Computing and Informatics, Vol. 45, 2026, 964–994, doi: 10.31577/cai_2026_4_964

LIGHTWEIGHT TIME-FREQUENCY NEURAL NETWORK FOR NON-INVASIVE GLUCOSE ESTIMATION ON WEARABLE DEVICES

Sandra Stankovic, Stevica Cvetkovic

University of Nis
Faculty of Electronic Engineering
Aleksandra Medvedeva 4
18104 Nis, Serbia
e-mail: s.sandra@elfak.rs, stevica.cvetkovic@elfak.ni.ac.rs

Abstract. Continuous glucose monitoring is essential for effective diabetes management, but most clinically validated systems rely on minimally invasive sensors that require subcutaneous insertion, periodic calibration, and frequent replacement. These limitations create a critical need for non-invasive glucose estimation methods that are accurate, wearable, and deployable on resource-constrained devices. In this study, we propose a lightweight neural network architecture for non-invasive glucose estimation using biomedical signals, including PPG, ECG, and accelerometer (ACC) signals. To address the non-stationary nature of physiological signals, we employ discrete cosine transform (DCT) representations that preserve both temporal and spectral characteristics. The model uses a multi-branch convolutional architecture with anisotropic kernels that independently capture temporal dynamics and spectral correlations, enabling efficient feature extraction while having only 20.5 k parameters. Evaluation on two public datasets demonstrates robust cross-modal performance, achieving over 98 % of predictions within zones A + B of the Clarke Error Grid. These results show that compact neural networks can enable accurate, real-time, non-invasive glucose monitoring on wearable devices.

Keywords: Non-invasive glucose estimation, time-frequency analysis, biomedical signals, TinyML, wearable health devices

Mathematics Subject Classification 2010: 68T05

1 INTRODUCTION

Continuous glucose monitoring is critical for effective diabetes management, enabling timely detection of hyperglycemia and hypoglycemia and guiding therapeutic interventions [1, 2]. Most clinically validated monitoring systems rely on minimally invasive sensors that measure interstitial glucose through subcutaneous probes. While accurate, these devices have inherent limitations, including high cost, user discomfort, periodic calibration requirements, and limited long-term wearability. These challenges have motivated research into non-invasive glucose estimation approaches that leverage physiological signals from wearable devices, which are safer, more convenient, and potentially suitable for continuous ambulatory monitoring [3].

Advances in wearable sensing technologies have enabled the collection of multimodal physiological signals, including photoplethysmography (PPG), electrocardiography (ECG), and accelerometer (ACC) signals. These modalities reflect cardiovascular, autonomic, and behavioral dynamics that are modulated by glucose metabolism [4]. Recent studies applying deep learning to this problem have demonstrated promising predictive performance. However, most existing architectures are computationally intensive, with millions of parameters, making them unsuitable for deployment on low-power microcontrollers typically found in wearable and edge devices. This computational bottleneck represents a key barrier to real-world implementation.

Physiological evidence supports the use of cardiovascular and autonomic signals for glucose estimation. Hypoglycemia is associated with electrocardiographic changes such as QT interval prolongation and altered repolarization dynamics [5]. Similarly, alterations in heart rate (HR) and heart rate variability (HRV) have been observed under both hypoglycemic and hyperglycemic conditions, with hypoglycemia typically associated with increased sympathetic activation and elevated heart rate, while hyperglycemia has been linked to reduced HRV [6]. ECG provides a direct measurement of cardiac electrical activity and remains the clinical reference standard for HR monitoring. In contrast, PPG offers a low-cost, single-sensor optical alternative widely integrated into consumer devices [7]. While ECG and PPG are governed by distinct physiological processes, they are tightly coupled through the cardiac cycle: ECG reflects electrical depolarization and repolarization of the myocardium, whereas PPG captures peripheral blood volume changes modulated by vascular tone, perfusion, and hemodynamics. Glucose concentration influences blood optical properties such as absorption and scattering coefficients [8], and it alters pulse wave morphology and peripheral perfusion patterns [9]. These interactions suggest that PPG signals contain latent glucose information, which can be extracted with appropriate computational methods.

Behavioral and contextual factors further influence glucose dynamics. Physical activity has been shown to increase glucose utilization, while stress-induced hormonal and autonomic activation tends to elevate glucose levels; in contrast, relaxation periods generally lead to a reduction in glucose concentration [10]. Pa-

tients with diabetes exhibit altered autonomic regulation, including elevated baseline HR, attenuated physiological responses, and slower recovery compared with normoglycemic individuals [11]. Incorporating accelerometer-derived activity signals has been shown to improve the prevention of exercise-induced hypoglycemia, highlighting the importance of contextual information beyond cardiovascular dynamics [12]. Despite this, ACC signals remain underexplored as a primary modality for glucose estimation, often treated solely as sources of motion artifacts.

A central challenge in non-invasive glucose estimation from physiological signals is their inherently non-stationary nature, with statistical properties that vary over time due to changes in autonomic regulation, vascular dynamics, motion, and environmental conditions. Conventional approaches based on raw time-domain representations often fail to capture transient frequency-specific patterns, while global spectral methods sacrifice temporal localization by implicitly assuming stationarity. As a result, many existing deep learning solutions rely on high-capacity architectures to implicitly learn these dependencies, often overlooking both the structured nature of the data and the constraints of real-world deployment.

In this work, we address these limitations through a structured and computationally efficient framework that explicitly aligns signal representation and model architecture with the underlying time-frequency characteristics of physiological signals. First, we employ windowed discrete cosine transformation (DCT) representations to encode signals in a joint time-frequency domain, preserving both temporal and spectral information. Second, we introduce a multi-branch convolutional architecture with anisotropic kernels that explicitly decouples temporal and spectral feature extraction. Horizontally oriented kernels capture localized temporal dynamics (e.g., cardiac cycles and motion transients), while vertically oriented kernels model spectral correlations associated with rhythmic physiological processes. This directional decomposition provides an effective inductive bias, enabling more structured feature learning compared to conventional isotropic convolutions.

The proposed design emphasizes computational efficiency and deployability, resulting in a compact model that achieves high clinical accuracy. Unlike prior work that prioritizes predictive performance at the expense of model complexity, the proposed approach demonstrates that accurate glucose estimation can be achieved with lightweight architectures, making it suitable for real-time deployment on resource-constrained wearable devices.

The key contributions of this work are as follows:

Lightweight architecture for edge deployment: A compact convolutional neural network with only 20 513 parameters that achieves over 98 % of predictions within Clarke Error Grid zones A + B.

Anisotropic time-frequency feature extraction: A multi-branch network design employing 1×3 and 3×1 kernels that explicitly separate temporal and spectral patterns in time-frequency representations, improving efficiency and predictive performance.

Hardware validation on edge devices: Feasibility demonstrated via post-training quantization and deployment on a Raspberry Pi 4, achieving low latency, small memory footprint, and high throughput under real-world constraints.

Analysis of time-frequency resolution trade-offs: A systematic study demonstrating that preserving temporal resolution is critical for modeling non-stationary physiological signals, while excessive compression degrades stability.

Cross-dataset multimodal validation: Consistent performance across two public datasets and multiple sensing modalities (PPG, ECG, ACC), confirming generalizability and robustness across sensor locations and acquisition conditions.

Overall, this work bridges a critical gap between signal representation, model design, and deployment constraints, demonstrating that physiologically informed, lightweight architectures can achieve accurate and clinically meaningful non-invasive glucose estimation in real-world wearable settings.

The remainder of this paper is organized as follows. Section 2 reviews the related work, covering traditional machine learning, deep learning, and time-frequencytime-frequency analysis. Section 3 presents the proposed methodology, including preprocessing steps and neural network design. Section 4 describes the implementation details, datasets, and evaluation metrics. Sections 5 and 6 present and discuss the results in detail, while Section 7 concludes the paper with directions for future work.

2 RELATED WORK

The development of non-invasive glucose estimation methods has evolved from traditional statistical models toward deep learning architectures capable of capturing complex, nonlinear relationships in high-dimensional biomedical signals [13, 14].

2.1 From Traditional Machine Learning to Deep Learning Methods

Early research relied heavily on manual feature engineering from PPG and ECG waveforms. Common temporal features include systolic/diastolic areas, peak-to-peak intervals, and pulse arrival times, while spectral features often include power spectral density (PSD) and various frequency-domain ratios [15, 16]. These approaches typically require extensive preprocessing, such as Gaussian filtering and baseline correction [17, 18], followed by dimensionality reduction techniques like ANOVA or mutual information [19, 20]. The resulting feature sets are then mapped to glucose levels using classical regressors, including SVM [21], Random Forest [22], and simple fully connected networks [23, 24]. While interpretable, these methods are limited by their reliance on manual feature extraction, which often fails to capture the subtle, non-stationary nature of physiological signals.

Deep learning has addressed these limitations by enabling automatic hierarchical feature learning. Recurrent neural networks (RNNs) and long short-term memory

(LSTM) models are commonly used for time-series analysis due to their ability to model temporal dependencies [25]. For example, encoder–decoder architectures augmented with LSTM and attention mechanisms have been applied to glucose prediction tasks, achieving RMSE values of 11.86 mg/dL on healthy subjects and 22.01 mg/dL for individuals with type 1 diabetes [26]. Convolutional neural networks (CNNs) have also been widely adopted for time-series analysis. For instance, ResNet models using 20 residual blocks have achieved a MAE of 14.8 mg/dL while demonstrating that shorter time windows can capture rapid glucose dynamics more effectively [27]. Similarly, 1D CNN architectures incorporating convolutional and max-pooling layers, tanh activations, kernel sizes up to 5, and 64–208 channels have been shown to achieve high prediction accuracy, with 95 % of samples exhibiting mean squared errors below 1.1 mmol/L [28].

