on the in-house dataset. For a fair comparison, we re-implemented state-of-the-art, end-to-end deep learning methods on our larger dataset having long-range PPG signals [19], [20], [28], [39]. They used the short-recorded PPG with a limited number of samples to train the models which eventually performed well on smaller data. Our dataset consists of 56,940 samples with a comprehensive range of BG trends. Most of them used the CNN and DNN with an increasing number of filters and time-frequency maps. Comparative results highlight that our proposed method gives the minimum RMSE of 1.67 mmol/L and the smallest MARD of 17.88%. Furthermore, the Clarke error grid analysis showed the largest points in zone A of 61.40% and Zone A+B of 98.80%. These results indicate that our method is suitable for BG monitoring.

V. DISCUSSION

BG monitoring is critical for the timely detection of hyperglycemic and hypoglycemic events, which is vital in active diabetes management. In this study, we developed a noninvasive

TABLE III
RESULTS COMPARISON OF DIFFERENT METHODS ON IN-HOUSE DATASET

Reference	Method	Features	RMSE (mmol/L)	MARD(%)	CEG A(%)	CEG A+B (%)
[20]	DLDRD	End-to-end	2.22	20.76	54.52	94.61
[28]	CSS-DNN	End-to-end	1.91	19.63	57.33	96.42
[39]	DBSCAN-CNN	End-to-end	1.89	19.41	58.12	97.07
[19]	VCCC	End-to-end	1.83	19.02	58.34	97.58
Ours	CBHFF	End-to-end	1.67	17.88	61.40	98.80

DLDRD: Deduction learning with dozen rounds of data; CSS-DNN: Central state sensing deep neural network; DBSCAN-CNN: Density-based spatial clustering of applications with noise and convolution neural networks; VCCC: Video-based cocktail causal container; CBHFF: Cascaded BiLSTM hierarchical feature fusion

BG monitoring approach based on multi-view attention and cascaded BiLSTM network with hierarchical feature fusion through cross-layer interaction. For validation of our proposed approach, we collected the PPG and BG data of a total 103 days from 21 participants including, healthy and diabetic patients. We aimed to record a broad spectrum of glycemic responses. Hence, we captured long-range signals at various glycemic states, including pre-meal, during-meal, and post-meal stages. This method produced a comprehensive dataset reflecting not just inter-individual variability but also the day-to-day variations in BG levels within the same individuals. The long-range signals provide a more detailed temporal context that improves the model's ability to discern subtle patterns and trends, which is crucial for accurate and reliable BG monitoring. This comprehensive temporal coverage ensures the model's robustness to the natural variability in BG levels and is better suited for practical applications where conditions are not tightly controlled. The PPG signals are easily influenced by noise, movement, and environmental factors, which leads to poor waveform. Also, the PPG signals have nonstationary behavior with complex dynamics. To address this, we adopted a 15-second segment from continuous PPG signals, which balances user acceptability with the need for accurate BG monitoring. This segment size has been empirically validated and is long enough to incorporate multiple heartbeats, thus reducing the impact of short and noisy signals. To facilitate the training and testing followed by a 10-fold cross-validation technique, we randomly split the complete data to prevent bias rather than division per subject or day. This division mitigates the risk of overfitting to particular subjects or temporal patterns. Furthermore, we partitioned the dataset on a per-subject basis to assess the proposed method, yielding good results. The subject-wise division is more suitable for personalized modeling. We also assessed our model on day-wise data division, which showed inconsistent results. Some folds showed good performance, and some folds were poor. The random division of the entire dataset is a reflection of a more generic and pragmatic approach suitable for real-world applications, considering that PPG readings can vary substantially among individuals.

