Evaluation of CNN Models Based on Photoplethysmography Signals Using Optimal Moving Average Filter for Non-Invasive Diabetic Classification

Zulhelmi Zulhelmi [1,2], Roslidar Roslidar [2], Sawal H. Md Ali [3], and Nasaruddin Nasaruddin [2,*]
[1] Doctoral Programs of Engineering, Postgraduate School, Universitas Syiah Kuala, Banda Aceh, Indonesia
[2] Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia
[3] Department of Electrical, Electronic and Systems Engineering, Faculty of Eng. and Built Environ., Universiti Kebangsaan, Malaysia
Email: zulhelmi@usk.ac.id (Z.Z.); roslidar@usk.ac.id (R.R.); sawal@ukm.edu.my (S.H.M.A.); nasaruddin@usk.ac.id (N.N.)

Manuscript received March 7, 2026; revised May 7, 2026; accepted May 16, 2026
*Corresponding author

Abstract—**The invasive glucose measurement and classification methods offer high diagnostic accuracy; however, they have several significant drawbacks, including patient discomfort and elevated costs due to the frequent use of glucose test strips. Consequently, previous research has explored non-invasive alternatives, which are more convenient and cost-effective. A prominent non-invasive approach involves the use of Photoplethysmography (PPG) signals processed with deep learning models to estimate blood glucose levels and classify diabetic status, but it still faces challenges in achieving optimal diagnostic accuracy. Therefore, this paper proposes a diabetic classification framework and evaluates the performance of CNN models in classifying diabetic and non-diabetic subjects using PPG signals. The research method employs a computer simulation, starting with the modification of PPG signals into five distinct datasets using a moving average filter that varies the number of points (M). The CNN models' performance is evaluated using transfer learning with pre-trained weights from ImageNet, utilizing architectures from the ResNet, EfficientNet, and MobileNet series. Each model is trained and tested separately to classify diabetic and non-diabetic. The results indicate that the size of the moving average window M is a crucial factor in determining model performance. Utilizing a dataset filtered with an average of $M = 10$ points, all evaluated architectures—including ResNet (34, 50), EfficientNet (V2-M, Lite), and MobileNet (V2-S, V3-L)—achieved perfect classification performance, yielding 100% accuracy, precision, recall, and F1-score across all test parameters. These findings indicate that the proposed model with $M = 10$ can effectively classify individuals as diabetic or non-diabetic, providing a promising non-invasive alternative for glucose monitoring.**

Index Terms—**convolutional neural network, deep learning, diabetes, non-invasive, photoplethysmography signals, moving average filter**

I. Introduction

Diabetes is a condition that impairs the body's ability to convert food into energy, often leading to severe complications such as heart disease, blindness, kidney failure, and other health issues [1]. The International Diabetes Federation's Diabetes Atlas underscores the escalating prevalence and burden of diabetes worldwide, drawing attention to its profound implications for public health systems and global healthcare policy. Recent estimates indicate that approximately 589 million adults aged 20–79 years are currently living with diabetes, reflecting the scale of this global health challenge. Moreover, diabetes was responsible for 3.4 million deaths in 2024, highlighting its severe impact on morbidity and mortality. By documenting these trends, the Atlas emphasizes the urgent need for improved monitoring strategies, preventive interventions, and innovative management approaches to mitigate the growing impact of this chronic disease [2]. The surge in diabetes cases is largely driven by a failure to address fundamental lifestyle habits. Effectively managing diabetes and prediabetes depends on several pillars: balanced nutrition, regular exercise, stress management, restorative sleep, strong social ties, and the avoidance of harmful substances. When these areas are overlooked, the body's ability to regulate blood sugar falters, drastically increasing the risk of metabolic disorders [3]. Most people are unaware of the symptoms of pre-diabetes, which, if untreated, can progress to diabetes. Preventing diabetes is essential, as no cure currently exists. Diabetes management primarily involves preventing further complications from associated health issues. Preventive measures include maintaining a balanced diet, engaging in regular exercise, and routinely monitoring blood sugar levels.

Diabetes can be identified by measuring blood glucose concentration. Normal individuals have concentrations in the range of 70 to 100 mg/dl for fasting blood glucose tests or less than 200 mg/dl for random tests; concentrations beyond these values are categorized as diabetes [4]. Since blood glucose control and diabetes diagnosis are critical, routinely monitoring blood glucose levels becomes imperative. Capillary blood glucose measurement is typically performed using an invasive finger-prick technique. This method requires obtaining a small drop of blood from the fingertip, which is then applied to a disposable test strip for immediate glucose analysis using a portable glucometer. Despite its widespread use and clinical reliability, the procedure is minimally invasive and may cause discomfort, highlighting the ongoing interest in developing non-invasive monitoring alternatives [5]. Although blood glucose meters provide reliable results, they require frequent blood sampling throughout the day for continuous glucose monitoring. This frequent sampling can lead to skin irritation, potentially causing additional adverse effects [6]. Consequently, the invasive method is not reliable for classification and continuous blood glucose monitoring.

On the other hand, the non-invasive approach offers subjects a low-cost test and a convenient procedure to detect and estimate the range of blood sugar levels in order to classify whether a subject is normal or diabetic. Despite the benefits, a non-invasive algorithm-based tool has not yet been introduced to the market since the methods have not received medical certification for blood glucose measurement [7]. Therefore, many scientists have been conducting research in this area in order to improve the methods. Several studies related to the non-invasive technique have been published [8–10]. The techniques in these studies utilize biosensors for measurement, which requires periodic calibration, and this is a drawback of the approaches. Researchers also utilized radio wave technology in blood glucose measurement; however, it is generally at a fundamental research stage [11–13]. In addition, the use of ECG signals to determine blood glucose concentration has been applied by medics; however, the system is also inconvenient since it requires a pair of electrodes on people with diabetes [14]. Optical spectroscopy approaches are techniques that have been widely used in recent years [15–19] including employing PPG signals [20–24].

The integration of DL with PPG signal analysis provides a pathway for non-invasive diabetes classification, as it reveals intricate vascular signatures not readily discernible using conventional techniques. Convolutional Neural Network (CNN) models have commonly become a promising choice in various deep learning applications, including PPG signal processing for diabetes classification. CNNs provide an efficient approach to data processing and can learn complex patterns within PPG signals. However, using CNN algorithms for PPG signal classification still presents certain challenges. Recent studies indicate that while CNNs simplify the classification process, the accuracy achieved remains suboptimal. For instance, research by [25] reported an accuracy of 76.34%, which remains considerably low. Therefore, advancing and evaluating the CNN algorithm to enhance accuracy in non-invasive blood glucose classification has become a pressing need. Existing PPG-based CNN studies rarely investigate how signal smoothing parameters influence model generalization and computational feasibility for embedded deployment. The selection of filtering parameters is often arbitrary and lacks systematic evaluation.

This paper proposes modifying and reconstructing the PPG-BP dataset into multiple datasets, applying a Moving-Average (MAV) filter with varying averaging window sizes, and using the new filtered datasets to evaluate CNN models such as ResNet, EfficientNet, and MobileNet for classifying diabetic and non-diabetic subjects. The research also highlights the appropriate CNN models that can be implemented in low-performance machines such as microcontroller devices and Raspberry Pi. The following are the main contributions of this work:

(1) We modify the PPG-BP dataset to evaluate the impact of Moving Average (MAV) filtering on CNN performance by systematically varying the number of filter points (M). This process identifies an optimal filtering configuration that balances high classification accuracy with computational efficiency. Additionally, this refined preprocessing approach mitigates inherent noise sensitivities in PPG signals, establishing a more robust benchmark for deep learning-based diabetic classification.

(2) We develop diabetic classification models by leveraging transfer learning from ResNet, EfficientNet, and MobileNet architectures pretrained on ImageNet to address the challenges of a limited dataset. This initialization strategy facilitates faster convergence and enhances model generalization, ensuring robust performance across diverse PPG-based inputs.

(3) We provide recommendations on the suitability of specific CNN models for deployment on resource-constrained hardware, including microcontrollers, Raspberry Pi, and other Single-Board Computers (SBCs), for PPG-based diabetic classification.

This paper is organized into four sections: Introduction, Methods of Research, Results and Discussion, and Conclusion. In the Introduction, we provide a detailed overview of the background, problem, and urgency of the research. In the second section, the method used in the research is followed by results and discussion. At the end of the text, we conclude the work.

II. Methods for Datasets, CNN Architectures, and Performance Evaluation

In this section, the processes of dataset preparation, MAV filter procedure, and model evaluation are described. The proposed evaluation of CNN models is explained, covering several algorithms for a diabetic classification task. Finally, the discussion ends with model performance evaluation using learning curves and evaluation metrics. The proposed framework for the evaluation of CNN models is depicted in Fig. 1.

A. Datasets

We used the PPG-BP database, which was downloaded from the Figshare Repository, as our baseline dataset. The dataset was provided in [26] and have been released in Scientific Data [27]. The PPG signal data were obtained from an open-source dataset comprising 219 adult subjects aged 20 to 89. The cohort includes 38 subjects diagnosed with diabetes, 28 of whom presented with comorbid hypertension, while the remaining subjects were healthy controls.

This dataset integrates physiological signals with clinical records of chronic conditions, all collected under standardized experimental protocols to ensure consistency and reliability. The dataset is particularly valuable for exploring the relationship between PPG waveforms and cardiovascular diseases. The dataset can be used to uncover and analyze latent characteristics within PPG signals, offering insights into cardiovascular health, as in previous researches [28, 29]. This dataset also supports research into early, noninvasive screening for common cardiovascular diseases, including hypertension and related conditions like diabetes. The PPG waveforms contain inherent noise, which must be mitigated to improve diagnostic accuracy, particularly for conditions like diabetes.