More recently, transformer-based architectures have emerged as powerful models for time-series prediction [29, 30]. Their multi-head self-attention mechanisms enable the modeling of long-range temporal dependencies without relying on recurrent structures. One notable example is GluTANN [31], which combines input and output embeddings with positional encodings to preserve temporal ordering of glucose measurements, while specialized attention blocks capture paired dependencies and suppress redundant disturbances. Depending on the dataset, this approach achieved RMSE values ranging from 8.27 to 21.85 mg/dL.

2.2 Lightweight Architectures and TinyML

Despite their strong performance, many deep learning models rely on millions of parameters, limiting their applicability in resource-constrained environments. This has motivated the development of lightweight architectures within the TinyML paradigm. Techniques such as depthwise separable convolutions, pruning, and quantization have been widely adopted to reduce computational complexity while maintaining accuracy [32, 33].

For instance, a lightweight convolutional autoencoder-inspired model [34] was proposed for blood glucose estimation from PPG signals. The model processes a 2D matrix of sequential PPG segments with extracted physiological features and uses a compact encoder–decoder structure to learn a compressed representation before regression. Despite containing only 11.8 k parameters, the model was successfully deployed on a microcontroller and achieved a MAE of 5.55 mg/dL.

Another example is a Lightweight Sequential Transformer, which processes input data incrementally rather than analyzing the full observation window at once. By focusing on the most recent and temporally relevant information, the model significantly reduces memory requirements. For a 30-minute prediction horizon, it requires only 0.49 MB of flash memory and 320 KB of RAM, enabling deployment on the microcontrollers used in modern continuous glucose monitors and insulin pumps [35].

2.3 Time-Frequency Analysis of Biomedical Signals

Although these approaches improve efficiency, they often do not fully exploit the inherent structure of biomedical signals. Signals such as PPG and ECG exhibit quasi-periodic behavior and are sampled at relatively low rates, with relevant information distributed across both time and frequency domains. Transient events, oscillatory components, and gradual trends can remain hidden in raw time-series representations. Time-frequency representations provide a more informative characterization by jointly capturing temporal and spectral dynamics.

Prior studies have demonstrated the benefits of transforming ECG or HR signals into two-dimensional representations using methods such as the continuous wavelet transform (CWT) or short-time Fourier transform (STFT). These representations enable the application of standard 2D CNNs, which have shown improved performance over 1D models in tasks such as diabetes diagnosis [36] and arrhythmia detection [37]. Other approaches further enhance performance by combining separate temporal and spectral processing branches [38].

In the context of glucose monitoring, wavelet-based methods have been used for signal decomposition and denoising, followed by dimensionality reduction and regression modeling, achieving RMSE values of 24.27 mg/dL [39]. Additional work has explored multimodal feature extraction and fusion strategies, combining ECG and PPG features using deep learning and classical machine learning models [40].

2.4 Research Gap and Motivation

Collectively, prior work highlights two key directions:

1. the need for lightweight models suitable for deployment on wearable and embedded platforms, and
2. the effectiveness of time-frequency representations for capturing non-stationary physiological dynamics.

However, existing approaches typically address these aspects independently. Lightweight models often rely on time-domain representations, while time-frequency methods are commonly paired with computationally intensive architectures.

In contrast, this work integrates both directions by combining a compact convolutional architecture with a structured time-frequency representation and domain-aligned feature extraction. This unified design enables efficient and accurate glucose estimation while maintaining suitability for real-time deployment on resource-constrained devices.

3 METHODOLOGY

This section presents the proposed framework for non-invasive glucose estimation from multimodal physiological signals, which was designed to achieve high predictive

performance while maintaining a low computational footprint suitable for wearable edge devices. The framework comprises three main stages (Figure 1):

1. temporal segmentation and preprocessing,
2. time-frequency feature extraction using DCT, and
3. regression-based glucose prediction using a lightweight convolutional neural network.

Each stage is optimized to preserve physiologically meaningful dynamics while ensuring computational efficiency.

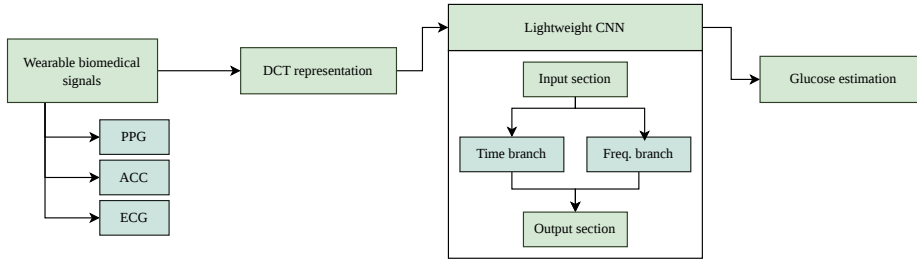

Figure 1. Overview of the proposed non-invasive glucose estimation pipeline. Multimodal physiological signals are transformed into time-frequency representations, processed by a multi-branch lightweight CNN with anisotropic kernels, and output high-accuracy glucose predictions suitable for real-time wearable edge deployment.

To ensure robust performance in unconstrained free-living environments, the preprocessing pipeline is deliberately designed to handle raw signals without an explicit artifact removal stage. Traditional filtering techniques often inadvertently eliminate subtle, low-frequency components and phase relationships that are heavily modulated by autonomic and metabolic changes. Instead of relying on manual denoising steps that risk stripping out latent physiological signatures, our framework shifts the burden of artifact separation directly into the hierarchical feature extraction layers of the neural network.

3.1 Signal Segmentation

Biomedical signals such as PPG and ACC are commonly acquired at sampling frequencies that far exceed the spectral bandwidth of the physiological processes of interest. Glucose dynamics, for example, change over the course of minutes rather than milliseconds, while many autonomic and metabolic variations are primarily contained within low-frequency components. In contrast, high-frequency components are often dominated by cardiac pulsatility, motion artifacts, and sensor noise.

To reduce computational complexity while preserving relevant physiological information, the signals were segmented into fixed-length temporal windows and downsampled to 4 Hz. This sampling rate was sufficient to retain the informative low-frequency content while substantially decreasing the input dimensionality. For each window, the target glucose value was defined as the average of all glucose measurements contained within that interval. Averaging reduces the effect of measurement noise and better reflects the slowly varying behavior of interstitial glucose dynamics.

The downsampled signals are then segmented into overlapping windows to capture localized temporal patterns. Let the window length be L, and the step size be $S = L(1 - O)$, where $O \in [0, 1)$ denotes the overlap ratio. The i^{th} windowed segment is defined as:

$$\mathbf{w}_i = [x_d[iS], x_d[iS + 1], \ldots, x_d[iS + L - 1]]^T. \tag{1}$$

In our experiments, we set $O = 75\,\%$, as we have found that an increased overlap ratio tends to lead to better validation accuracies due to more available training data, while preserving temporal continuity and capturing transient features across the signal.

3.2 Time-Frequency Feature Extraction

Physiological signals are inherently non-stationary, with spectral characteristics that evolve over time due to variations in autonomic tone, motion, vascular properties, and metabolic state. Time-frequency representations therefore provide a more informative characterization than purely time-domain or global frequency-domain analyses [41]. To capture these evolving patterns, we adopt a time-frequency representation based on the DCT, which provides a compact, real-valued spectral decomposition well-suited for efficient and low-complexity implementations.

Unlike Fourier-based representations that produce complex-valued outputs, the DCT yields purely real coefficients while maintaining strong energy compaction properties, making it particularly suitable for embedded and quantized pipelines. While complex-valued CNNs have been proposed [42, 43], we use a real-valued representation because conventional CNNs and optimization methods are more established, computationally efficient, and widely supported, yet still capable of learning equivalent feature representations through multi-branch parallel processing. Each signal is segmented into overlapping or non-overlapping fixed-length windows, and a DCT-II is applied independently within each segment.

For a segment $x[n]$ of length L, the DCT-II is defined as:

$$X[k] = \sum_{n=0}^{L-1} x[n] \cos\left(\frac{\pi}{L}\left(n + \frac{1}{2}\right)k\right), \quad k = 0, 1, \ldots, L - 1. \tag{2}$$

To ensure practical deployment on resource-constrained platforms, the preprocessing pipeline was designed for deterministic low-latency execution and minimal computational overhead. Rather than relying on iterative runtime computations or floating-point trigonometric evaluations, the implementation employs a hardware-conscious optimization strategy based entirely on vectorized integer arithmetic.

First, the DCT-II basis matrix is computed once during system initialization and converted into an 8-bit signed fixed-point representation via scaled integer quantization, significantly reducing execution latency and memory bandwidth requirements. Second, overlapping temporal segments are generated using zero-copy strided windowing operations. Instead of explicitly allocating and copying memory blocks for each sliding window, the implementation constructs virtual overlapping views through memory stride manipulation. This minimizes heap allocation overhead and eliminates unnecessary memory-transfer operations during streaming execution.

3.3 Proposed Network Architecture

We propose a lightweight two-dimensional CNN with a multi-branch time-frequency design and residual fusion, specifically tailored for processing time-frequency physiological signal representations. The architecture (Figure 2) integrates: progressive spatial downsampling, anisotropic multi-branch feature extraction, and residual feature fusion.

This design explicitly incorporates the distinct statistical properties of the temporal and spectral dimensions while maintaining a low computational footprint suitable for edge deployment. The network is parameterized by a scaling factor M, which controls channel depth and overall model complexity. The model builds upon multi-branch convolutional architectures and residual learning frameworks, which are known to improve feature representation and training stability [44]. Processing inputs through parallel convolutional branches enables the extraction of complementary features at different scales, even when the branches are structurally identical, because small differences in initial parameters produce diverse learned representations [45, 46].

To exploit the structure of time-frequency representations, we use anisotropic (direction-specific) convolutional kernels that separately model temporal and spectral dependencies. Horizontal kernels capture temporal dynamics (e.g., heartbeats and motion transients), while vertical kernels capture spectral correlations (e.g., oscillatory patterns and HRV components). This decomposition allows the network to explicitly model temporal and spectral patterns without requiring deeper or more computationally intensive architectures. Residual connections are incorporated to facilitate gradient propagation and stabilize training, enabling the network to refine informative features rather than relearning redundant representations.