Specifically, our proposed approach is based on convolutional multi-view attention, cascaded BiLSTM network, and cross-layer bilinear polling with hierarchical feature fusion. To the best of our knowledge, this is the first approach using a cascaded BiLSTM network with multi-view attention and hierarchical feature fusion for PPG signals to estimate the BG levels. Initially, the convolutional layer extracts the temporal feature map from PPG, which subsequently passed to multi-view attention to generate the channel attention map and spatial attention map, respectively. The attention mechanism is implemented through a separate branch flow that employs dilated convolutions within

the spatial branch to broaden the receptive field and focus on wide spatial features. These feature maps are refined through a four-stage convolutional multi-view attention block, which is designed to extract robust and optimal features. Subsequently, a cascaded BiLSTM network is implemented to extract a wide range of time-dependent features. Each BiLSTM block targets features at varying scales through bidirectional learning. The implementation of BiLSTM improved the limitations associated with unidirectional LSTMs that struggle with dependencies in one direction. After that, bilinear pooling is applied at cross-layer interaction on BiLSTM features to attain more fine-grained features. Finally, a hierarchical feature fusion approach is developed to fuse these fine-grained features to construct the higher-order features for BG monitoring.

The proposed CBHFF approach showed excellent numerical and clinical accuracy compared to the existing BG monitoring approaches. In 10-fold cross-validation, the CBHFF gives MARD of 17.88%, RMSE of 1.67 mmol/L and 98.80% predicted values fell in clinically acceptable A+B regions of CEG, which highlight the significance of the proposed approach.

The rapid development of smart wearable devices enables the smooth acquisition of PPG from health bands, rings, and watches. Therefore, PPG-based BG monitoring can provide a noninvasive, cost-effective, and readily attainable solution for diabetes management. This technology revolutionizes traditional BG monitoring and positively impacts high-risk diabetic populations. Despite the potential clinical significance and promising performance of CBHFF in BG monitoring, this work has three limitations that need to be noted and addressed in the future. First, the ratio of diabetic participants was low, which needs to be increased to develop a more robust dataset to analyze the feasibility of CBHFF. Second, the population of this study was based on uni-region (Chinese). A further study of multi-region populations is required to assess the validity of the proposed approach on different populations. Third, the experiment was conducted in a controlled environment. So, the external factors (emotion, physical activity), blood diseases, and heart diseases were not considered, which may have caused the variations in the PPG signal.

VI. CONCLUSION

A novel CBHFF approach has been developed for noninvasive blood glucose monitoring using PPG signals. The CBHFF is based on a convolutional multi-view attention block followed by a cascaded BiLSTM hierarchical feature fusion block. In particular, the convolutional multi-view attention block processes the PPG signal to extract the temporal features. It generates channel and spatial attention maps, constructing refined feature

maps by aggregating contextual information. The subsequent cascaded BiLSTM network efficiently captures the long-range temporal relationships related to BG levels by employing bidirectional learning at various scales. This is followed by bilinear pooling, which performs cross-layer interactions to construct robust feature vectors. Finally, hierarchical feature fusion is implemented to obtain higher-order features for BG monitoring. Compared with various existing BG monitoring approaches, the CBHFF demonstrates superior performance by showing an average MARD of 17.88% and an average RMSE of 1.67 mmol/L. Moreover, the predicted BG values fell 98.80% in zone A+B of Clarke error grid analysis, highlighting the excellent clinical relevance of our proposed approach. In future, we plan to conduct a more robust and comprehensive study based on a multi-regional population, with a higher ratio of diabetic participants in real-life environments by considering physical activities. Further, we have a plan to investigate the influence of cardiac and blood diseases on our PPG based CBHFF approach.