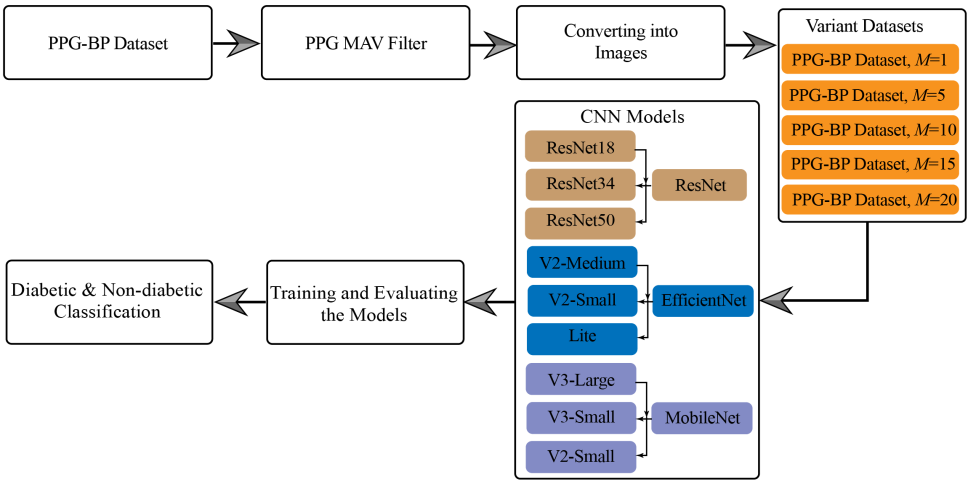


Fig. 1. The proposed diabetic classification framework uses various CNN models using the MAV filter PPG-BP dataset.

B. The PPG Signal and MAV Filter

The key aspect of evaluating CNN models depends on various PPG-BP datasets generated by applying the MAV filter with different averaging points (*M*). In our study, this process resulted in a total of five extended datasets, as shown in Fig. 1. Previous research has explored numerous filtering methods that effectively remove noise [29, 30]. Our purpose is to select a method that is computationally feasible for hardware implementation. Therefore, we chose the MAV filter, as in [30],

$$y[n] = \frac{1}{M}\sum_{k}^{M-1} x[n-k] \tag{1}$$

In Eq. (1), M denotes the window size (number of points), while $x[n]$ and $y[n]$ represent the input and output signals at the discrete-time index n, respectively. The variable k serves as the summation index representing the temporal lag for each data point within the moving window, where $k = 0, 1, \ldots, M-1$. To evaluate the various models discussed previously, filtering configurations of $M = 1, 5, 10, 15,$ and 20 were employed.

The original PPG-BP signals were used as the baseline dataset before MAV filtering. Each signal segment was converted into a waveform image using a standardized plotting procedure. The MAV-filtered versions were then generated using window sizes of $M = 5, 10, 15,$ and 20 to evaluate the influence of signal smoothing on CNN classification performance.

To achieve reproducibility and consistency in the signal-to-image conversion process, each segment of the PPG signal was processed using the same plotting procedure outlined in Algorithms 1 and 2. After applying the MAV filter with the selected window size, each signal segment consisting of 2100 sampled data points was normalized using min–max normalization to scale the amplitude values into a comparable range. The normalized PPG signal was then plotted using Matplotlib with identical figure size, axis limits, line width, and sampling range for all samples. This procedure ensured that differences observed in the generated images were primarily due to signal morphology rather than variations in plotting scale. Each plot was saved as an image file with a resolution of 600 dpi and subsequently used as input for the CNN models. The same conversion procedure was applied consistently to all datasets, including the unfiltered dataset and the MAV-filtered datasets with $M = 5, 10, 15,$ and 20.

Algorithm 1: PPG signal processing and image plotting for diabetic subjects

```
BEGIN
 LOAD data from 'diabetes_data_ppg.csv'

 SET x_axis = sample numbers (SamplesNo)
 DEFINE mav_filter(signal, M):
  FOR each index i in signal:
   IF i >= length(signal) - M:
    output_value = previous output
   ELSE:
    output_value=average (signal[i: i+M])
     APPEND output_value to output_array
   RETURN output_array

 DEFINE save_plot(base_file, index):
  filename = base_file+ "_" + index + ".png"
```

```
    SAVE the current plot to the directory
    PRINT confirmation message

  FOR sbjct_idx = 1 TO 38:
   signal = data column[sbjct_idx]
   filtered_signal = mav_filter(signal, M)
    CREATE new plot (size 10x6)
    PLOT x_axis vs filtered_signal
    CALL save_plot("Diabetes", sbjct_idx)
    CLOSE plot
END
```

Algorithm 2: PPG signal processing and image plotting for non-diabetic subjects

```
BEGIN
  LOAD data from 'non-diabetes_data_ppg.csv'
  SET x_axis = sample numbers (SamplesNo)
  DEFINE mav_filter(signal, M):
   FOR each index i in signal:
    IF i >= length(signal) - M:
     output_value = previous output
    ELSE:
     output_value=average (signal[i: i+M])
      APPEND output_value to output_array
    RETURN output_array

  DEFINE save_plot(base_file, index):
   filename = base_file+ "_" + index + ".png"
    SAVE the current plot to the directory
    PRINT confirmation message

  FOR sbjct_idx = 1 TO 181:
   signal = data column[sbjct_idx]
   filtered_signal = mav_filter(signal, M)
    CREATE new plot (size 10x6)
    PLOT x_axis vs filtered_signal
    CALL save_plot("non-diabetes", sbjct_idx)
    CLOSE plot
END
```

C. Variant Datasets

This stage creates datasets for evaluating CNN models, with the baseline dataset, referred to as the PPG-BP database or unfiltered dataset (PPG-BP Dataset-M1), processed using the MAV window size M set to 5. The resulting output signals are then collected, converted into image files, and stored in a new dataset, labeled PPG-BP Dataset-M5, where the suffix M5 denotes the use of $M = 5$ in the MAV filter. The same procedure is repeated to generate additional datasets, PPG-BP Dataset-M10, PPG-BP Dataset-M15, and PPG-BP Dataset-M20, by applying the MAV filter with M values of 10, 15, and 20, respectively. As a result, we have one baseline dataset, PPG-BP Dataset-M1, and four filtered datasets, each labeled according to the M values of the MAV filter, as shown in Table I. The datasets are prepared and fed into deep learning models for evaluation, where each dataset is divided into training, validation, and testing at an 80:10:10 ratio.

TABLE I: GENERATING DATASETS BY APPLYING THE MAV FILTER

MAV Filter (M)	Dataset	Note
1	PPG-BP Dataset-M1	unfiltered dataset
5	PPG-BP Dataset-M5	Filtered dataset
10	PPG-BP Dataset-M10	Filtered dataset
15	PPG-BP Dataset-M15	Filtered dataset
20	PPG-BP Dataset-M20	Filtered dataset

While PPG signals are fundamentally one-dimensional (1D) time-series data, this study utilizes an image-based representation to leverage transfer learning with established 2D CNN models that have been pretrained on ImageNet. This strategy is particularly useful for a limited dataset, as pretrained CNNs provide robust initial feature representations and can improve convergence during training. In addition, converting PPG segments into waveform images preserves relevant morphological characteristics, such as peak shape, amplitude variation, periodicity, and smoothness after MAV filtering. These visual features can be captured by 2D convolutional kernels and used to distinguish diabetic and non-diabetic patterns. Furthermore, the use of 2D CNNs allows direct evaluation of lightweight pretrained architectures, such as EfficientNet-Lite and MobileNet-V2-Small, which are widely available and suitable for resource-constrained deployment. Nevertheless, we acknowledge that 1D CNNs are generally more computationally efficient for raw time-series processing, and comparative evaluation between 1D CNN and 2D CNN approaches is recommended for future work.

D. CNN Architectures

Various CNN architectures have been extensively utilized to develop deep learning models for mobile and edge computing applications [31, 32]. These architectures are designed to prioritize hardware efficiency and minimize operational complexity, ensuring high-performance execution on resource-constrained devices. Furthermore, the real-world feasibility of these models has been validated through implementation on high-performance edge platforms, such as the NVIDIA Jetson Nano and Raspberry Pi [33].

The selected architectures, in this study, emphasize computational efficiency, balancing key metrics such as floating-point operations (FLOPs) and parameter count. To meet these criteria, we focus on lightweight variants of ResNet, EfficientNet, and MobileNet, which are known for their low FLOPs and minimal parameter requirements. The FLOPs and parameter count for the selected models are summarized in Table II.

TABLE II: PARAMETERS AND FLOPS OF VARIOUS CNN MODELS

Model	Parameters (~ M)	FLOPs (~ G)	Category
ResNet18 [34]	11.7	1.8	Deep CNN
ResNet34 [34]	21.8	3.6	Deep CNN
ResNet50 [34]	26	4.1	Deep CNN
EfficientNet-V2-M [35]	54.1	24.7	Efficient CNN
EfficientNet-V2-S [35]	22	8.4	Efficient CNN
EfficientNet-Lite [35]	5	0.511	Tiny CNN models
MobileNet-V3-L [31]	5.4	0.219	Lightweight CNN
MobileNet-V3-S [31]	2.5	0.066	Lightweight CNN
MobileNet-V2-S [32]	3.4	0.3	Lightweight CNN

1) ResNet Architecture

The ResNet architecture, introduced by Kaiming He and colleagues [34], is renowned for its residual blocks, which enable the training of extremely deep networks by mitigating the vanishing gradient problem. Fig. 2

illustrates the architecture of a 34-layer Residual Neural Network (ResNet34), a deep convolutional model widely employed in image classification and biomedical signal analysis. The network begins with a 7×7 convolutional layer followed by a pooling operation, progressively extracting hierarchical features through stacked 3×3 convolutional layers. Residual blocks—depicted as shortcut connections—enable identity mappings that bypass one or more layers, facilitating gradient flow and mitigating vanishing gradient issues during backpropagation. This design allows the network to maintain high representational capacity while preserving training stability, even at substantial depth.

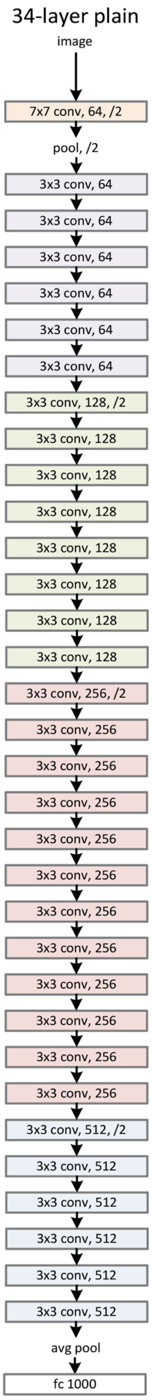


Fig. 2. The ResNet34 architecture [34].