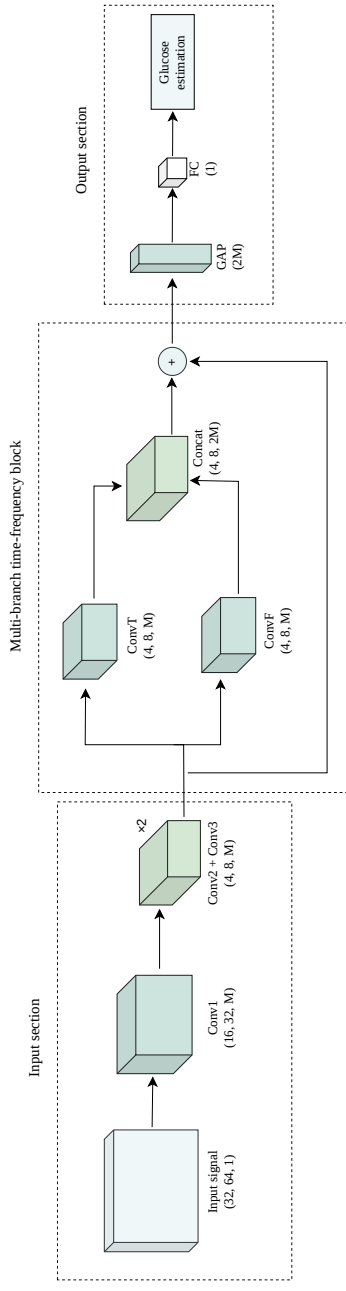

Figure 2. Proposed lightweight 2D CNN for time-frequency input signals, combining strided convolutions for feature extraction, a multi-branch time-frequency block with anisotropic kernels, residual fusion, and global average pooling for final glucose prediction

3.3.1 Input Layers

The network's feature extraction stage consists of three convolutional layers, each applied with a stride of 2 along both time and frequency dimensions. This design has two objectives. First, the initial layers capture localized patterns in the time-frequency plane, which, in physiological spectrograms, may correspond to transient oscillatory bursts or gradual frequency shifts. By stacking multiple layers, the receptive field grows hierarchically, enabling the network to integrate these low-level patterns into more abstract representations. Second, strided convolutions progressively downsample the feature maps, reducing their dimensions by a factor of eight along both axes. This reduces computational and memory requirements while retaining essential information for multi-scale feature extraction [47]. Compared to traditional pooling layers, strided convolutions have been shown to provide more efficient spatial reduction, offering a learnable and content-aware downsampling mechanism [48, 49].

To stabilize training, batch normalization is applied after each convolution, normalizing intermediate feature distributions and mitigating sensitivity to initialization and gradient instability [50]. Nonlinear activation functions (ReLU) follow, allowing the network to model complex, non-linear interactions across time and frequency while maintaining numerical robustness.

3.3.2 Multi-Branch Time-Frequency Block

Following the initial convolutional sequence, the feature map is processed by a multi-branch architecture that explicitly separates temporal and spectral feature extraction. Time-frequency representations of physiological signals have different types of information along each axis. Along the temporal axis, the signal captures changes over time, such as variations in heart rate, autonomic responses, or movement patterns. Along the frequency axis, the signal reflects rhythmic or oscillatory processes, such as periodic components in heart rate variability or vascular pulsations.

Using standard square convolutional kernels treats the time and frequency dimensions the same, applying the same filtering across both axes. While straightforward and commonly used in computer vision, isotropic filters operate under the strict assumption of spatial homogeneity and 2D translation invariance, implying that moving along the horizontal axis represents the same physical transformation as moving along the vertical axis. This assumption fails when applied to physiological spectrograms, which are fundamentally non-isotropic topological spaces governed by distinct physical phenomena along each coordinate axis. Convolving these dimensions with an isotropic square filter forces an artificial coupling between time and frequency, inducing spectral–temporal blurring and imposing an incorrect inductive bias.

Along the temporal axis, the signal exhibits non-stationary translation invariance, capturing localized transient variations, autonomic nervous system fluctuations, and macro-level kinematic artifacts. Conversely, along the spectral axis,

the representation exhibits frequency-scaling invariance, mapping highly correlated, periodic oscillatory structures such as cardiac harmonics, vascular pulsations, and respiration-driven rhythms. Convolving these orthogonal properties with a symmetric, non-separable square kernel forces a joint 2D cross-correlation that artificially couples the axes, inducing spectral-temporal blurring and imposing an isotropic inductive bias on a physically anisotropic space.

To preserve the structural integrity of the signal topology, we propose a decoupled multi-branch architecture that enforces structural separability via parallel, anisotropic (direction-specific) convolutions. This formulation mirrors the mathematical properties of the underlying windowed transforms:

- **Time branch:** uses kernels elongated along the temporal axis but narrow in frequency (1×3). This allows the network to capture local temporal dynamics such as transient changes in heart rate, autonomic responses, or motion-induced variations in the signal.
- **Frequency branch:** uses kernels elongated along the frequency axis but narrow in time (3×1). This focuses on oscillatory or periodic components such as HRV, vascular pulsations, or respiration-related rhythms.

At the end of the multi-branch block, the feature maps from both branches are concatenated along the channel dimension, producing a fused representation that combines temporal and spectral features. To facilitate gradient flow and improve learning efficiency, a residual shortcut connection adds the original input of the multi-branch block to the fused output, as shown in the following expression:

$$Y_{\text{out}} = \text{Concat}(Y_{\text{freq}}, Y_{\text{time}}) + X. \tag{3}$$

3.3.3 Output Layers

To achieve translation invariance across the time-frequency plane and reduce parameter count, global average pooling (GAP) is applied after residual fusion. GAP replaces fully connected spatial flattening with channel-wise averaging. This operation significantly reduces the number of trainable parameters, enhances generalization, and enforces spatial robustness by aggregating distributed evidence across the entire time-frequency representation. Formally, given a feature map $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$, the GAP operation computes:

$$z_c = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{i,j,c}, \quad \text{for } c = 1, \ldots, C. \tag{4}$$

To further mitigate overfitting, a dropout layer (rate 0.4) is applied during training. By randomly deactivating a subset of features, dropout encourages redundancy in learned representations and improves generalization performance, particularly in settings with limited biomedical datasets.

For regression tasks, the final dense layer has a single neuron with a linear activation, producing a continuous output corresponding to the estimated glucose concentration. This streamlined design focuses the network entirely on glucose prediction, leveraging the fused temporal and spectral features without additional classification overhead.

4 EXPERIMENTAL EVALUATION

4.1 Implementation Details

Model development and training were performed using the TensorFlow framework, with subsequent optimization carried out using TensorFlow Lite, which is designed for deployment on resource-constrained edge devices. The experiments were conducted on the Google Colab platform, providing virtual machines equipped with Intel Xeon CPUs operating at 2.2 GHz and 12 GB of RAM.

A batch size of 64 was used, and model parameters were optimized using the AdamW optimizer. The initial learning rate was set to 0.002 and dynamically adjusted using an exponential decay scheduler, reducing the learning rate to 85 % of its current value every three epochs. This schedule enables larger gradient updates during the early training stages, facilitating rapid convergence, while allowing finer adjustments near local optima. To ensure statistical robustness and mitigate overfitting, a 5-fold cross-validation strategy was employed. Final performance metrics are reported as the mean and standard deviation across folds, providing a reliable estimate of real-world generalization performance.

4.2 Datasets

The proposed model was evaluated on two publicly available datasets: D1NAMO and PhysioCGM.

The **D1NAMO dataset** [51] includes physiological recordings from 29 participants, comprising 20 healthy individuals and 9 patients with Type 1 Diabetes (T1D). Data was collected under free-living conditions to capture realistic variability in physiological signals. Each participant wore a Zephyr BioHarness 3 chest strap, which recorded ECG signals at 250 Hz and tri-axial ACC signals at 100 Hz. Additional metrics, including HR and respiratory rate, were derived from the raw signals at 1 Hz. Glucose measurements were obtained using standard continuous glucose monitoring (CGM) devices. For the present study, only recordings from T1D participants were used.

The multimodal **PhysioCGM dataset** [52] consists of recordings from 10 individuals with T1D over periods of up to 17 days in free-living conditions. Participants wore multiple wearable devices simultaneously. A Zephyr BioHarness captured ECG, ACC, HR, and respiratory signals, while an Empatica E4 wristband recorded PPG at 64 Hz, EDA at 4 Hz, and tri-axial ACC at 32 Hz. Continuous

glucose measurements were recorded every 5 minutes using standard CGM devices. This setup provides synchronized multimodal physiological data from both chest- and wrist-worn sensors.

Both datasets consist of raw, unfiltered sensor signals and were not collected under strictly controlled laboratory conditions. The use of publicly available datasets ensures comparability with prior studies while enabling evaluation under realistic, real-world conditions. In this study, only recordings from participants with T1D were included, with target values reflecting real-world glycemic fluctuations ranging from severe hypoglycemia (40 mg/dL) to extreme hyperglycemia (400 mg/dL) across both datasets.

4.3 Evaluation Metrics

To comprehensively assess model performance, we employ complementary metrics capturing both numerical accuracy and clinical relevance.

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE): MAE quantifies the average magnitude of prediction errors without considering their direction, providing an intuitive estimate of deviation between predicted and reference glucose values. RMSE penalizes larger errors more heavily due to the squaring operation, making it more sensitive to outliers. Together, these metrics characterize overall numerical accuracy and error consistency across the glucose range.

Pearson Correlation Coefficient (PCC): PCC measures the linear correlation between predicted and reference glucose values:

$$\mathrm{PCC} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}},$$

where y_i denotes the reference glucose values, $\hat{y}_i$ denotes the predicted values, $\bar{y}$ is the mean of reference values, and $\bar{\hat{y}}$ is the mean of predicted values. The coefficient ranges from -1 to 1, with higher values indicating stronger linear agreement and better tracking of glucose trends over time.

