MIBINET: Real-time Proctoring of Cardiovascular Inter-Beat-Intervals using a Multifaceted CNN from mm-Wave Ballistocardiography Signal

Rafid U Murshed¹, Md. Abrar Istiak², Md. Toufiqur Rahman³, Zulqarnain B Ashraf³, Md Saheed Ullah³, Mohammad Saquib⁴

¹Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, 1000, Bangladesh
²Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, 75080, Texas, USA

Abstract

The recent pandemic has refocused the medical world’s attention on the diagnostic techniques associated with cardiovascular disease. Heart rate provides a real-time snapshot of cardiovascular health. A more precise heart rate reading provides a better understanding of cardiac muscle activity. Although many existing diagnostic techniques are approaching the limits of perfection, there remains potential for further development. In this paper, we propose MIBINET, a convolutional neural network for real-time proctoring of heart rate via inter-beat-interval (IBI) from millimeter wave (mm-wave) radar ballistocardiography signals. This network can be used in hospitals, homes, and passenger vehicles due to its lightweight and contactless properties. It employs classical signal processing prior to fitting the data into the network. Although MIBINET is primarily designed to work on mm-wave signals, it is found equally effective on signals of various modalities such as PCG, ECG, and PPG. Extensive experimental results and a thorough comparison with the current state-of-the-art on mm-wave signals demonstrate the viability and versatility of the proposed methodology.

Keywords: Cardiovascular disease, contactless measurement, heart rate, IBI, mm-wave radar, neural network

1. Introduction

Human vital signs like Heart Rate (HR), Heart Rate Variability (HRV), Respiration Rate (RR), and Oxygen Saturation (SpO₂) are important physiological indicators that reflect the physical and mental well-being of the human body. The heart pumps oxygenated and nutrient-rich blood all over the body. As the cardiac output is intimately associated with HR and stroke volume, HR is central to the cardiovascular process. HR measurement is crucial for a health monitoring system [1]. According to WHO Global Health Estimates, heart diseases such as myocardial infarction (MI), sudden cardiac death, heart attack, coronary artery disease, arrhythmia, heart valve disease, and heart infection have been the leading causes of death over the last few decades. In 2020, about 697,000 people died alone from heart disease in the United States; one in every five deaths [2] [3]. Clinical biofeedback practice heavily emphasizes the control of cardiac dynamics. To analyze biofeedback, HR, the number of heartbeats per minute is the most often measured metric. However, a healthy heart does not beat uniformly; it changes its rhythm with each beat. While HR focuses on the average number of beats per minute, HRV measures the specific variations in time between successive heartbeats. The period between heartbeats is measured in milliseconds (ms) and is referred to as the R-R interval or the inter-beat interval (IBI). The sympathetic and parasympathetic branches of the autonomic nervous system (ANS) control HRV. Especially for diabetic and post-infarction patients, it is a crucial parameter in analyzing the behavior of the sympathetic and parasympathetic functions of the ANS [4]. An effective way of estimating HRV is measuring the variation in IBI [5], which can be used to detect probable cardiovascular disorders [6, 7].

Electrocardiogram (ECG) [8] [9] and Photoplethysmography (PPG) are among the recognized methods for measuring HRV indices [10]. These systems require the patients to be in constant communication with humans. However, this can be challenging for remote patient monitoring, especially for elderly people living alone or recently discharged patients. Another disadvantage of the aforementioned systems is the limitation of mobility and freedom of the patients. In addition, the electrodes used by ECG and PPG may not only cause unnecessary discomfort but also potentially exacerbate symptoms. Moreover, the palpation of the probes could lead to false alarms and cause alarm fatigue [11] [12]. Radar-based contactless telemeasurements offer a very comfortable means of continuous vital signal monitoring, which is essential for detecting early onsets of cardiovascular diseases. In this paper, we pursue that line of inquiry.