In biomedical applications, such as diabetic classification using PPG signals, ResNet34 offers a robust framework for capturing subtle temporal and morphological variations. The residual connections enhance feature learning by allowing the model to focus on discriminative patterns associated with vascular anomalies, which are often indicative of diabetic pathology. The final layers, comprising average pooling and a fully connected output, consolidate learned features for classification tasks, making ResNet34 a suitable backbone for non-invasive diagnostic systems.

2) EfficientNet architectures

Fig. 3 depicts the architecture of an EfficientNet-based convolutional neural network, optimized for high performance and computational efficiency in image classification tasks. The model begins with a 3×3 convolutional layer followed by batch normalization and the Swish activation function, establishing initial feature extraction at a spatial resolution of 112×112×32. It then progresses through a series of MBConv blocks—Mobile Inverted Bottleneck Convolutions—with varying kernel sizes (3×3 and 5×5), expansion ratios, and output dimensions. These blocks incorporate depthwise separable convolutions and linear projections, enabling the network to balance representational power with reduced parameter count and memory footprint.

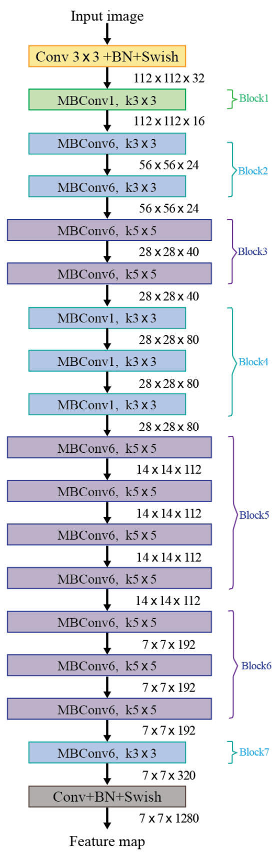


Fig. 3. The structural layout of EfficientNet-B0 is illustrated, featuring seven distinct stages or blocks differentiated by color for clarity [36].

The architecture's progressive downsampling—from 112×112 to 7×7 spatial dimensions—facilitates hierarchical feature learning across multiple scales. Each MBConv stage is carefully designed to preserve spatial information while increasing channel depth, culminating in a final convolutional layer that outputs a 7×7×1280 feature map. This structure is particularly well-suited for biomedical applications such as diabetic classification using PPG signals, where subtle temporal and morphological patterns must be captured efficiently. EfficientNet's compound scaling strategy and lightweight design make it an attractive backbone for deployment in resource-constrained environments.

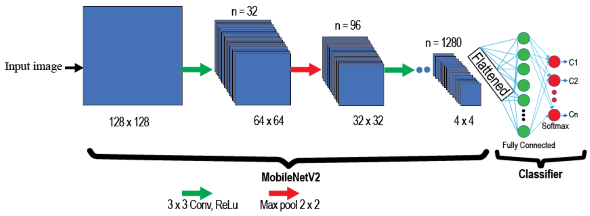


Fig. 4. The MobileNetV2 architecture in a classification task [37].

3) *MobileNet architectures*

MobileNet is designed for mobile and embedded vision applications, using depthwise separable convolutions to reduce computational costs [31]. Fig. 4 describes a Convolutional Neural Network (CNN) architecture based on MobileNetV2, tailored for image classification tasks. This architecture exemplifies a lightweight and efficient deep learning model optimized for mobile and embedded vision applications. The flow of data from input to output is systematically structured, beginning with a 128×128 input image and culminating in a softmax-based classification output. Each stage of the network is designed to progressively extract and condense spatial and semantic features, enabling accurate categorization with minimal computational overhead.

The initial layer applies a 3×3 convolution followed by a ReLU activation function, producing 32 feature maps of size 64×64. This operation captures local patterns such as edges and textures while reducing the spatial resolution. The subsequent max pooling layer, using a 2×2 filter, further downsamples the feature maps to 32×32 and increases the depth to 96 channels. This pooling operation not only reduces dimensionality but also enhances translational invariance, a critical property for robust image recognition.

As the data flows through deeper layers of the MobileNetV2 backbone, the number of feature maps increases significantly to 1280 while the spatial dimensions shrink to 4×4. This transformation reflects the model's ability to encode high-level abstractions in a compact representation. MobileNetV2 employs depthwise separable convolutions and inverted residual blocks, which are not explicitly shown in the diagram but are integral to its efficiency. These components reduce the number of parameters and computational cost while preserving representational power.

The final stages of the architecture transition from feature extraction to classification. The 4×4×1280 feature maps are flattened into a 1D vector of size 1280, which serves as input to a fully connected layer. This dense layer integrates the learned features and feeds into a softmax activation function, producing a probability distribution over the target classes (C_1, C_2, ..., C_n). The softmax layer ensures that the output values are interpretable as class probabilities, facilitating decision-making in classification tasks.

Overall, the figure encapsulates a streamlined CNN pipeline that balances accuracy and efficiency. By leveraging MobileNetV2's architectural innovations, the model is well-suited for deployment in resource-constrained environments such as mobile devices. The clear delineation of convolutional, pooling, and classification stages underscores the modularity and scalability of modern CNNs, making them adaptable to a wide range of computer vision applications.

E. Set up Parameters for Training and Testing Models

All generated PPG waveform images were resized to meet the input requirements of each CNN architecture before being used. The ResNet models, including ResNet18, ResNet34, and ResNet50, used input images with a dimension of 224×224×3 pixels. The MobileNet models, including MobileNet-V3-Large, MobileNet-V3-Small, and MobileNet-V2-Small, were also configured with input images of 224×224×3 pixels. For the EfficientNet models, the input dimensions were adjusted according to the corresponding architecture configuration, namely 224×224×3 pixels for EfficientNet-Lite, 300×300×3 pixels for EfficientNet-V2-S, and 480×480×3 pixels for EfficientNet-V2-M. All images were converted into three-channel RGB format to match the input format of the pretrained ImageNet models.

As we have five modified datasets from the MAV filter, the model development was conducted sequentially. We first selected a ResNet18 model to train and test each dataset, "PPG-BP Dataset-M1" up to "PPG-BP Dataset-M20", and collected confusion matrix parameters such as accuracy, precision, recall, and F1-score. Afterwards, the second model, ResNet34, is selected for training and testing all datasets and collecting all the confusion matrices. This procedure is done until the last model, MobileNet-Lite. Finally, all classification performance metrics corresponding to each model are analyzed to find out which model is reliable for diabetic classification.

The models utilized in diabetic classification are adjusted by adjusting several parameters and defining the utilized architecture as indicated in Table III. The first parameter is epoch, which is a single run of the whole training dataset that enables the model to learn from the data repeatedly. In this study, 100 epochs were used, which indicates that data training was repeated 100 times. Afterwards, select an appropriate batch size, which refers to the number of training instances utilized in a single model training cycle; more frequent model parameter updates will be the outcome of a smaller batch size. To improve model accuracy, the optimizer parameter must be set up in order to modify the model parameters (weights) in order to minimize the loss function. improves model correctness by modifying the model's weights and parameters to minimize the loss function.

The last parameter required in developing a classification model is the learning rate, which is a hyperparameter determining how quickly or slowly the model learns by controlling the step size throughout the optimization process. In our work, we use epoch, batch size, optimizer, and learning rate as indicated in Table III.

TABLE III. SET UP PARAMETERS FOR TRAINING THE MODELS

Parameter	Values
Epoch	100
Batch size	32
Optimizer	RMSprop
Learning Rate	0.0001

F. Model Performance Evaluation

Model performance is evaluated using commonly accepted classification metrics, including accuracy, precision, recall, and F1-score. Accuracy provides an overall measure of correct classification, while precision and recall capture the balance between false positives and false negatives, which is particularly important in medical screening contexts. The F1-score offers a harmonic mean of precision and recall, serving as a robust summary metric. Confusion matrices are also analyzed to provide detailed insight into classification behavior for each model and preprocessing configuration. These metrics collectively ensure that performance claims are comprehensive and clinically interpretable. The metrics are assessed as follows [38]:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+TN}} \tag{4}$$

$$\text{F1\_score} = \frac{2\times\text{Recall}\times\text{Precision}}{\text{Recall+Precision}} \tag{5}$$

This measure takes into account to evaluate CNN models.

III. RESULTS AND DISCUSSION

In this section, we address three complementary objectives to advance non-invasive diabetic classification using PPG signals. First, we preprocess the PPG-BP dataset using the MAV filter with varying M values to evaluate its impact on signal quality and feature retention. Second, we address the challenge of limited labeled data by developing classification models through transfer learning, leveraging ResNet, EfficientNet, and MobileNet architectures pretrained on the ImageNet dataset. Third, we provide recommendations on the suitability of these convolutional neural network (CNN) models for deployment on resource-constrained hardware, thereby ensuring practical applicability in real-world screening scenarios. Collectively, these objectives aim to optimize both the data preparation pipeline and the model design strategy, while also considering the feasibility of implementation in portable and low-power diagnostic systems.

A. Modification of PPG-BP Dataset

The original PPG-BP dataset presents several limitations, including susceptibility to noise and artifacts, reduced feature discriminability, challenges in model training, and concerns regarding clinical reliability. Consequently, preprocessing through filtering techniques, such as the application of a moving average (MAV) filter with varying window sizes, is essential to enhance signal quality and ensure the dataset's suitability for robust model development.