Mean Absolute Relative Difference (MARD): MARD quantifies the average relative error between predicted and reference glucose values:

$$\mathrm{MARD} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\,\%,$$

where y_i and $\hat{y}_i$ denote reference and predicted glucose values, respectively. As a percentage-based metric, it provides an interpretable measure of performance across varying glucose ranges.

Clarke Error Grid Analysis (CEG): CEG evaluates clinical reliability by mapping predicted versus reference glucose values onto five risk zones (A–E). Zone A represents clinically accurate predictions within ±20 % of the reference (or ±20 mg/dL when glucose is below 70 mg/dL). Zone B includes benign errors unlikely to affect treatment decisions, while Zones C–E represent progressively more severe clinical risk, with Zone E indicating potentially dangerous misclassifications.

Together, these metrics provide a comprehensive evaluation of model performance: MAE and RMSE quantify numerical accuracy, PCC assesses trend consistency, MARD evaluates relative error across glucose ranges, and CEG captures clinical safety and reliability.

5 RESULTS

5.1 Effect of Time-Frequency Resolution

We first investigate the influence of the time-frequency input resolution on model performance by varying the allocation between temporal and spectral dimensions while keeping the total input size constant. Experiments were conducted on the D1NAMO dataset using ACC signals, and the results are summarized in Table 1.

Resolution (Freq. × Time)	**MAE (mg/dL)**	**RMSE (mg/dL)**	**PCC**	**MARD (%)**	**CEG A + B (%)**
32×64	24.65 ± 1.99	33.38 ± 1.83	0.90 ± 0.02	18.56 ± 2.01	93.50
64×32	13.54 ± 1.51	17.51 ± 1.82	0.98 ± 0.01	10.29 ± 0.75	98.33
128×16	15.01 ± 2.48	19.05 ± 2.82	0.97 ± 0.01	11.44 ± 2.37	96.89
127×32 (50 % overlap)	13.11 ± 0.82	16.99 ± 1.05	0.98 ± 0.01	9.61 ± 0.38	98.39

Table 1. Influence of time-frequency resolution on regression performance for the ACC signal ($M = 16$, D1NAMO dataset)

The results demonstrate a strong dependence of glucose prediction accuracy on the selected time-frequency representation. Among the non-overlapping configurations, the 64×32 resolution achieved the best overall performance, yielding a MARD of 10.29 ± 0.75 %, together with low MAE and RMSE values and a high correlation coefficient. Compared with the 32×64 and 128×16 configurations, this representation provides a more effective balance between temporal and spectral information, enabling the model to capture both short-term signal variations and discriminative frequency-domain characteristics relevant to glucose dynamics.

Although the overlapping 127×32 configuration achieved slightly better performance across most evaluation metrics, the improvement was relatively marginal considering the increased computational complexity and redundancy introduced by

the overlapping windows. Consequently, the 64×32 representation was selected for the subsequent experiments, as it offers the most favorable trade-off between prediction accuracy, computational efficiency, and model complexity.

5.2 Impact of the Time-Frequency Representation and Anisotropic Kernel Design

To evaluate the contribution of the proposed time-frequency representation and anisotropic convolutional design, we compare the model against several baseline approaches: a lightweight 1D CNN applied directly to raw time-domain signals, a deeper 1D CNN with increased capacity, gradient-boosted tree models trained on raw 1D features, and gradient-boosted trees trained on DCT-transformed representations. In addition, we include time-frequency CNN variants using isotropic 3×3 kernels (single-branch and dual-branch configurations) to isolate the effect of kernel anisotropy.

In contrast, the proposed model uses a multi-branch time-frequency representation with anisotropic 1×3 and 3×1 kernels designed to explicitly decouple temporal and spectral feature extraction. All neural architectures were kept approximately comparable in scale to ensure a fair comparison.

Model	**Params.**	**MAE (mg/dL)**	**RMSE (mg/dL)**	**PCC**	**MARD (%)**	**CEG A + B (%)**
Anisotropic Branches ($1 \times 3 + 3 \times 1$)	5 393	18.24 ± 0.93	23.76 ± 1.21	0.96 ± 0.03	13.21 ± 1.05	97.44
Single Isotropic Branch (3×3)	5 761	18.33 ± 0.41	23.57 ± 0.60	0.95 ± 0.03	14.51 ± 0.73	96.60
Dual Isotropic Branches (3×3)	7 769	18.98 ± 2.08	24.72 ± 2.39	0.95 ± 0.02	13.96 ± 1.15	96.74
1D CNN (light)	8 321	44.60 ± 34.67	52.55 ± 34.81	0.91 ± 0.04	34.46 ± 30.94	83.33
1D CNN (deeper)	107 137	15.38 ± 3.44	19.65 ± 4.10	0.97 ± 0.01	1.85 ± 2.19	97.78
GBTrees (raw)	3 000 splits	58.03 ± 2.06	75.96 ± 2.75	0.13 ± 0.04	48.34 ± 2.20	92.15
GBTrees (DCT)	3 000 splits	38.48 ± 1.06	51.09 ± 1.70	0.80 ± 0.01	31.75 ± 1.43	96.18

Table 2. Performance comparison of the proposed time-frequency model with equivalent 1D time-domain and tree-based models (ACC signals, D1NAMO dataset)

As shown in Table 2, the results highlight a clear interaction between signal representation, model capacity, and inductive bias. The strongest overall performance is achieved by the deeper 1D CNN, indicating that end-to-end time-domain learning remains highly effective when sufficient representational capacity is available. From a signal processing perspective, this suggests that a deep temporal convolutional hierarchy can implicitly approximate both filtering and feature extraction operations across multiple scales, effectively learning a data-driven filter bank directly from raw signals. However, this gain comes at a substantial increase in model complexity, reinforcing the classical trade-off between expressive power and efficiency in purely time-domain approaches.

It is also worth pointing out that deeper variants of our proposed models, obtained by increasing the M parameter, achieve even stronger performance than the 1D CNN while requiring substantially fewer parameters. This indicates that the proposed architecture scales more parameter-efficiently, suggesting that its inductive bias enables more effective utilization of model capacity compared to purely time-domain convolutional approaches.

In contrast, shallow time-domain modeling is insufficient: the lightweight 1D CNN and the XGBoost 1D model fail to capture the nonlinear and nonstationary structure of the physiological signals. The severe failure of the raw-feature tree model (XGBoost 1D, PCC = 0.13) underscores that gradient-boosted trees possess zero structural priors for sequence data; they treat adjacent time steps as independent orthogonal tabular features. Consequently, they cannot inherently compute spatial or temporal invariance, making them highly susceptible to phase shifts, alignment errors, and localized high-frequency noise.

Introducing structured transformations significantly improves gradient-boosted tree performance. In particular, the DCT-based reparameterization maps the sequence into a static frequency-energy spectrum. By optimizing decision trees on fixed orthogonal spectral components rather than raw temporal points, the XGBoost DCT model successfully bypasses its lack of convolutional priors, allowing it to isolate low-frequency macro-trends related to metabolic changes. This confirms that explicit spectral decomposition can partially compensate for the lack of learned hierarchical feature extraction, although it remains fundamentally limited by its rigid, non-adaptive basis representation.

The time-frequency CNN variants further reinforce the importance of multi-domain signal representation. Isotropic convolutional designs already benefit from operating in a joint time-frequency space, suggesting that localized filtering in this domain provides a more informative inductive bias than raw, shallow time-domain processing alone. However, increasing architectural complexity via multiple isotropic branches yields diminishing returns, indicating that redundancy rather than specialization dominates in such configurations.

The proposed anisotropic architecture addresses this limitation by introducing a structured separation of roles across convolutional directions. Rather than treating time-frequency representations as homogeneous 2D textures, the 1×3 and 3×1 kernels impose directional sensitivity: one axis emphasizes local temporal evolution while the other captures spectral structure and cross-frequency interactions. This effectively emulates a factorized filter bank, reducing cross-axis interference and encouraging more interpretable feature decomposition.

Overall, the results suggest that performance gains are not solely driven by increased model capacity, but by how well the architecture aligns with the underlying structure of biomedical signals. Deep 1D CNNs achieve strong performance through implicit hierarchical filtering at the cost of parameter overhead, while the proposed anisotropic time-frequency design achieves highly competitive accuracy (18.24 mg/dL MAE) using only half the parameter envelope (5 393 parameters) by explicitly encoding signal-processing priors into the convolutional layout.

5.3 Effect of Model Capacity

We next analyze the impact of model capacity by varying the channel depth parameter M, which controls the number of convolutional filters in each layer and total parameter count in the model. Increasing M enhances the model's representational capacity, enabling it to capture more complex patterns. However, larger models increase the risk of overfitting, particularly with limited data, and impose higher computational and memory demands.

As shown in Table 3 and Figure 3, performance improves consistently as model capacity increases from $M = 4$ to $M = 16$. The smallest configuration ($M = 4$) exhibits underfitting, with higher MAE/RMSE and lower correlation metrics. Increasing capacity enhances the model's ability to capture complex temporal and spectral patterns of the ACC signal.