REFERENCES

- [1] A. T. Kharroubi and H. M. Darwish, "Diabetes mellitus: The epidemic of the century," *World J. Diabetes*, vol. 6, no. 6, pp. 850–867, 2015.
- [2] H. Sun et al., "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes Res. Clin. Pract.*, vol. 183, Jan. 2022, Art. no. 109119.
- [3] F. Aikaeli et al., "Prevalence of microvascular and macrovascular complications of diabetes in newly diagnosed type 2 diabetes in low-and-middle-income countries: A systematic review and meta-analysis," *PLOS Glob. Public Health*, vol. 2, no. 6, Jun. 2022, Art. no. e0000599.
- [4] T. S. Group, "Long-term complications in youth-onset type 2 diabetes," *New England J. Med.*, vol. 385, no. 5, pp. 416–426, Jul. 2021.
- [5] D. J. Magliano et al., "What is diabetes?," in *Proc. IDF DIABETES ATLAS [Internet]*, 2021.
- [6] L. Olansky and L. Kennedy, "Finger-stick glucose monitoring - issues of accuracy and specificity," *Diabetes Care*, vol. 33, no. 4, pp. 948–949, Apr. 2010.
- [7] G. Cappon, M. Vettoretti, G. Sparacino, and A. Facchinetti, "Continuous glucose monitoring sensors for diabetes management: A review of technologies and applications," *Diabetes Metab. J.*, vol. 43, no. 4, pp. 383–397, Aug. 2019.
- [8] W. Animaw and Y. Seyoum, "Increasing prevalence of diabetes mellitus in a developing country and its related factors," *Plos One*, vol. 12, no. 11, Nov. 2017, Art. no. e0187670.
- [9] A. A. Alian and K. H. Shelley, "Photoplethysmography," *Best Pract. Res., Clin. Anaesthesiol.*, vol. 28, no. 4, pp. 395–406, 2014.
- [10] J. Allen, D. Zheng, P. A. Kyriacou, and M. Elgendy, "Photoplethysmography (PPG): State-of-the-art methods and applications," *Physiol. Meas.*, vol. 42, no. 10, pp. 1–4, Oct. 2021.
- [11] A. Guemes and P. Georgiou, "Review of the role of the nervous system in glucose homeostasis and future perspectives towards the management of diabetes," *Bioelectron. Med.*, vol. 4, no. 1, pp. 1–18, Jul. 2018.
- [12] C. E. Tabit, W. B. Chung, N. M. Hamburg, and J. A. Vita, "Endothelial dysfunction in diabetes mellitus: Molecular mechanisms and clinical implications," *Rev. Endocr. Metabolic Disord.*, vol. 11, pp. 61–74, 2010.
- [13] E. Monte-Moreno, "Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques," *Artif. Intell. Med.*, vol. 53, no. 2, pp. 127–138, 2011.
- [14] C. Le Devehat, M. Vimeux, and T. Khodabandehlou, "Blood rheology in patients with diabetes mellitus," *Clin. Hemorheol. Microcirculation*, vol. 30, no. 3/4, pp. 297–300, 2004.
- [15] S. Zanelli, M. Ammi, M. Hallab, and M. A. El Yacoubi, "Diabetes detection and management through photoplethysmographic and electrocardiographic signals analysis: A systematic review," *Sensors*, vol. 22, no. 13, pp. 1–15, 2022.
- [16] P. H. Charlton and V. Marozas, "Wearable photoplethysmography devices," in *Photoplethysmography Technology, Signal Analysis and Applications*. Academic, Jan. 2022, pp. 401–439.
- [17] J. Park, H. S. Seok, S. S. Kim, and H. Shin, "Photoplethysmogram analysis and applications: An integrative review," *Front. Physiol.*, vol. 12, Mar. 2022, Art. no. 2511.
- [18] J. Zhang et al., "Diagnostic features and potential applications of PPG signal in healthcare: A systematic review," *Healthcare*, vol. 10, no. 3, Mar. 2022, Art. no. 547.
- [19] C. H. Zhang et al., "Video based cocktail causal container for blood pressure classification and blood glucose prediction," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 1118–1128, Feb. 2023.
- [20] W. R. Lu, W. T. Yang, J. Chu, T. H. Hsieh, and F. L. Yang, "Deduction learning for precise noninvasive measurements of blood glucose with a dozen rounds of data for model training," *Sci. Rep.*, vol. 12, no. 1, pp. 1–12, 2022.