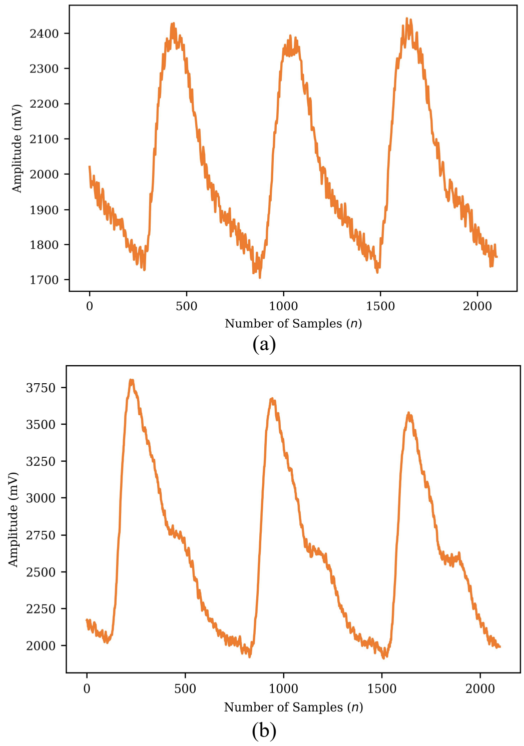
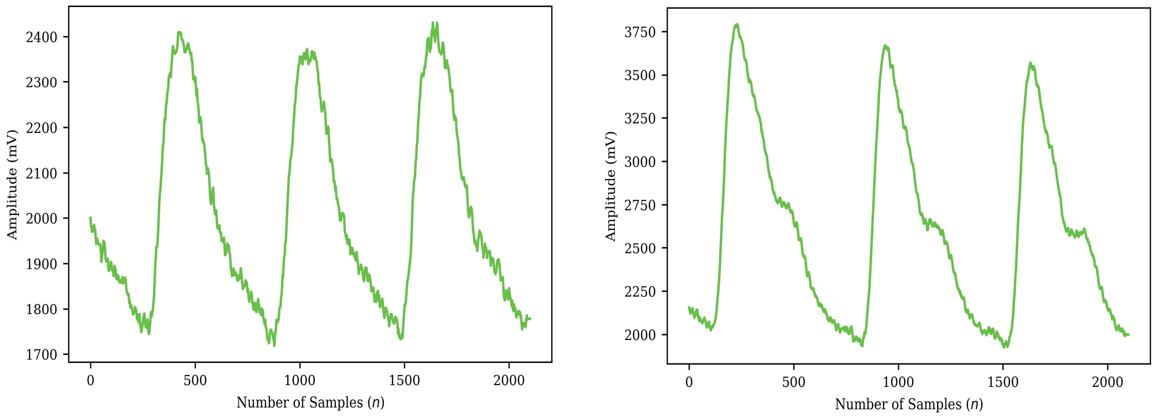


Fig. 5. PPG signals filtered with MAV (M = 5) for (a) diabetic and (b) non-diabetic subjects.

We extend the original PPG-BP dataset into four versions by applying the MAV filter with four different window sizes (M = 5, 10, 15, and 20). This modification enables comparative evaluation of how varying degrees of signal smoothing influence feature extraction and subsequent model performance. Fig. 5 presents filtered PPG signals obtained using a MAV window size M = 5, comparing Fig. 5 (a) a diabetic subject and Fig. 5 (b) a non-diabetic subject. Both signals exhibit periodic waveforms with distinguishable peaks and troughs, reflecting the pulsatile nature of blood volume changes. The diabetic signal Fig. 5 (a) shows reduced amplitude variability and a less pronounced waveform morphology, which may indicate vascular stiffness or impaired peripheral circulation commonly associated with diabetes. In contrast, the non-diabetic signal Fig. 5 (b) displays sharper peaks and greater amplitude dynamics, suggesting healthier vascular responsiveness. The MAV filtering enhances signal smoothness while preserving essential morphological features, facilitating more reliable feature extraction for downstream classification tasks.

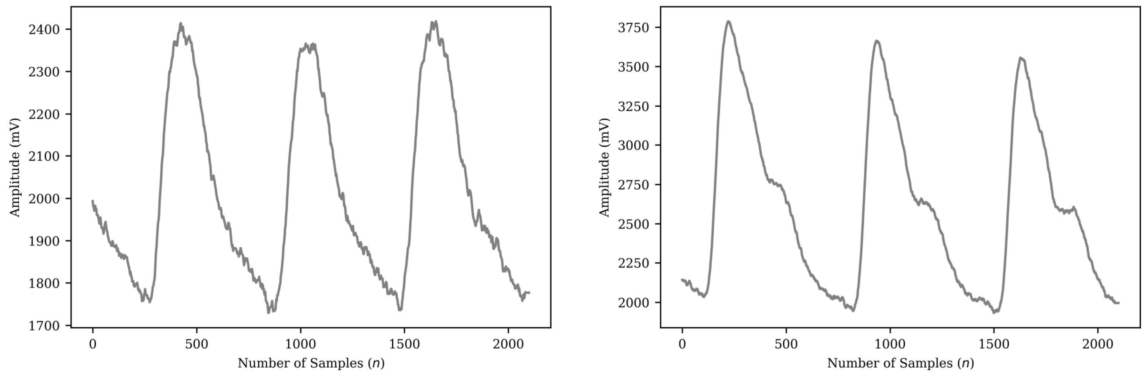
Fig. 6 displays filtered PPG signals with MAV window size M = 10, comparing Fig. 6 (a) diabetic and Fig. 6 (b) non-diabetic subjects. Relative to the previous figure filtered with $M = 5$, the signals in this graph exhibit increased smoothness and reduced high-frequency fluctuations. The diabetic signal Fig. 6 (a) shows a more attenuated waveform with less pronounced peaks, suggesting that a larger MAV window may suppress subtle morphological features critical for classification. Meanwhile, the non-diabetic signal Fig. 6 (b) retains its periodic structure but with slightly dampened amplitude dynamics compared to the $M = 5$ version.

This comparison highlights the trade-off introduced by increasing the MAV window size: while $M = 10$ improves noise suppression and baseline stability, it may also reduce the visibility of fine-grained vascular features, particularly in diabetic subjects where signal morphology is already compromised. These observations underscore the importance of selecting an optimal MAV parameter that balances denoising with feature preservation to support effective CNN-based diabetic classification.



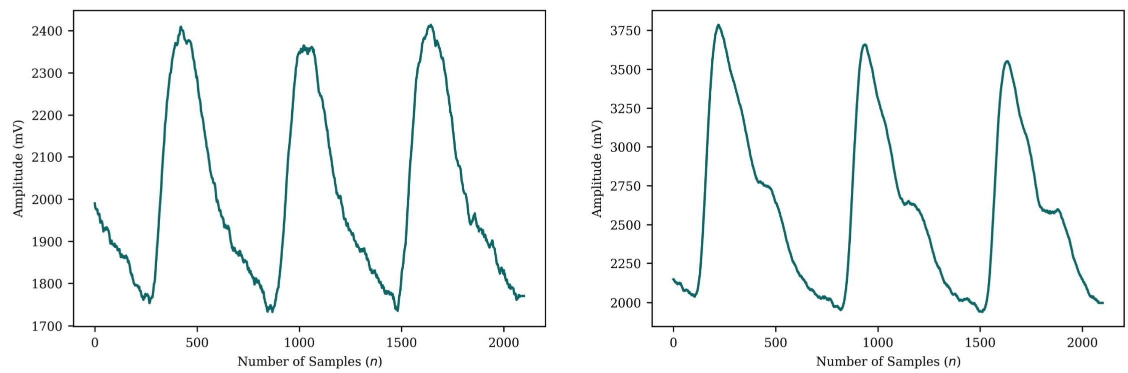
(a) (b)

Fig. 6. PPG signals filtered with MAV (M = 10) for (a) diabetic and (b) non-diabetic subjects.



(a) (b)

Fig. 7. PPG signals filtered with MAV (M = 15) for (a) diabetic and (b) non-diabetic subjects.



(a) (b)

Fig. 8. PPG signals filtered with MAV (M = 20) for (a) diabetic and (b) non-diabetic subjects.

Increasing M values to 15 and 20, we obtained PPG signals as illustrated in Figs. 7 and 8. Compared to the earlier figures with $M = 5$ and $M = 10$, these signals exhibit significantly smoother waveforms with reduced amplitude fluctuations and sharper noise suppression. In particular, the diabetic signals show a pronounced flattening effect, where peak structures become less distinct, and waveform morphology is increasingly attenuated. This may hinder the model's ability to extract fine-grained features relevant to vascular abnormalities.

In contrast, non-diabetic signals retain more of their periodic structure even at higher MAV values, though the amplitude dynamics are visibly dampened. The comparison across all four window sizes reveals a clear trade-off: smaller MAV windows (e.g., $M = 5$) preserve more physiological detail but allow residual noise, while larger windows (e.g., $M = 20$) enhance smoothness at the cost of morphological fidelity. These observations suggest that moderate filtering (e.g., $M = 10$) may offer a balanced compromise between noise reduction and feature preservation, particularly for CNN-based diabetic classification tasks where waveform integrity is critical.

B. Evaluation of ResNet Models

We begin by evaluating three ResNet architectures: ResNet18, ResNet34, and ResNet50. These models were selected due to their feasibility of deployment on low-performance machines, thereby making them suitable candidates for prototyping. Furthermore, ResNet architectures are widely recognized in the literature and have consistently demonstrated strong performance across a variety of classification tasks.

In deep learning–based classification, model performance is typically assessed using a confusion matrix, which provides a detailed summary of predicted versus actual class distributions. Figs. 9–11 show the confusion matrix results of ResNet18, ResNet34, and ResNet50, respectively. These three bar charts present a comparative analysis of the performance metrics of accuracy, precision, recall, and F1-score, for ResNet18, ResNet34, and ResNet50 models, respectively, across datasets filtered using a moving average filter (MAV) with varying parameter values (denoted as M=5, $M = 10$, $M = 15$, and $M = 20$), where $M = 1$ is corresponds to the unfiltered dataset. Each chart illustrates how the performance of the respective ResNet architecture varies as the MAV window size M increases, reflecting the impact of dataset smoothing on model evaluation metrics. For all three models, a general trend is observable: as the MAV window size M increases from 1 to 20, the performance metrics initially exhibit fluctuations, but tend to stabilize or slightly improve for moderate values of M. Specifically, ResNet18 demonstrates relatively high and stable performance across all metrics, with minor variations as M increases. The ResNet34 model shows a similar pattern, maintaining robust performance in accuracy, precision, recall, and F1-score, though a slight decline in precision and recall is noticeable at higher M values. ResNet50, being the deepest architecture among the three, exhibits the most consistent performance across all metrics, with only marginal degradation in precision and recall at the highest M value (M=20). The results suggest that dataset filtering using MAV can influence model performance, with moderate filtering (e.g., $M = 5$ to $M = 10$) potentially enhancing or maintaining evaluation metrics by reducing noise in the data. However, excessive filtering (e.g., $M = 20$) may lead to a slight reduction in precision and recall, likely due to the loss of fine-grained information in the dataset. Overall, these findings underscore the importance of selecting an appropriate MAV.