M	**Params.**	**MAE (mg/dL)**	**RMSE (mg/dL)**	**PCC**	**MARD (%)**	**CEG A + B (%)**
4	1 481	30.27 ± 1.89	40.60 ± 2.58	0.86 ± 0.02	22.64 ± 2.53	88.80
8	5 393	18.24 ± 0.93	23.76 ± 1.21	0.96 ± 0.03	13.21 ± 1.05	97.44
12	11 737	14.45 ± 1.09	18.77 ± 1.57	0.97 ± 0.01	10.81 ± 0.92	98.09
16	20 513	13.54 ± 1.51	17.51 ± 1.82	0.98 ± 0.01	10.29 ± 0.75	98.33
24	45 361	13.31 ± 1.40	17.22 ± 1.67	0.98 ± 0.01	9.71 ± 0.92	97.47

Table 3. Influence of model parameter count on regression performance (ACC signals, D1NAMO dataset)

The best performance is achieved for $M = 16$, corresponding to approximately 20 k parameters. Further increasing capacity ($M = 24$) yields no additional improvement and slightly degrades performance, indicating diminishing returns and potential overfitting. Notably, even the compact $M = 8$ configuration (approx. 5.5 k parameters) achieves strong performance for glucose estimation, demonstrating that the proposed architecture is highly parameter-efficient and suitable for resource-constrained deployment.

5.4 Evaluation on Edge Devices

To evaluate the feasibility of real-time deployment in resource-constrained environments, the complete signal processing and inference pipeline was benchmarked on a Raspberry Pi 4 Model B equipped with a Broadcom BCM2711 quad-core ARM Cortex-A72 processor operating at 1.5 GHz and running a 64-bit Debian Linux kernel. The models were executed using the optimized AI Edge LiteRT runtime under single-threaded CPU inference to emulate realistic low-power edge conditions without hardware acceleration.

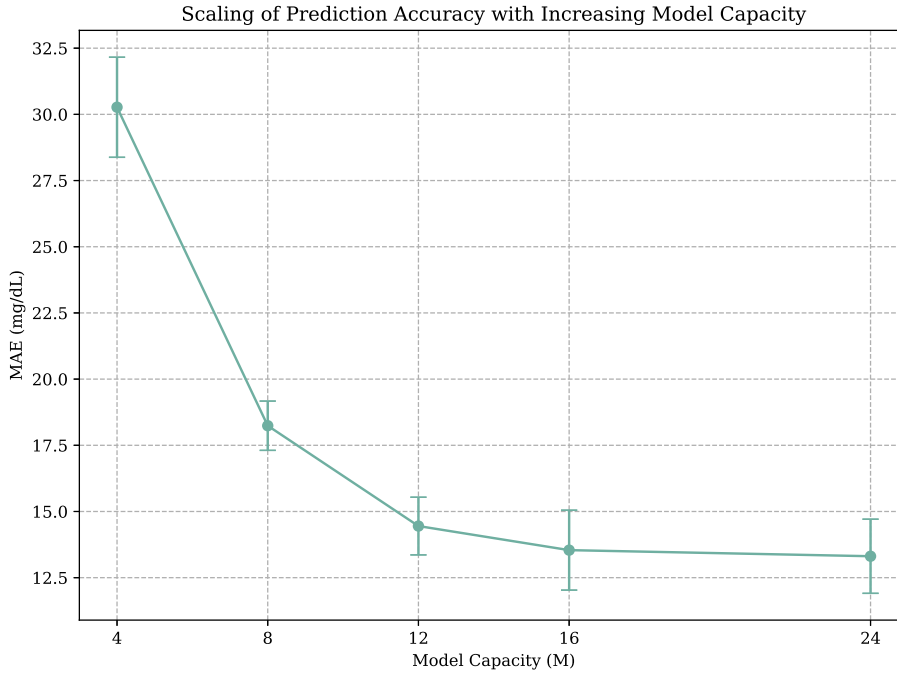

Figure 3. Effect of model capacity (M) on regression performance. Performance improves up to $M = 16$, after which it saturates and slightly degrades at $M = 24$.

Post-training quantization (PTQ) was applied to convert the trained floating-point model weights to 8-bit integer precision, reducing memory footprint and improving inference efficiency with negligible impact on predictive performance.

Table 4 summarizes the runtime characteristics of the compact and higher-capacity configurations. The reported latency reflects the total end-to-end processing cost, including both preprocessing operations and neural network inference, rather than model execution time alone.

M	**Model size (KB)**	**Total latency (ms)**	**Peak RAM (MB)**	**Peak CPU (%)**
8	18.39	1.24	43.23	24.6
16	36.28	1.72	43.26	25.5

Table 4. Performance, parameter count, and end-to-end inference latency comparison across evaluated models on Raspberry Pi 4 device

Despite the increase in representational capacity between configurations, memory utilization remained effectively constant. The peak resident set size (RSS) sta-

bilized at approximately 43.2 MB for both models, corresponding to less than 2.2 % of the available memory on a standard 2 GB embedded system. This indicates that the dominant memory footprint is driven by the runtime environment rather than the model parameters themselves. Additional overhead from signal preprocessing and temporary buffering remained minimal.

In terms of computational demand, processor utilization averaged around 25 % during execution. Importantly, inference occurs only once every 128 seconds ($\sim$ 2.1 minutes), meaning the processing time per inference is negligible relative to the acquisition interval. With an inference latency below 2 ms, the system spends less than 0.002 % of its operational time performing computation, leaving substantial headroom for concurrent system tasks such as wireless communication, sensor acquisition, and background processes.

Overall, these results demonstrate that the proposed pipeline is highly compact in both memory and compute requirements and is well suited for continuous real-time physiological monitoring on low-power embedded hardware.

5.5 Performance Across Physiological Modalities

To evaluate the generalizability of the proposed model across diverse signal modalities, we extend our evaluation to ECG, HR, and HRV signals in order to determine how effectively the model captures glucose-relevant features from different input signals. The results for the D1NAMO dataset and $M = 16$ are summarized in Table 5.

Signal type	**MAE (mg/dL)**	**RMSE (mg/dL)**	**PCC**	**MARD (%)**	**CEG A + B (%)**
ECG	22.19 ± 3.75	28.76 ± 4.85	0.96 ± 0.01	17.40 ± 3.47	94.68
HR	$\mathbf{18.27 \pm 2.26}$	$\mathbf{24.11 \pm 2.67}$	$\mathbf{0.94 \pm 0.03}$	$\mathbf{13.49 \pm 1.62}$	**97.58**
HRV	46.24 ± 3.34	69.78 ± 6.12	0.52 ± 0.02	31.78 ± 2.66	70.75

Table 5. Performance of the proposed model on different physiological signal types ($M = 16$, D1NAMO dataset)

The results show a clear performance gap across physiological modalities. Both ECG and HR provide strong and reliable glucose prediction, with HR achieving the best overall performance in terms of error metrics and clinical agreement, followed closely by ECG. In contrast, HRV consistently underperforms across all metrics, with substantially higher errors and notably lower correlation with ground-truth glucose values.

Overall, HR and ECG demonstrate clinically acceptable and robust predictive behavior, while HRV remains considerably weaker even at higher model capacity. This suggests that increasing model complexity further benefits ECG- and HR-based inputs, whereas HRV does not meaningfully benefit from additional capacity. These results indicate that glucose-related information is more directly reflected in primary

cardiac activity signals, likely due to their stronger coupling with autonomic and cardiovascular responses to glucose fluctuations. In comparison, HRV appears more sensitive to external physiological influences and indirect effects, which may dilute glucose-specific information and limit its predictive utility.

5.6 Cross-Dataset Evaluation

To assess robustness across datasets and sensor modalities, we evaluated the model on the PhysioCGM dataset, which includes both wrist-worn (Empatica E4) and chest-mounted (Zephyr BioHarness) sensors. Performance metrics for $M = 16$ are presented in Table 6.

Device	Signal type	MAE (mg/dL)	RMSE (mg/dL)	PCC	MARD (%)	CEG A + B (%)
Empatica E4 (wrist)	**PPG**	$\mathbf{12.56 \pm 0.69}$	$\mathbf{16.60 \pm 1.04}$	$\mathbf{0.94 \pm 0.01}$	$\mathbf{13.87 \pm 0.61}$	**98.55**
	HR	17.63 ± 1.24	24.19 ± 1.30	0.86 ± 0.02	13.86 ± 1.05	95.86
	EDA	13.84 ± 0.46	18.26 ± 0.58	0.91 ± 0.01	11.13 ± 0.38	98.28
	ACC	13.87 ± 0.61	19.03 ± 1.00	0.91 ± 0.02	10.97 ± 0.41	97.95
Zephyr BioHarness (chest)	ECG	13.49 ± 0.33	17.91 ± 0.38	0.91 ± 0.01	11.20 ± 0.45	98.28
	HR	12.39 ± 0.27	16.37 ± 0.48	0.93 ± 0.01	10.22 ± 0.30	98.90
	ACC	$\mathbf{12.04 \pm 0.45}$	$\mathbf{15.78 \pm 0.64}$	$\mathbf{0.94 \pm 0.01}$	$\mathbf{9.89 \pm 0.33}$	**98.64**

Table 6. Performance of the proposed model on PhysioCGM signals ($M = 16$)

As shown in Table 6 and Figure 4, performance varies across both sensor placement and physiological modality, with a clear overall advantage for chest-mounted sensing. Signals acquired from the Zephyr BioHarness consistently exhibit more stable and reliable predictive behavior than those from the wrist-worn Empatica E4, suggesting that chest placement provides cleaner physiological measurements with reduced susceptibility to motion artifacts and environmental noise.

Within the wrist-based signals, PPG emerges as the most informative modality, with low errors and high correlation. In contrast, HR and EDA appear more sensitive to noise and physiological variability, leading to less stable performance. ACC falls in between, offering moderate predictive utility but still trailing behind PPG in terms of robustness. Overall, this pattern indicates that wrist-based sensing is highly modality-dependent, with optical signals being substantially more reliable than cardiac or electrodermal measurements in this placement.

In the chest-mounted setup, performance is more uniform across modalities, reflecting the greater stability of signals recorded in this region. ACC and HR in particular show strong and consistent predictive behavior, while ECG also maintains solid performance, albeit slightly less competitive than motion- and heart-rate-derived signals. The relatively small variation between modalities suggests that chest-based sensing provides a more robust physiological context overall.