- [21] A. Hina and W. Saadeh, "A 186 μ W photoplethysmography-based noninvasive glucose sensing SoC," *IEEE Sensors J.*, vol. 22, no. 14, pp. 14185–14195, Jul. 2022.
- [22] G. B. Zhang et al., "A noninvasive blood glucose monitoring system based on smartphone PPG signal processing and machine learning," *IEEE Trans. Ind. Inform.*, vol. 16, no. 11, pp. 7209–7218, Nov. 2020.
- [23] A. Prabha, J. Yadav, A. Rani, and V. Singh, "Intelligent estimation of blood glucose level using wristband PPG signal and physiological parameters," *Biomed. Signal Process. Control*, vol. 78, 2022, Art. no. 103876.
- [24] G. Hammour and D. P. Mandic, "An in-ear PPG-Based blood glucose monitor: A proof-of-concept study," *Sensors*, vol. 23, no. 6, pp. 1–14, 2023.
- [25] T. T. Islam, M. S. Ahmed, M. Hassanuzzaman, S. A. B. Amir, and T. Rahman, "Blood glucose level regression for smartphone PPG signals using machine learning," *Appl. Sci. - Basel*, vol. 11, no. 2, pp. 1–20, 2021.
- [26] J. Li et al., "Noninvasive blood glucose monitoring using spatiotemporal ECG and PPG feature fusion and weight-based choquet integral multi-model approach," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 08, 2023.
- [27] J. S. Chu, W. T. Yang, W. R. Lu, Y. T. Chang, T. H. Hsieh, and F. L. Yang, "90% accuracy for photoplethysmography-based non-invasive blood glucose prediction by deep learning with cohort arrangement and quarterly measured HbA1c," *Sensors*, vol. 21, no. 23, pp. 1–13, 2021.
- [28] E. Lee and C. Y. Lee, "PPG-Based smart wearable device with energy-efficient computing for mobile health-care applications," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13564–13573, Jun. 2021.
- [29] J. Park, S. Woo, J. Y. Lee, and I. S. Kweon, "BAM: Bottleneck attention module," in *Proc. Brit. Mach. Vis. Conf.*, Jul. 2018, pp. 1–14.
- [30] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Dec. 2017, pp. 6154–6162.
- [31] J. H. Kim, K. W. On, W. Lim, J. Kim, J. W. Ha, and B. T. Zhang, "Hadamard product for low-rank bilinear pooling," in *Proc. 5th Int. Conf. Learn. Representations, ICLR 2017 - Conf. Track Proceedings. Int. Conf. Learn. Representations, ICLR*, Oct. 2016, pp. 1–14.
- [32] Y. Gao, O. Beijbom, N. Zhang, and T. Darrell, "Compact bilinear pooling," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Dec. 2016, pp. 317–326.
- [33] B. Tutuko et al., "DAE-ConvBiLSTM: End-to-end learning single-lead electrocardiogram signal for heart abnormalities detection," *Plos One*, vol. 17, no. 12, Dec. 2022, Art. no. e0277932.
- [34] D. Gao et al., "CSF-GTNet: A novel multi-dimensional feature fusion network based on Convnext-GeLU-BiLSTM for EEG-signals-enabled fatigue driving detection," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 5, pp. 2558–2568, May 2024.
- [35] J. Yang, X. Huang, H. Wu, and X. Yang, "EEG-based emotion classification based on bidirectional long short-term memory network," *Procedia Comput. Sci.*, vol. 174, pp. 491–504, Jan. 2020.
- [36] F. Reiterer et al., "Significance and reliability of MARD for the accuracy of CGM systems," *J. Diabetes Sci. Technol.*, vol. 11, no. 1, Jan. 2017, Art. no. 59.
- [37] W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter, and S. L. Pohl, "Evaluating clinical accuracy of systems for self-monitoring of blood glucose," *Diabetes Care*, vol. 10, no. 5, pp. 622–628, Sep. 1987.
- [38] H. K. Akturk, R. Dowd, K. Shankar, and M. Derdzinski, "Real-world evidence and glycemic improvement using dexcom G6 features," *Diabetes Technol. Therapeutics*, vol. 23, no. 1, Mar. 2021, Art. no. S-21.
- [39] J. Li, L. Tobore, Y. Liu, A. Kandwal, L. Wang, and Z. Nie, "Non-invasive monitoring of three glucose ranges based on ECG by using DBSCAN-CNN," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 9, pp. 3340–3350, Sep. 2021.