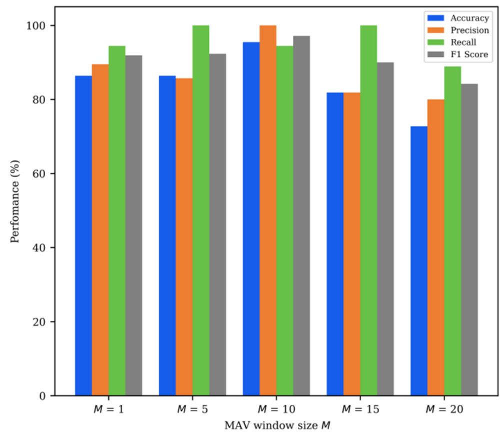


Fig. 9. Classification performance metrics of ResNet18.

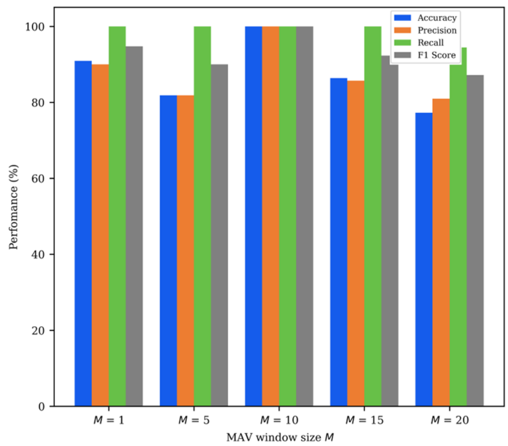


Fig. 10. Classification performance metrics of ResNet34.

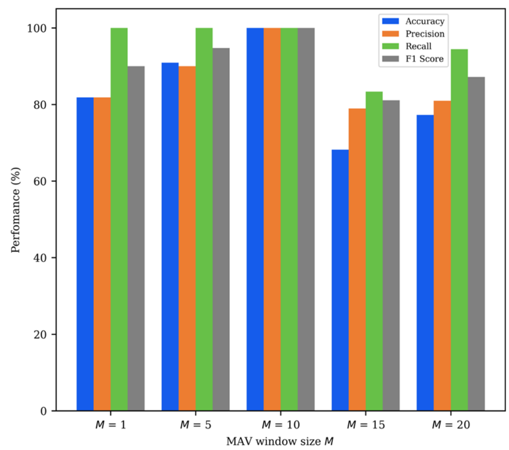


Fig. 11. Classification performance metrics of ResNet50.

In addition to the confusion matrix, learning curves are employed as graphical tools to illustrate a model's performance (e.g., accuracy, loss, or error rate) relative to the number of training iterations, epochs, or sample size. These curves play a critical role in evaluating and diagnosing the effectiveness of deep learning models. In this simulation, we present the learning curve results of models trained, validated, and tested using the PPG-BP Dataset-M10, which is the filtered dataset obtained with $M = 10$ in the MAV filter. This dataset was selected because it demonstrated perfect performance in the confusion matrix.

Figs. 12–Fig. 14 present the learning curves of three variants of the ResNet architecture—ResNet18, ResNet34, and ResNet50—indicating the training and validation loss over 100 epochs. Each graph illustrates the progression of loss values for both the training (blue line) and validation (orange line) datasets, offering insight into the models' learning dynamics and generalization capabilities.

In terms of loss plots, a similar trend is observable: both the training and validation loss values exhibit a rapid decline in the initial epochs, followed by a gradual stabilization as training progresses.

This pattern is characteristic of effective model training, where the loss decreases sharply at the beginning due to the model learning general features, and then plateaus as it refines its parameters. The close alignment between the training and validation loss curves across all models suggests that there is no significant overfitting, as the loss on the validation set remains comparable to that of the training set throughout the epochs.

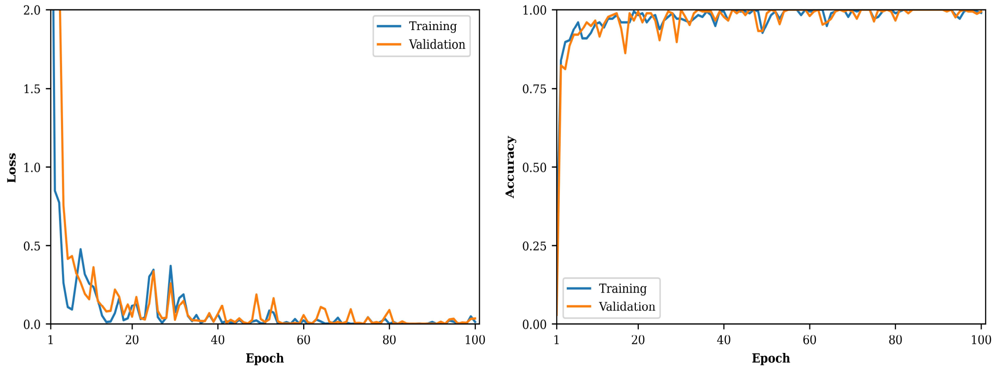


Fig. 12. Learning curve of ResNet18 model using PPG-BP Dataset-M10.

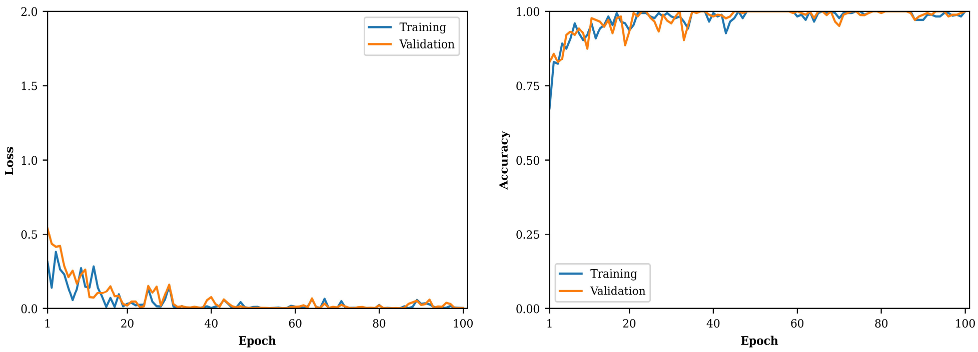


Fig. 13. Learning curve of the ResNet34 model using PPG-BP Dataset-M10.

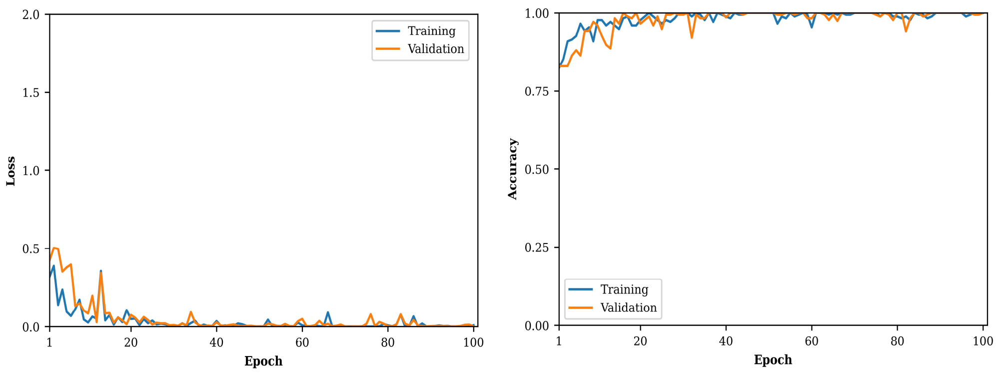


Fig. 14. Learning curve of the ResNet50 model using PPG-BP Dataset-M10

The training and validation loss curves employing PPG-BP Dataset-M10 for three variants of the ResNet architecture—ResNet18, ResNet34, and ResNet50—over 100 epochs. Each graph illustrates the progression of loss values for both the training (blue line) and validation (orange line) datasets, offering insight into the models' learning dynamics and generalization capabilities. Remarkably, the ResNet18 model appears to achieve slightly lower loss values compared to ResNet34 and ResNet50, which may indicate its efficiency in converging faster with fewer parameters.

The ResNet50 model, while deeper and more complex, does not show a substantial advantage in terms of loss reduction, implying that the increased depth does not necessarily translate to better performance on this particular dataset. Overall, these curves reflect the models' ability to generalize well to unseen data, with minimal divergence between training and validation losses, thus demonstrating robust learning and effective regularization.

In the case of training and validation accuracy across all three models, a rapid increase in accuracy is observed within the initial epochs, followed by a plateau as training continues. This trend indicates that the models quickly learn discriminative features from the data, achieving high accuracy early in the training process. The close overlap between the training and validation accuracy curves for each model suggests effective generalization, with minimal signs of overfitting. Notably, all three architectures ultimately converge to similar accuracy levels, hovering around 0.95 to 1.00, which underscores their capability to achieve high performance on the given task. The ResNet18 model, despite its relatively simpler architecture, demonstrates competitive accuracy compared to the deeper ResNet34 and ResNet50 models. The consistency between training and validation accuracy further reinforces the robustness of these models in maintaining performance across unseen data, aligning with expectations for well-regularized deep learning architectures.

C. Evaluation of EfficientNet Models

Confusion matrix results of the EfficientNet models' evaluation are shown in Figs. 15–17, summarizing the performance of three EfficientNet variants, EfficientNet V2-M, EfficientNet V2-S, and EfficientNet-Lite, across datasets filtered using MAV with varying parameter values. The evaluation metrics include accuracy, precision, recall, and F1-score, offering insights into how each model responds to different levels of dataset smoothing. Overall, the results indicate that moderate dataset filtering ($M = 10$) enhances model performance across all EfficientNet variants, particularly in terms of accuracy and recall. However, excessive filtering ($M = 20$) tends to reduce accuracy and precision, suggesting that an optimal MAV parameter is crucial for balancing data smoothing and the preservation of informative features. EfficientNet Lite consistently demonstrates resilience to varying MAV parameters, particularly in precision and recall, making it a robust choice for tasks requiring high generalization and reliability.

These findings highlight the importance of selecting an appropriate MAV parameter to maximize model performance while mitigating the loss of critical data features.

Similar learning curve patterns are observed in the evaluation of EfficientNet models, as illustrated in Fig. 18, where EfficientNet-Lite is used as an example. In contrast to the ResNet models, which achieved stable and high accuracy with minimal divergence between training and validation metrics, the current model displays more erratic behavior in both loss and accuracy.