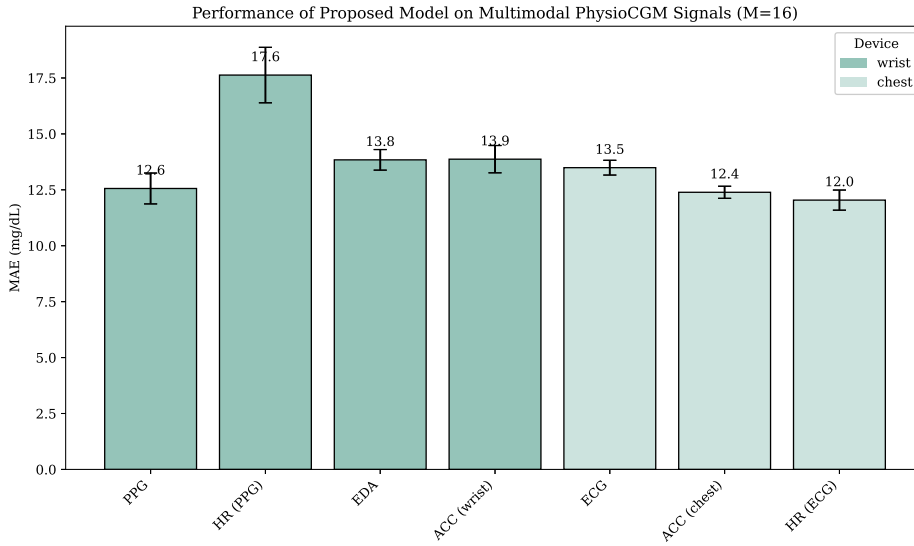

Figure 4. Comparison of model performance across physiological signal types and wearable placements, showing MAE for wrist- and chest-mounted sensors

Taken together, these findings highlight that sensor placement plays a more dominant role than modality choice in determining predictive performance. While certain modalities such as PPG can partially compensate for noisier placements, chest-mounted signals generally provide a more reliable foundation for modeling due to improved signal quality and reduced artifact contamination.

6 DISCUSSION

This study presents a lightweight deep learning framework for non-invasive glucose estimation from multimodal physiological signals, with a particular focus on deployability in wearable edge devices. The results demonstrate that accurate and clinically meaningful glucose prediction can be achieved using compact neural architectures when combined with appropriate signal representations and domain-aligned model design.

6.1 Architectural Design

A key finding of this work is the important role of time-frequency representations in capturing glucose-related dynamics. As shown in Table 2, both raw time-domain and global frequency-domain approaches are insufficient, primarily due to their inability to jointly model temporal localization and spectral structure. In contrast,

the time-frequency representation enables the model to capture transient, non-stationary patterns characteristic of physiological processes influenced by glucose metabolism.

Notably, the degradation in performance observed under substantial reductions in temporal resolution indicates that temporal structure has a slightly more dominant role, particularly for motion-derived signals (Table 1). This suggests that while spectral information is necessary for robust modeling, preserving fine-grained temporal dynamics is especially critical for capturing the full complexity of the underlying physiological processes.

Beyond predictive performance, the time-frequency formulation aligns naturally with convolutional architectures. In time-frequency representations (e.g., 32×64), convolutional filters operate over localized two-dimensional neighborhoods, enabling joint modeling of spectral and temporal dependencies. Spatial downsampling is also more efficient: a stride of 2 across both dimensions reduces feature map size by a factor of four in a single operation, facilitating rapid hierarchical abstraction. In contrast, equivalent 1D representations (e.g., length-2048 vectors) require more sequential operations and provide limited flexibility for learning joint time-frequency patterns, making feature extraction less structured and less parameter-efficient.

Another central contribution of this work is the use of anisotropic convolutional kernels, which explicitly decouple temporal and spectral feature extraction. By enforcing directionally distinct receptive fields, the model is encouraged to learn complementary representations along each axis, rather than relying on generic, isotropic filtering. The empirical results in Table 2 suggest that this structured decomposition provides a stronger inductive bias than conventional isotropic kernels, leading to improved predictive performance even under reduced parameter budgets. This finding further underscores the importance of aligning architectural design with the inherent structure of physiological signals, which exhibit distinct dynamics across time and frequency dimensions.

The analysis of model capacity in Table 3 further supports the suitability of the proposed approach for edge deployment scenarios. Strong performance is already achieved with as few as 5.5 k parameters, while optimal performance is observed around 20 k parameters. Importantly, performance saturates beyond this regime, with additional capacity yielding only marginal gains. This suggests that the proposed architecture is highly parameter-efficient and can effectively capture the most relevant signal structure within a compact representation. Such behavior is particularly desirable for wearable and on-device applications, where strict constraints on memory, computation, and energy consumption must be balanced against the need for robust inference performance.

6.2 Physiological Insights from Multimodal Signals

Evaluation across multiple physiological modalities (Tables 5 and 6) provides insight into the relationship between glucose dynamics and cardiovascular regulation.

Signals directly reflecting cardiac activity, such as ECG and HR, consistently outperform derived measures such as HRV. This suggests that primary cardiovascular signals more directly encode glucose-related autonomic effects, whereas higher-order variability measures may be more susceptible to confounding factors.

This observation is consistent with prior findings showing a positive correlation between glucose levels and heart rate [53]. In contrast, HRV reflects second-order statistics of beat-to-beat intervals and is influenced by numerous confounders, including respiration, psychological stress, and physical activity. As a result, subtle glucose-related effects may be attenuated or obscured in HRV features.

The evaluation in Table 6 highlights the importance of sensor placement and signal quality. Chest-mounted sensors consistently outperform wrist-worn devices, likely due to reduced motion artifacts and closer proximity to central physiological processes. These findings suggest that sensor configuration plays a critical role in the reliability of non-invasive glucose estimation.

Interestingly, accelerometer signals also demonstrate strong predictive capability. In some cases, chest accelerometry even outperforms ECG-based metrics. This suggests that glucose-related physiological changes are not limited to autonomic regulation but also manifest through biomechanical processes, including respiration, posture, and subtle movement patterns. Prior work has shown that hypoglycemia can induce enhanced physiological tremor and impair motor function [22], supporting the hypothesis that glucose fluctuations produce detectable mechanical signatures.

Wrist-worn sensors are more susceptible to motion artifacts and local biomechanical noise, which can degrade signal quality in free-living conditions [54]. In contrast, chest-mounted sensors provide more stable and physiologically representative measurements, capturing central cardiovascular and respiratory dynamics with higher fidelity.

Differences between sensing modalities further support this conclusion. ECG-derived heart rate appears substantially more informative than wrist-based PPG-derived heart rate. ECG directly captures the electrical depolarization of the myocardium, producing sharp R-peaks that enable precise beat-to-beat interval estimation [55]. In contrast, PPG measures peripheral blood volume changes and is more susceptible to motion artifacts and tissue-dependent variability. Additionally, prior research has reported higher rates of inaccuracy in PPG-based heart rate measurements (up to 15 % more frequent) in individuals with darker skin tones, likely due to increased light absorption by melanin [56].

6.3 Limitations and Future Work

Despite these promising results, several limitations should be acknowledged. First, the datasets used in this study, although collected under free-living conditions, are relatively limited in size and population diversity. Second, while the proposed model demonstrates strong cross-dataset generalization, further validation on larger and more heterogeneous cohorts is necessary to confirm clinical applicability.

Although the proposed approach demonstrates high parameter efficiency and robust cross-dataset tracking performance, several challenges must be addressed before deployment in practical wearable medical devices. One issue is the effect of long-term sensor drift and variations at the skin–sensor interface. During prolonged continuous monitoring, physiological sensing systems are susceptible to gradual baseline drift caused by factors such as electrode degradation, sweat accumulation, minor sensor displacement, and changes in skin temperature or tissue impedance.

Future work should also explore broader demographic representation, improved robustness to real-world noise conditions, and integration into fully deployed wearable systems.

7 CONCLUSIONS

In this work, we addressed the critical need for accessible and continuous diabetes management by developing a lightweight, non-invasive framework for glucose estimation from multimodal physiological signals. The proposed approach is explicitly designed to balance predictive performance with computational efficiency, enabling deployment on resource-constrained wearable devices. Unlike conventional deep learning models with high computational costs, the proposed architecture leverages the inherent structure of physiological signals through 2D time-frequency representations. By employing the windowed discrete cosine transform (DCT), the model preserves both temporal and spectral characteristics of non-stationary PPG, ECG, and ACC signals, which are often lost in purely 1D or fully automated feature extraction approaches.

The results demonstrate that accurate and clinically meaningful glucose estimation can be achieved using compact neural architectures when combined with appropriate signal representations and domain-aligned design. The proposed model achieves strong predictive performance, with over 98 % of estimates falling within Zones A and B of the Clarke Error Grid, while maintaining a compact size of approximately 20 k parameters.

Comparative evaluations confirm that the proposed approach consistently outperforms both time-domain and isotropic convolutional baselines, highlighting the effectiveness of the overall framework. Furthermore, cross-dataset validation demonstrates robust generalization across different signal modalities and acquisition settings, supporting its applicability in real-world scenarios.

Overall, this work demonstrates that efficient, non-invasive glucose estimation is feasible using lightweight neural architectures, bridging the gap between algorithmic performance and practical deployment. The proposed framework provides a strong foundation for real-time, wearable glucose monitoring systems. Future work will focus on validation in larger and more diverse populations, as well as further optimization for real-time embedded implementation and personalized modeling.

Acknowledgements

This work has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under Grant No. 451-03-34/2026-03/200102.

REFERENCES

[1] Wu, X.—Zhao, X.—Chen, W.—Chen, Q.—Kong, L.—Li, P.: A Systematic Review of Continuous Glucose Monitoring Sensors: Principles, Core Technologies and Performance Evaluation. Sensors and Actuators Reports, Vol. 10, 2025, Art. No. 100361, doi: 10.1016/j.snr.2025.100361.

[2] Kwon, S. Y.—Moon, J. S.: Advances in Continuous Glucose Monitoring: Clinical Applications. Endocrinology and Metabolism, Vol. 40, 2025, No. 2, pp. 161–173, doi: 10.3803/enm.2025.2370.

[3] Jain, P.—Joshi, A. M.—Mohanty, S. P.—Cenkeramaddi, L. R.: Non-Invasive Glucose Measurement Technologies: Recent Advancements and Future Challenges. IEEE Access, Vol. 12, 2024, pp. 61907–61936, doi: 10.1109/access.2024.3389819.

[4] Min, S.—Geng, H.—He, Y.—Xu, T.—Zhang, X.: Minimally and Non-Invasive Glucose Monitoring: The Road Toward Commercialization. Sensors and Diagnostics, Vol. 4, 2025, No. 5, pp. 370–396, doi: 10.1039/d4sd00360h.

[5] Christensen, T. F.—Tarnow, L.—Randløv, J.—Kristensen, L. E.—Struijk, J. J.—Eldrup, E.—Hejlesen, O. K.: QT Interval Prolongation During Spontaneous Episodes of Hypoglycaemia in Type 1 Diabetes: The Impact of Heart Rate Correction. Diabetologia, Vol. 53, 2010, No. 9, pp. 2036–2041, doi: 10.1007/s00125-010-1802-0.

[6] Nguyen, L. L.—Su, S.—Nguyen, H. T.: Identification of Hypoglycemia and Hyperglycemia in Type 1 Diabetic Patients Using ECG Parameters. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2012, pp. 2716–2719, doi: 10.1109/EMBC.2012.6346525.

[7] Rasti, T.—Zhu, Q.—Zhou, K.—Soni, A.—Zhao, S.—Sleeman, D.: A Summative Survey on ECG and PPG Signal Analysis. International Journal of Computer Applications in Technology, Vol. 73, 2023, No. 2, pp. 148–162, doi: 10.1504/IJCAT.2023.134762.

[8] Rachim, V. P.—Chung, W. Y.: Wearable-Band Type Visible-Near Infrared Optical Biosensor for Non-Invasive Blood Glucose Monitoring. Sensors and Actuators B: Chemical, Vol. 286, 2019, pp. 173–180, doi: 10.1016/j.snb.2019.01.121.

[9] Zavriyev, A. I.—Kaya, K.—Wu, K. C.—Pierce, E. T.—Franceschini, M. A.—Robinson, M. B.: Measuring Pulsatile Cortical Blood Flow and Volume During Carotid Endarterectomy. Biomedical Optics Express, Vol. 15, 2024, No. 3, pp. 1355–1369, doi: 10.1364/boe.507730.

[10] Kajisa, T.—Kuroi, T.—Hara, H.—Sakai, T.: Correlation Analysis of Heart Rate Variations and Glucose Fluctuations During Sleep. Sleep Medicine, Vol. 113, 2024, pp. 180–187, doi: 10.1016/j.sleep.2023.11.038.

[11] Valensi, P.—Extramiana, F.—Lange, C.—Cailleau, M.—Haggui, A.—Maison Blanche, P.—Tichet, J.—Balkau, B.: Influence of Blood Glucose on Heart Rate and Cardiac Autonomic Function. the DESIR Study. Diabetic Medicine, Vol. 28, 2011, No. 4, pp. 440–449, doi: 10.1111/j.1464-5491.2010.03222.x.

[12] Stenerson, M.—Cameron, F.—Wilson, D. M.—Harris, B.—Payne, S.—Bequette, B. W.—Buckingham, B. A.: The Impact of Accelerometer and Heart Rate Data on Hypoglycemia Mitigation in Type 1 Diabetes. Journal of Diabetes Science and Technology, Vol. 8, 2014, No. 1, pp. 64–69, doi: 10.1177/1932296813516208.

[13] Chan, P. Z.—Jin, E.—Jansson, M.—Chew, H. S. J.: AI-Based Noninvasive Blood Glucose Monitoring: Scoping Review. Journal of Medical Internet Research, Vol. 26, 2024, Art. No. e58892, doi: 10.2196/58892.

[14] Jiang, H.—Yao, T.—Ding, C.: PPG-Based Glucose Sensors: A Review. Artificial Intelligence Review, Vol. 58, 2025, No. 12, Art. No. 391, doi: 10.1007/s10462-025-11379-4.

[15] Němcová, A.—Svozilová, V.—Bucsuházy, K.—Smíšek, R.—Mézl, M.—Hesko, B.—Belák, M.—Bilík, M.—Maxera, P.—Seitl, M.—Dominik, T.—Semela, M.—Šucha, M.—Kolář, R.: Multimodal Features for Detection of Driver Stress and Fatigue: Review. IEEE Transactions on Intelligent Transportation Systems, Vol. 22, 2021, No. 6, pp. 3214–3233, doi: 10.1109/TITS.2020.2977762.

[16] Mejía-Mejía, E.—Allen, J.—Budidha, K.—El-Hajj, C.—Kyriacou, P. A.—Charlton, P. H.: Photoplethysmography Signal Processing and Synthesis. In: Allen, J., Kyriacou, P. (Eds.): Photoplethysmography: Technology, Signal Analysis and Applications. 2022, pp. 69–146, doi: 10.1016/b978-0-12-823374-0.00015-3.

[17] Jian, S. E.—Lo, Y. L.—Chuang, Y. T.—Kuo, S. H.: Using Machine Learning to Predict Blood Glucose Level Based on Photoplethysmography. Measurement, Vol. 253, Part A, 2025, Art. No. 117421, doi: 10.1016/j.measurement.2025.117421.

[18] Islam, T. T.—Ahmed, M. S.—Hassanuzzaman, M.—Bin Amir, S. A.—Rahman, T.: Blood Glucose Level Regression for Smartphone PPG Signals Using Machine Learning. Applied Sciences, Vol. 11, 2021, No. 2, Art. No. 618, doi: 10.3390/app11020618.

[19] Ali, N. F.—Elfadel, I. M.—Atef, M.: Multi-Datasets Transfer Multitask Learning for Simultaneous Blood Glucose and Blood Pressure Monitoring Using Common PPG Features. Computers in Biology and Medicine, Vol. 195, 2025, Art. No. 110434, doi: 10.1016/j.compbiomed.2025.110434.

[20] Dave, D.—Vyas, K.—Branan, K.—McKay, S.—DeSalvo, D. J.—Gutierrez-Osuna, R.—Cote, G. L.—Erraguntla, M.: Detection of Hypoglycemia and Hyperglycemia Using Noninvasive Wearable Sensors: Electrocardiograms and Accelerometry. Journal of Diabetes Science and Technology, Vol. 18, 2024, No. 2, pp. 351–362, doi: 10.1177/19322968221116393.

[21] Kermani, A.—Esmaeili, H.: Photoplethysmography-Based Non-Invasive Blood Glucose Estimation Using Systolic-Diastolic Framing MFCC Features and Machine Learning Regression. BioImpacts, Vol. 15, 2025, Art. No. 30589, doi: 10.34172/bi.30589.

[22] Alhaddad, A. Y.—Aly, H.—Gad, H.—Al-Ali, A.—Sadasivuni, K. K.—Cabibihan, J. J.—Malik, R. A.: Sense and Learn: Advances in Wearable Sensing and Machine Learning for Blood Glucose Monitoring and Trend-Detection. Frontiers in Bioengineering and Biotechnology, Vol. 10, 2022, Art. No. 876672, doi: 10.3389/fbioe.2022.876672.

[23] Haque, M. R.—Raju, S. M. T. U.—Golap, M. A. U.—Hashem, M. M. A.: A Novel Technique for Non-Invasive Measurement of Human Blood Component Levels from Fingertip Video Using DNN Based Models. IEEE Access, Vol. 9, 2021, pp. 19025–19042, doi: 10.1109/ACCESS.2021.3054236.

[24] Wang, M.—Liu, X.—Han, W.—Lin, X.—Chen, X.—Zhao, S.—Zhuang, Z.—Zhang, L.—Su, P.: Design of a Wearable Finger PPG-Based Blood Glucose Monitor. Annals of Biomedical Engineering, Vol. 53, 2025, No. 10, pp. 2580–2593, doi: 10.1007/s10439-025-03809-9.

[25] Faouzi, J.: Time Series Classification: A Review of Algorithms and Implementations. In: Rocha, J., Viana, C. M., Oliveira, S. (Eds.): Time Series Analysis – Recent Advances, New Perspectives and Applications. IntechOpen, London, 2024, doi: 10.5772/intechopen.1004810.

[26] Langarica, S.—de la Vega, D.—Cariman, N.—Miranda, M.—Andrade, D. C.—Núñez, F.—Rodriguez-Fernandez, M.: Deep Learning-Based Glucose Prediction Models: A Guide for Practitioners and a Curated Dataset for Improved Diabetes Management. IEEE Open Journal of Engineering in Medicine and Biology, Vol. 5, 2024, pp. 467–475, doi: 10.1109/ojemb.2024.3365290.

[27] Zeynali, M.—Alipour, K.—Tarvirdizadeh, B.—Ghamari, M.: Non-Invasive Blood Glucose Monitoring Using PPG Signals with Various Deep Learning Models and Implementation Using TinyML. Scientific Reports, Vol. 15, 2025, No. 1, Art. No. 581, doi: 10.1038/s41598-024-84265-8.

[28] Hossain, S.—Debnath, B.—Biswas, S.—Al-Hossain, M. J.—Anika, A.—Zaman Navid, S. K.: Estimation of Blood Glucose from PPG Signal Using Convolutional Neural Network. 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), 2019, pp. 53–58, doi: 10.1109/BECITHCON48839.2019.9063187.

[29] Wen, Q.—Zhou, T.—Zhang, C.—Chen, W.—Ma, Z.—Yan, J.—Sun, L.: Transformers in Time Series: A Survey. In: Elkind, E. (Ed.): Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23). International Joint Conferences on Artificial Intelligence Organization, 2023, pp. 6778–6786, doi: 10.24963/ijcai.2023/759.