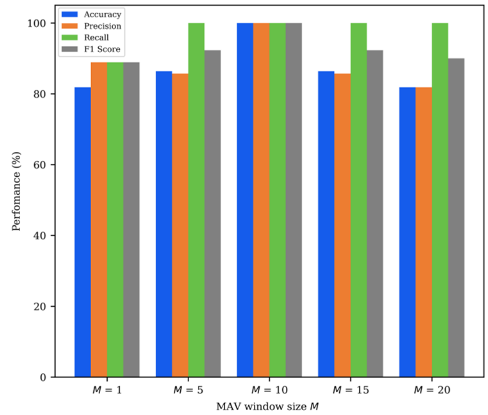


Fig. 15. Classification performance metrics of EfficientNet V2-M.

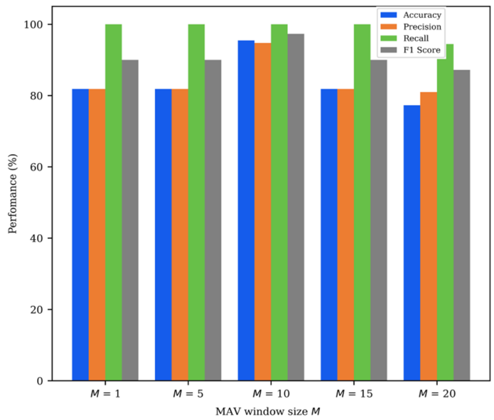


Fig. 16. Classification performance metrics of EfficientNet V2-S.

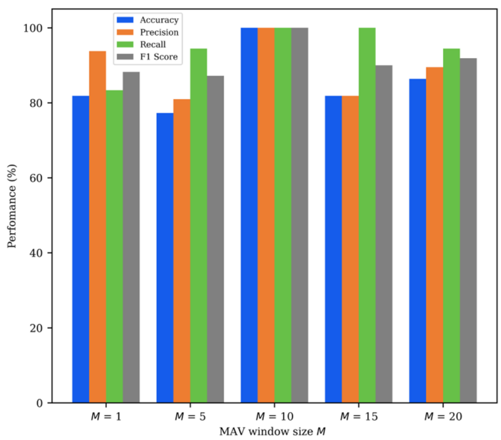
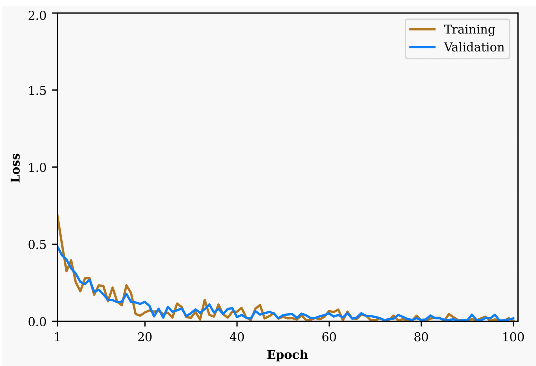
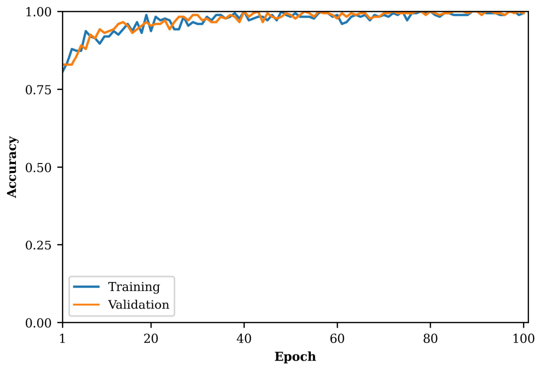


Fig. 17. Classification performance metrics of EfficientNet-Lite.



(a)



(b)

Fig. 18. Learning curve of the EfficientNet-Lite model using PPG-BP Dataset-M10 for (a) loss and (b) Accuracy trends.

D. Evaluation of MobileNet Models

The metric evaluation of the MobileNet series is shown in Figs. 19–21. The different performance of the MobileNet variants, MobileNet-V3-Large, MobileNet-V3-Small, and MobileNet-V2-Small, is illustrated in the confusion matrix.

Each figure presents the performance metrics of accuracy, precision, recall, and F1-score across various values of the MAV window size M, which represents the average number of the moving averaging filter. This parameter influences the smoothing effect on the model's predictions and, consequently, its performance metrics.

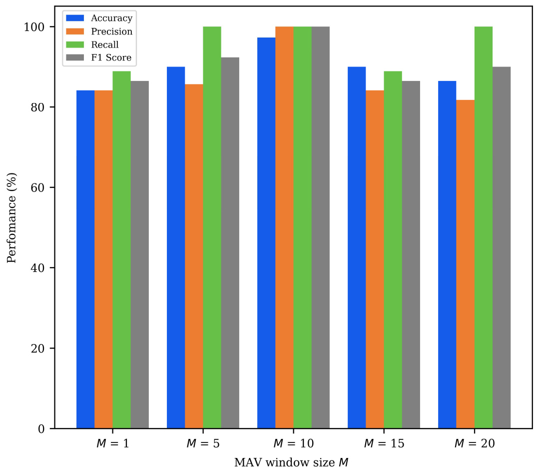


Fig. 19. Classification performance metrics of MobileNet-V3-Large.

Fig. 19 shows the performance of MobileNet-V3-Large Evaluation. The performance metrics exhibit distinct trends as the MAV window size M varies. At $M = 1$, the model starts with relatively lower performance across all metrics, likely due to minimal smoothing. As M increases to $M = 5$ and $M = 10$, there is a significant improvement in accuracy, precision, recall, and F1-score, with all metrics peaking at $M = 10$. This suggests that an intermediate level of smoothing enhances the model's ability to generalize and perform optimally. However, beyond $M = 10$, particularly at $M = 15$ and $M = 20$, the performance metrics decline. This indicates that excessive smoothing may introduce lag or over-smoothing, thereby degrading the model's performance.

Fig. 20 indicates the MobileNet-V3-Small model evaluation, demonstrating a similar pattern in its performance metrics as the MAV window size M changes. Initially, at $M = 1$, the metrics are lower, reflecting insufficient smoothing. However, as M increases to $M = 5$ and $M = 10$, the performance metrics improve significantly, reaching their highest values at $M = 10$. This indicates that the model benefits from an optimal level of smoothing at this point. Beyond $M = 10$, the performance metrics decline at $M = 15$ and $M = 20$, suggesting that further increasing the moving average parameter leads to diminished returns and potential over-smoothing.

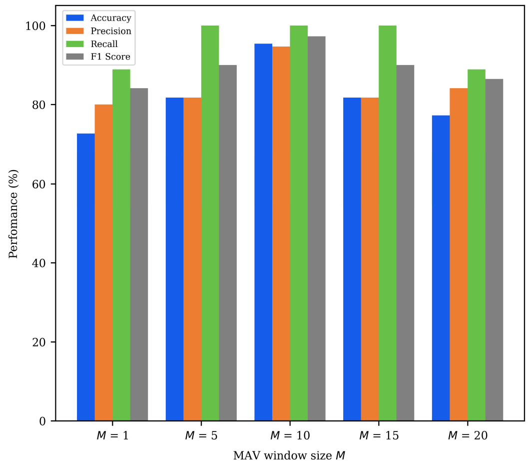


Fig. 20. Classification performance metrics of MobileNet-V3-Small.

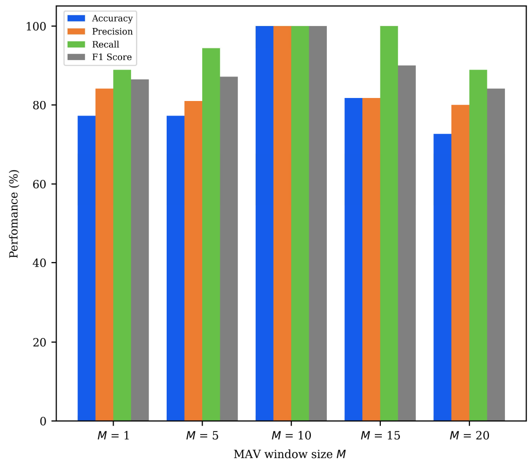


Fig. 21. Classification performance metrics of MobileNet-V2-Small.

The selection $M = 10$ for the MAV filter is considered optimal because it balances noise suppression with the preservation of physiologically relevant signal features. Smaller window sizes, such as $M = 1$ or $M = 5$, fail to adequately smooth high-frequency noise, leaving residual fluctuations that obscure waveform morphology. Larger windows, on the other hand, tend to over-smooth the signal, distorting critical features such as systolic peaks and dicrotic notches. With $M = 10$, the filter achieves sufficient smoothing to reduce random noise while maintaining temporal resolution, ensuring accurate detection of PPG characteristics essential for biomedical analysis.

Similarly, the evaluation of the MobileNet-V2-Small model (Fig. 21) also reveals a comparable trend in performance metrics as the MAV window size M varies. The model starts with lower values for accuracy, precision, recall, and F1-score at $M = 1$. These metrics improve as M increases to $M = 5$ and $M = 10$, with the highest performance observed at M=10. However, further increasing M to $M = 15$ and $M = 20$ results in a decline in performance metrics. This pattern is consistent with the evaluations of the other MobileNet variants, reinforcing the idea that $M = 10$ is the optimal value for achieving the best balance between smoothing and model performance.

Among the MobileNet series, MobileNet-V2-Small was selected based on its favorable learning curve characteristics. The training plots demonstrate effective convergence and strong generalization performance over 100 epochs. In the loss curve, both training and validation losses show a steep decline during the initial epochs, indicating rapid learning and successful optimization of model parameters. The losses stabilize near zero in later epochs, suggesting that the model has minimized error without overfitting, as evidenced by the close alignment between training and validation trajectories.

The accuracy curve further supports this observation. Both training and validation accuracies increase sharply in the early stages and plateau near 1.0, reflecting high classification performance across both datasets. The minimal gap between training and validation accuracy implies that the model generalizes well to unseen data, a critical requirement for reliable diabetic classification. This behavior is particularly notable given the limited size of the dataset, highlighting the effectiveness of transfer learning and the architectural efficiency of MobileNet-V2-Small.

Overall, the learning curves suggest that MobileNetV2S achieves a favorable balance between model complexity and performance. Its lightweight design enables fast convergence and high accuracy while maintaining robustness against overfitting—making it a strong candidate for deployment in resource-constrained environments such as portable diagnostic devices.