[30] Nayak, G. H. H.—Alam, M. W.—Avinash, G.—Kumar, R. R.—Ray, M.—Barman, S.—Singh, K. N.—Naik, B. S.—Alam, N. M.—Pal, P.—Rathod, S.—Bisen, J.: Transformer-Based Deep Learning Architecture for Time Series Forecasting. Software Impacts, Vol. 22, 2024, Art. No. 100716, doi: 10.1016/j.simpa.2024.100716.

[31] Xiong, S.—Xu, Y.—Tang, C.—Wang, J.—Liu, S.—Shimada, A.: GluTANN: Transformer-Based Continuous Glucose Monitoring Model with ANN Attention. 2025 IEEE Conference on Artificial Intelligence (CAI), 2025, pp. 543–548, doi: 10.1109/cai64502.2025.00099.

[32] Rokh, B.—Azarpeyvand, A.—Khanteymoori, A.: A Comprehensive Survey on Model Quantization for Deep Neural Networks in Image Classification. ACM Transactions on Intelligent Systems and Technology, Vol. 14, 2023, No. 6, Art. No. 97, doi: 10.1145/3623402.

[33] Wei, L.—Ma, Z.—Yang, C.—Yao, Q.: Advances in the Neural Network Quantization: A Comprehensive Review. Applied Sciences, Vol. 14, 2024, No. 17, Art. No. 7445, doi: 10.3390/app14177445.

[34] Ali, N. F.—Aldhaheri, A.—Wodajo, B.—Alshamsi, M.—Alshamsi, S.—Atef, M.: Non-Invasive Continuous Real-Time Blood Glucose Estimation Using PPG Features-Based Convolutional Autoencoder with TinyML Implementation. 2024 IEEE International Symposium on Circuits and Systems (ISCAS), 2024, pp. 1–5, doi: 10.1109/ISCAS58744.2024.10558453.

[35] Barbato, M. P.—Rigamonti, G.—Marelli, D.—Napoletano, P.: Lightweight Sequential Transformers for Blood Glucose Level Prediction in Type-1 Diabetes. IEEE Journal of Biomedical and Health Informatics, Vol. 30, 2026, No. 6, pp. 5465–5475, doi: 10.1109/JBHI.2025.3633194.

[36] Yildirim, O.—Talo, M.—Ay, B.—Baloglu, U. B.—Aydin, G.—Acharya, U. R.: Automated Detection of Diabetic Subjects Using Pre-Trained 2D-CNN Models with Frequency Spectrum Images Extracted from Heart Rate Signals. Computers in Biology and Medicine, Vol. 113, 2019, Art. No. 103387, doi: 10.1016/j.compbiomed.2019.103387.

[37] Lilda, S. D.—Jayaparvathy, R.: Enhancing Cardiovascular Disease Classification in ECG Spectrograms Using Multi-Branch CNN. Computers in Biology and Medicine, Vol. 186, 2025, Art. No. 109737, doi: 10.1016/j.compbiomed.2025.109737.

[38] Kwon, S. Y.—Park, D. H.—Kim, H. N.: IPFSCNN: A Time–Frequency Fusion CNN for Wideband Spectrum Sensing. Sensors, Vol. 25, 2025, No. 23, Art. No. 7134, doi: 10.3390/s25237134.

[39] Shankar, A.—Chang, S.—Wang, X.—Zhao, Y.: Wavelet-Based Machine Learning Approaches Towards Precision Medicine in Diabetes Mellitus. The FASEB Journal, Vol. 36, 2022, No. S1, doi: 10.1096/fasebj.2022.36.S1.R6003.

[40] Li, J.—Ma, J.—Omisore, O. M.—Liu, Y.—Tang, H.—Ao, P.—Yan, Y.—Wang, L.—Nie, Z.: Noninvasive Blood Glucose Monitoring Using ECG and PPG Feature Fusion and Weight-Based Choquet Integral Multimodel Approach. IEEE Transactions on Neural Networks and Learning Systems, Vol. 35, 2024, No. 10, pp. 14491–14505, doi: 10.1109/TNNLS.2023.3279383.

[41] Hammond, J. K.—White, P. R.: The Analysis of Non-Stationary Signals Using Time-Frequency Methods. Journal of Sound and Vibration, Vol. 190, 1996, No. 3, pp. 419–447, doi: 10.1006/jsvi.1996.0072.

[42] Chakraborty, M.—Aryapoor, M.—Daneshtalab, M.: Frequency Domain Complex-Valued Convolutional Neural Network. Expert Systems with Applications, Vol. 295, 2026, Art. No. 128893, doi: 10.1016/j.eswa.2025.128893.

[43] Yadav, S.—Jerripothula, K. R.: FCCNs: Fully Complex-Valued Convolutional Networks Using Complex-Valued Color Model and Loss Function. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 10655–10664, doi:

10.1109/ICCV51070.2023.00981.

[44] Zhang, H.—Shao, J.—Salakhutdinov, R.: Deep Neural Networks with Multi-Branch Architectures Are Less Non-Convex. 2018, doi: 10.48550/arXiv.1806.01845 (arXiv Preprint arXiv:1806.01845).

[45] Ahmed, K.—Torresani, L.: Connectivity Learning in Multi-Branch Networks. 2017, doi: 10.48550/arXiv.1709.09582 (arXiv Preprint arXiv:1709.09582).

[46] Cvetkovic, S.—Stankovic, S.—Stancic, G.: Modular Deep Residual Network for Driver Stress Detection Based on Photoplethysmography. Journal of Electrical Engineering, Vol. 76, 2025, No. 4, pp. 344–355, doi: 10.2478/jee-2025-0035.

[47] Nirthika, R.—Manivannan, S.—Ramanan, A.—Wang, R.: Pooling in Convolutional Neural Networks for Medical Image Analysis: A Survey and Empirical Study. Neural Computing and Applications, Vol. 34, 2022, No. 7, pp. 5321–5347, doi: 10.1007/s00521-022-06953-8.

[48] Ayachi, R.—Afif, M.—Said, Y.—Atri, M.: Strided Convolution Instead of Max Pooling for Memory Efficiency of Convolutional Neural Networks. In: Bouhlel, M. S., Rovetta, S. (Eds.): Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18), Vol. 1. Springer, Cham, Smart Innovation, Systems and Technologies, Vol. 146, 2019, pp. 234–243, doi: 10.1007/978-3-030-21005-2_23.

[49] Springenberg, J. T.—Dosovitskiy, A.—Brox, T.—Riedmiller, M.: Striving for Simplicity: The All Convolutional Net. 2015, doi: 10.48550/arXiv.1412.6806 (arXiv Preprint arXiv:1412.6806).

[50] Ioffe, S.—Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Bach, F., Blei, D. (Eds.): Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR), Vol. 37, 2015, pp. 448–456, https://proceedings.mlr.press/v37/ioffe15.html.

[51] Dubosson, F.—Ranvier, J. E.—Bromuri, S.—Calbimonte, J. P.—Ruiz, J.—Schumacher, M.: The Open D1NAMO Dataset: A Multi-Modal Dataset for Research on Non-Invasive Type 1 Diabetes Management. Informatics in Medicine Unlocked, Vol. 13, 2018, pp. 92–100, doi: 10.1016/j.imu.2018.09.003.

[52] Quamer, W.—Tseng, M. R.—Vyas, K.—Dave, D.—Villegas, C.—McKay, S.—DeSalvo, D. J.—Erraguntla, M.—Cote, G.—Gutierrez-Osuna, R.: A Multimodal Physiological Dataset for Non-Invasive Blood Glucose Estimation. Scientific Data, Vol. 12, 2025, Art. No. 1822, doi: 10.1038/s41597-025-06090-6.

[53] Fergencs, T.—Illés, D.—Pilawka, O.—Meier, F.: Human-Centered Humanities: Using Stimulus Material for Requirements Elicitation in the Design Process of a Digital Archive. Digital Humanities in the Nordic and Baltic Countries Publications, Vol. 3, 2020, No. 1, pp. 235–246, doi: 10.5617/dhnbpub.11193.

[54] Castaneda, D.—Esparza, A.—Ghamari, M.—Soltanpur, C.—Nazeran, H.: A Review on Wearable Photoplethysmography Sensors and Their Potential Future Applications in Health Care. International Journal of Biosensors & Bioelectronics, Vol. 4, 2018, No. 4, pp. 195–202, doi: 10.15406/ijbsbe.2018.04.00125.

[55] Coste, A.—Millour, G.—Hausswirth, C.: A Comparative Study Between ECG- and PPG-Based Heart Rate Sensors for Heart Rate Variability Measurements: Influence of Body Position, Duration, Sex, and Age. Sensors, Vol. 25, 2025, No. 18, Art. No. 5745, doi: 10.3390/s25185745.

[56] Bent, B.—Goldstein, B. A.—Kibbe, W. A.—Dunn, J. P.: Investigating Sources of Inaccuracy in Wearable Optical Heart Rate Sensors. npj Digital Medicine, Vol. 3, 2020, Art. No. 18, doi: 10.1038/s41746-020-0226-6.

Sandra Stankovic is Junior Researcher at the Faculty of Electronic Engineering, University of Nis, Serbia. She earned her B.Sc. in electrical and computer engineering in 2024, followed by her M.Sc. in 2025, the same year she started her doctoral studies. Her research focuses on the development of efficient and lightweight deep learning models, digital signal processing, biomedical signal analysis, and image processing.

Stevica Cvetkovic is Associate Professor at the Faculty of Electronic Engineering, University of Nis. He received his M.Sc. in computer science in 2007 and subsequently engaged in research collaborations with the Polytechnic University of Madrid and the Technical University of Munich. He earned his Ph.D. in 2018 in the field of computer vision. His research interests include deep learning, computer vision, and generative artificial intelligence. He has authored over 70 scientific publications, including 11 articles in leading international journals.