TABLE IV. THE SELECTED MODEL FOR M = 10 (OPTIMAL VALUE)

Model	Size (MB)	Accuracy	Precision	Recall	F1-score
ResNet34	81.3	100	100	100	100
EfficientNet-Lite	13.1	100	100	100	100
MobileNet-V2-Small	**8.73**	100	100	100	100

Considering computational complexity and model size, it is also important to evaluate the feasibility of implementing the model as a prototype within practical resource constraints. We propose that the top-performing models from each family—specifically ResNet34, EfficientNet-Lite, and MobileNet-V2-Small—are suitable for implementation as prototypes on microcontrollers, Raspberry Pi, and other single-board computers (SBCs). The memory footprints of these models are summarized in Table IV. As indicated by the data, MobileNet-V2-Small and EfficientNet-Lite are particularly well-suited for deployment on resource-constrained microcontroller devices.

The comparative results presented in the table highlight a critical dimension in modern machine learning research: the balance between model performance and computational efficiency. While ResNet34, EfficientNet-Lite, and MobileNet-V2-Small all achieve perfect scores across Accuracy, Precision, Recall, and F1-score, their model sizes differ substantially. This discrepancy underscores the importance of considering not only predictive capability but also the computational footprint of a model when evaluating its suitability for real-world applications. From a deployment perspective, smaller models such as EfficientNet-Lite (13.1 MB) and MobileNet-V2-Small (8.73 MB) offer significant advantages.

The perfect classification performance across ResNet34, EfficientNet-Lite, and MobileNet-V2-Small can be attributed to the effectiveness of transfer learning. By initializing the models with pretrained weights, the networks inherited robust feature representations that generalized seamlessly to our dataset. This enabled the classifier to achieve complete separability of the classes, as reflected in the flawless confusion matrix values. The consistency across architectures of varying sizes further demonstrates that the discriminative features required for this task were adequately captured by the pretrained models, highlighting the efficiency of transfer learning in biomedical signal classification. Their reduced memory requirements and faster inference times make them particularly well-suited for environments with limited resources, such as mobile devices, embedded systems, and edge computing platforms. In contrast, ResNet34, with a size of 81.3 MB, may pose challenges in such contexts despite its flawless performance. This observation aligns with broader trends in the field, where lightweight architectures are increasingly favored for practical implementation without compromising accuracy.

The fact that EfficientNet-Lite and MobileNet-V2-Small match ResNet34's performance demonstrates the success of these innovations in bridging the gap between efficiency and effectiveness. This raises important questions about scalability: whether similar efficiency gains can be maintained across more complex datasets, larger-scale tasks, or domains requiring higher variability in input data. The ability to achieve high accuracy with minimal resource demands is increasingly valued. Thus, the table provides compelling evidence that lightweight models are not merely alternatives but viable frontrunners in the pursuit of efficient, scalable, and accessible artificial intelligence systems.

The dataset used in this study is the same as that employed in previous work [28]. No new dataset was added, as expanding the dataset was constrained by the need for ethical clearance, the time-intensive process of subject recruitment, and the logistical challenges associated with obtaining diabetic measurements. To reduce the risk of overfitting, this study employed transfer learning using pretrained CNN models rather than training all models from scratch. However, we acknowledge that transfer learning alone is not sufficient to guarantee model generalizability.

Table V presents a comparative analysis of various methodologies for diabetes detection using PPG signals, contrasting established literature with the authors' proposed approach. While previous studies have utilized a range of techniques, including time-frequency analysis (Susana et al.), lightweight 1D Convolutional Neural Networks (CNNs) [23, 25], and feature-based machine learning [28], they often rely on raw temporal data or manual feature extraction. In contrast, the proposed study introduces a novel pipeline that converts PPG-BP signals into waveform images processed through a MAV filter and pretrained 2D CNNs. This methodology achieves a perfect 100% in accuracy, precision, recall, and F1-score at a window size of M=10, effectively addressing the performance gaps identified in prior research.

TABLE V. COMPARISON OF THE PROPOSED METHOD WITH PREVIOUSLY PUBLISHED PPG-BASED DIABETES CLASSIFICATION STUDIES

Ref.	Dataset	Method	Main Result	Remarks
[10]	PPG signal	Time–frequency analysis	Promising classification performance	Relies on feature/time–frequency representation
[23]	Raw PPG wave	Light CNN	Effective type-2 diabetes detection	Uses a lightweight CNN on raw/single PPG wave
[25]	PPG features	Machine learning-based diabetes detection	Demonstrated feasibility of PPG-based diabetes detection	Uses a feature-based ML approach
[28]	Short-time PPG signal	1D CNN	Diabetes detection using temporal PPG representation	Directly processes a 1D time-series signal
Proposed	PPG-BP converted into waveform images	MAV filter + pretrained 2D CNNs	Demonstrated superior classification performance metrics at $M = 10$	Systematically evaluates MAV window size and lightweight CNN models

E. Limitations and Future Work

This study has several limitations, as it relies on a single public dataset (PPG-BP) comprising 219 subjects, of whom 38 were diagnosed with diabetes. The relatively small sample size, along with the use of a binary classification framework (diabetic vs. non-diabetic), may limit the generalizability of the findings. Expanding the dataset was constrained by ethical clearance requirements, the time-intensive nature of subject recruitment, and the logistical challenges of obtaining diabetic measurements. Although transfer learning with pretrained CNN models was employed to reduce overfitting, this approach alone cannot fully guarantee model generalizability.

Future work should validate the proposed framework using larger and more diverse datasets, independent external datasets, and subject-wise evaluation strategies. More rigorous validation methods, such as cross-validation, longitudinal testing, and real-world pilot studies under varied sensor conditions, are needed to confirm robustness. In addition, comparative studies between image-based 2D CNNs, raw-signal-based 1D CNNs, and hybrid architectures would help evaluate trade-offs in classification performance, computational efficiency, and deployment feasibility. Broader validation across different populations and acquisition conditions will be essential to establish the clinical and consumer applicability of the proposed method.

IV. CONCLUSIONS

This paper investigated a non-invasive diabetes classification framework based on PPG signals and convolutional neural networks (CNNs), with particular emphasis on the role of preprocessing optimization through moving average filtering. The results demonstrate that preprocessing is not merely a supportive step but a decisive component that substantially influences classification performance. Across all evaluated CNN families—ResNet, EfficientNet, and MobileNet—moderate signal smoothing consistently improved predictive accuracy, while insufficient or excessive smoothing led to degraded performance.

The central finding of this work is that a MAV window size $M = 10$ provides an optimal balance between noise suppression and preservation of diagnostically relevant signal morphology. At this configuration, multiple CNN architectures achieved near-perfect or perfect classification performance, as reflected by accuracy, precision, recall, and F1-score metrics. These results empirically support the theoretical expectation that moderate smoothing enhances feature learning by reducing high-frequency noise and motion-related artifacts without obscuring subtle waveform characteristics linked to diabetes-related vascular changes.

The analysis shows that both deep and lightweight CNN models can achieve strong classification performance with proper preprocessing, with compact architectures like EfficientNet-Lite and MobileNet offering an effective balance of accuracy and efficiency for resource-constrained devices. However, the study's dependence on a single dataset and binary classification limits generalizability, underscoring the need for broader validation across diverse populations, sensor setups, and real-world conditions, as well as more rigorous evaluation strategies to strengthen confidence in practical deployment.

Future research should therefore explore multi-class classification scenarios, such as differentiating stages of diabetes or related comorbidities, and investigate robustness under uncontrolled measurement environments typical of wearable use. Integrating explainability methods to enhance transparency and clinician trust, as well as addressing ethical and regulatory considerations, will also be critical for translation into healthcare practice.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Zulhelmi Zulhelmi conducted original draft preparation, development of the model, software coding, and verification; Roslidar Roslidar formulated tasks, analysis, and verification; and Sawal Hamid bin Md Ali designed conceptualization and methodology. Nasaruddin Nasaruddin reviewed and edited the manuscript.

FUNDING

This research was supported by Universitas Syiah Kuala with the research contract number: 8850/UN11.F4/PG.01.03/15408-PTNBH/2025.

This research was also partly funded by the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science, and Technology and managed under the Equity Program (Contract No. 4318/B3/DT.03.08/2025 and No. 491/UN11/HK.02.06/2025.

REFERENCES

[1] WHO. Diabetes. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes

[2] International Diabetes Federation. IDF Diabetes Atlas. 2024. [Online]. Available: https://diabetesatlas.org/

[3] R. M. Rosenfeld, M. L. Grega, M. C. Karlsen *et al.*, "Lifestyle interventions for treatment and remission of type 2 diabetes and prediabetes in adults: A clinical practice guideline from the American College of Lifestyle Medicine," *American Journal of Lifestyle Medicine,* vol. 19, no. 2_suppl, pp. 10S–131S, 2025.

[4] L. Riley. Mean fasting blood glucose. [Online]. Available: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/2380

[5] M. Ghosh and V. Bora, "Evolution in blood glucose monitoring: A comprehensive review of invasive to non-invasive devices and sensors," *Discover Medicine,* vol. 2, no. 1. 2025. doi: 10.1007/s44337-025-00273-1

[6] F. Attivissimo, L. De Palma, A. Di Nisio *et al.*, "Photoplethysmography signal wavelet enhancement and novel features selection for non-invasive cuff-less blood pressure monitoring," *Sensors,* vol. 23, no. 4, 8342, 2023.

[7] L. Tang, S. J. Chang, C.-J. Chen *et al.*, "Non-invasive blood glucose monitoring technology: A review," *Sensors,* vol. 20, no. 23, 6925, 2020. doi: 10.3390/s20236925

[8] Y. Lin, M. Bariya, H. Y. Y. Nyein *et al.*, "Porous enzymatic membrane for nanotextured glucose sweat sensors with high stability toward reliable noninvasive health monitoring," *Advanced Functional Materials,* vol. 29, no. 33, 1902521, 2019.

[9] P. Makaram, D. Owens, and J. Aceros, "Trends in nanomaterial-based non-invasive diabetes sensing technologies," *Diagnostics,* vol. 4, no. 2, pp. 27–46, 2014. doi: 10.3390/diagnostics4020027.

[10] E. Susana, K. Ramli, P. D. Purnamasari *et al.*, "Non-invasive classification of blood glucose level based on photoplethysmography using time–frequency analysis," *Information,* vol. 14, no. 3, 145, 2023.

[11] M. C. Cebedio, L. A. Rabioglio, I. E. Gelosi *et al.*, "Analysis and design of a microwave coplanar sensor for non-invasive blood glucose measurements," *IEEE Sensors Journal,* vol. 20, no. 18, pp. 10572–10581, 2020. doi: 10.1109/JSEN.2020.2993182

[12] S. Y. Huang, Omkar, Y. Yoshida *et al.*, "Microstrip line-based glucose sensor for noninvasive continuous monitoring using the main field for sensing and multivariable crosschecking," *IEEE Sensors Journal,* vol. 19, no. 2, pp. 535–547, 2019. doi: 10.1109/JSEN.2018.2877691

[13] A. E. Omer, G. Shaker, S. Safavi-Naeini *et al.*, "Non-invasive real-time monitoring of glucose level using novel microwave biosensor based on triple-pole CSRR," *IEEE Trans. on Biomedical Circuits and Systems,* vol. 14, no. 6, pp. 1407–1420, 2020. doi: 10.1109/TBCAS.2020.3038589

[14] J. Li, I. Tobore, Y. Liu, *et al.*, "Non-invasive monitoring of three glucose ranges based On ECG by using DBSCAN-CNN," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3340–3350, Sept. 2021.

[15] A. Shanmugapriya, N. B. Selvan, N. H. Veyannie *et al.*, "Non-invasive optical glucose monitoring using laser spectroscopy," in *Proc. 2023 7th Int. Conf. on Electronics, Communication and Aerospace Technology*, Nov. 2023, pp. 1030–1035.

[16] M. Shokrekhodaei, D. P. Cistola, R. C. Roberts *et al.*, "Non-invasive glucose monitoring using optical sensor and machine learning techniques for diabetes applications," *IEEE Access,* vol. 9, pp. 73029–73045, 2021. doi: 10.1109/ACCESS.2021.3079182

[17] K. V. Pozhar, E. L. Litinskaia, M. O. Mikhailov *et al.*, "Reflection near-infrared spectroscopic method for non-invasive blood glucose measurement," in *Proc. 2020 IEEE Conf. of Russian Young Researchers in Electrical and Electronic Engineering*, Jan. 2020, pp. 2524–2527. doi: 10.1109/EIConRus49466.2020.9039005.

[18] M. Althobaiti and I. Al-Naib, "Optimization of dual-channel near-infrared non-invasive glucose level measurement sensors based on monte-carlo simulations," *IEEE Photonics Journal,* vol. 13, no. 3, pp. 1–9, 2021. doi: 10.1109/JPHOT.2021.3079408

[19] S. Ghosal, A. Kumar, V. Udutalapally *et al.*, "gluCam: Smartphone based blood glucose monitoring and diabetic sensing," *IEEE Sensors Journal,* vol. 21, no. 21, pp. 24869–24878, 2021. doi: 10.1109/JSEN.2021.3116191

[20] S. Gupta, A. Singh, A. Sharma *et al.*, "dSVRI: A PPG-based novel feature for early diagnosis of type-II diabetes mellitus," *IEEE Sensors Letters,* vol. 6, no. 9, pp. 1–4, 2022. doi: 10.1109/LSENS.2022.3203609

[21] S. Chen, F. Qin, X. Ma *et al.*, "Multi-view cross-fusion transformer based on kinetic features for non-invasive blood glucose measurement using PPG signal," *IEEE Journal of Biomedical and Health Informatics,* vol. 28, no. 4, pp. 1982–1992, 2024.

[22] J. Li, J. Ma, O. M. Omisore *et al.*, "Noninvasive blood glucose monitoring using spatiotemporal ECG and PPG feature fusion and weight-based choquet integral multimodel approach," *IEEE Trans. on Neural Networks and Learning Systems,* vol. 35, no. 10, pp. 14491–14505, 2024. doi: 10.1109/TNNLS.2023.3279383

[23] S. Zanelli, M. A. E. Yacoubi, M. Hallab *et al.*, "Type 2 diabetes detection with light CNN from single raw PPG wave," *IEEE Access*, vol. 11, pp. 57652–57665, 2023.

[24] A. Hina and W. Saadeh, "A noninvasive glucose monitoring SoC based on single wavelength photoplethysmography," *IEEE Trans. on Biomedical Circuits and Systems,* vol. 14, no. 3, pp. 504–515, 2020. doi: 10.1109/TBCAS.2020.2979514

[25] F. A. C. Oliveira, F. M. Dias, M. A. F. Toledo *et al.*, "Machine learning-based diabetes detection using photoplethysmography signal features," in *Proc. the Brazilian Symposium on Computing Applied to Health (SBCAS)*. doi: 10.5753/sbcas.2024.1889

[26] Y. Liang, G. Liu, Z. Chen *et al.*, *PPG-BP Database*, Figshare. doi: 10.6084/m9.figshare.5459299.v5

[27] M. Elgendi, V. Galli, C. Ahmadizadeh *et al.*, "Dataset of psychological scales and physiological signals collected for anxiety assessment using a portable device," *Data,* vol. 7, no. 9, 132, 2022. doi: 10.3390/data7090132

[28] M. S. I. Sumon, T. Ahmed, M. M. Khan *et al.*, "Detection of diabetes using 1D convolutional neural network from short time PPG signal," in *Proc. 2023 Int. Conf. on Information and Communication Technology for Sustainable Development*, Sept. 2023, pp. 372–376. doi: 10.1109/ICICT4SD59951.2023.10303575

[29] N. J. Papri, A. Ahmed, and A. Chowdhury, "IoT and cloud-based non-invasive diabetes detection system from photoplethysmogram," *Discover Internet of Things,* vol. 5, no. 1, 57, 2025. doi: 10.1007/s43926-025-00158-w

[30] Y. Chen, D. Li, Y. Li *et al.*, "Use moving average filter to reduce noises in wearable PPG during continuous monitoring," in *eHealth 360°*, Cham, K. Giokas, L. Bokor, and F. Hopfgartner, Eds., 2017, Springer International Publishing, pp. 193–203.

[31] A. Howard, M. Sandler, B. Chen *et al.*, "Searching for MobileNetV3," in *Proc. 2019 IEEE/CVF Int. Conf. on Computer Vision* , Nov. 2019, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140

[32] M. Sandler, A. Howard, M. Zhu *et al.*, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. 2018 IEEE/CVF Conf. on*

Computer Vision and Pattern Recognition, June 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474

[33] M. Kazeneza, A. S. Bosman, D. K. Amenyedzi *et al.*, "Balancing complexity and performance of machine learning models for avian pests sound detection in agricultural environments," *IEEE Access,* vol. 13, pp. 105813–105827, 2025.

[34] K. He, X. Zhang, S. Ren *et al.*, "Deep residual learning for image recognition," in *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition*, June 2016, pp. 770–778.

[35] K. Han, Y. Wang, J. Guo *et al.*, "ParameterNet: Parameters are all you need for large-scale visual pretraining of mobile networks," in *Proc. of 2024 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, June 2024, pp. 15751–15761.

[36] A. Zhou, Y. Ma, W. Ji *et al.*, "Multi-head attention-based two-stream EfficientNet for action recognition," *Multimedia Systems,* vol. 29, no. 2, pp. 1–12, 06/24 2022. doi: 10.1007/s00530-022-00961-3

[37] M. Akay, Y. Du, C. L. Sershen *et al.*, "Deep learning classification of systemic sclerosis skin using the MobileNetV2 model," *IEEE Open Journal of Engineering in Medicine and Biology,* vol. 2, no. 2021, pp. 104–110, 2021. doi: 10.1109/OJEMB.2021.3066097

[38] K. Riehl, M. Neunteufel, and M. Hemberg, "Hierarchical confusion matrix for classification performance evaluation," *Journal of the Royal Statistical Society: Series C (Applied Statistics),* vol. 72, no. 5, pp. 1394–1412, 2023. doi: 10.1093/jrsssc/qlad057

Zulhelmi Zulhelmi was born in Sigli, Aceh, on July 2, 1979. He received the B.Eng. degree from Universitas Syiah Kuala, Banda Aceh, Indonesia in 2003, and the M.Sc. degrees in electrical engineering from King Saud University (KSU), Riyadh Saudi Arabia in 2013. He has been a Lecturer at the Electrical and Computer Engineering Department, Universitas Syiah Kuala since 2003. Now, he is conducting the Ph.D. Program at Doctoral Programs of Engineering, Universitas Syiah Kuala. His research interests include electronic design, FPGAs architecture design and implementation, Microcontroller based on internet of things, biomedical electronics, machine learning and deep learning.

Roslidar Roslidar received the bachelor's degree in electrical engineering from Universitas Syiah Kuala, in 2002, the master's degree in telecommunication engineering from the University of Arkansas, USA, in 2009, and the Ph.D. degree in January 2022. Since 2003, she has been a Lecturer with Universitas Syiah Kuala. She has been actively doing research in science and technology. She was a Visiting Scholar with the Faculty of System Design, Tokyo Metropolitan University, Japan, in May 2014. She dedicated her time to develop the e-health monitoring system based on a non-invasive technique. She was granted a Fulbright Scholarship for the master's degree from the University of Arkansas. . She is currently a Full Professor with the Electrical Engineering Department, Universitas Syiah Kuala.

Sawal Hamid Md Ali received the bachelor's degree in electronic and computer engineering from Universiti Putra Malaysia, Serdang, Selangor, Malaysia, in 1998, and the master's degree in microelectronic system design and the Ph.D. degree in electrical and electronics from the University of Southampton, Southampton, U.K., in 2004 and 2010, respectively. He is currently the Deputy Director of the Centre for Innovation and Technology Transfer. He is also a Professor with the Department of Electrical, Electronics and System Engineering, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia.

Nasaruddin Nasaruddin received the B.Eng. degree in electrical engineering from the Sepuluh Nopember Institute of Technology, Surabaya, Indonesia, in 1997, and the M.Eng. and D.Eng. degrees in physical electronics and informatics from the Graduate School of Engineering, Osaka City University, Japan, in 2006 and 2009, respectively. From 2016 to 2020, he was the Head of the Electrical and Computer Engineering Department, Faculty of Engineering, Universitas Syiah Kuala. He is currently Director of the Directorate of Education and Learning, and a Full Professor with the Electrical and Computer Engineering Department, Universitas Syiah Kuala. He has published numerous articles on cooperative communication networks, communication technologies, and deep learning applications. His research interests include digital communications, information theory, computer and communication networks, and deep learning.