Millimeter wave (mm-wave) based technologies are ideal alternatives for contactless, continuous measurement of human vital signals [13]. The shorter wavelengths of mm-waves (30–300 GHz) allow for greater spatial resolution. To this end, a number of approaches have been explored to implement mm-wave systems in different biological and medical applications. The most notable applications include contactless measurement of arterial pulses [14], cancer diagnosis [15] [16] and dental diagnosis [17]. In addition, the use of high-resolution mm-wave array beamformers has increased in medical imaging, gesture recognition, and navigation in recent years. Non-thermal, low-intensity electromagnetic radiation is used in mm-wave ther-

arXiv:2211.07776v1 [eess.SP] 14 Nov 2022

Preprint submitted to Biomedical Signal Processing and Control November 16, 2022
aply (MWT), a novel and revolutionary method \cite{18} of treating patients. The implementation of mm-wave FMCW radar \cite{19} is used to measure breathing and HRs. To enhance the effectiveness and feasibility of these technologies in real-time, their implementation must take into account both computational efficiency and optimal precision. The above studies do not analyze both the computational efficiency and optimal precision of the used mm-wave technologies simultaneously. Unlike those, our proposed mm-wave-based technique, MIBINET, can estimate the IBI values with optimal precision in real-time.

To the best of our knowledge, MIBINET is the first neural network-based approach that operates on pre-processed mm-wave data to estimate the instantaneous IBI. It adopts a rolling window-based pre-processing approach to confine the input data to a predetermined size for conveniently feeding into a neural network’s input layer. The network’s output is also reprocessed with a novel rolling window averaging approach followed by various traditional post-processing filters to improve the estimation accuracy; see Figure 1. Our numerical results demonstrate that the proposed MIBINET is capable of outperforming the state-of-the-art techniques in terms of IBI estimation precision by more than 5%. Below is a summary of the other major contributions of this investigation:

1. Our model operates on a fixed-length input, which requires zero padding. To make the model unbiased towards zero-padded tails, a novel zero-padding augmentation is introduced via arbitrarily shifting the original signal inside the input data.

2. MIBINET model architecture is based on lightweight 1D-CNN for real-time use. It achieves an especially high cumulative correlation coefficient and produces a very meager root-mean-square error (RMSE) over 11 different test subjects compared to the existing state-of-the-art.

3. A customized loss function is introduced to optimize the convergence of training. This loss function, categorically designed to reduce outliers, improves the correlation coefficient significantly and substantially lowers the RMSE.

4. In order to generalize our proposed strategy across a variety of patients, the data set is enriched with synthetic IBI values. This augmentation enhances the correlation coefficient by a significant margin and noticeably reduces the RMSE. Our experiments demonstrate the robustness of the technique across a variety of patients and show promise even in anomalous cases.

5. Proposed MIBINET is a flexible network that also operates efficiently on signal modalities, such as PPG, phonocardiograph (PCG), and ECG, in which it offers nearly ideal performance.

The rest of this paper is organized as follows: Section 2 contains a review of the current state of contact-free vital sign monitoring. MIBINET is proposed in Section 3, along with the description of pre and post-processing techniques. Section 4 includes the numerical results. Finally, a brief discussion of the findings, limitations, and strengths of MIBINET is provided in Section 5. Regarding notation, scalars, and vectors are represented in lower and bold lower cases, respectively. The \( n \)th element of vector \( \mathbf{a} \) is denoted by \( a(n) \).

2. Related Works

A significant amount of research has focused on non-invasive contactless monitoring of vital signs (e.g., HR, RR, HRV) due to its many advantages. Non-invasive and contactless optical approaches for heartbeat monitoring based on optical Doppler interferometry and laser have been proposed in \cite{20, 21}. Here, the photo-EMF pulsed laser vibrometer (PPLV) is studied where the subjects were instructed to first exhale and then hold their breath for as long as they could after inhaling. This is to avoid any potential muffling of the cardiac signals by the subjects’ respirations. However, it might cause unnecessary discomfort to the patients. Fengyu Wang et al. \cite{22} proposed contactless HRV monitoring using mm-wave radio. First, they developed a user-locating target detector without calibration. Heartbeat signal extractors then optimize the decomposition of chest-movement-modulated channel information to uncover the desired signal. Now the pulse signal’s peak position can be used to evaluate HRV parameters for each target utilizing the IBI values. This method, mmHRV, can assess HRV with a median IBI error of 28 ms (w.r.t 96.16\% accuracy) for 11 players in the line of sight (LOS). For non-LOS, it is 31.71 ms. However, this proposed method exhibited considerably uneven errors among the participant subjects. Zhang et al. \cite{23} suggested radio signal-based contactless MI detection. This work establishes MI detection using RF signals, providing contactless, non-intrusive, continuous home monitoring for MI hazards. They also proposed heartbeat signal segmentation and MI detection algorithms. Extensive evaluations have been conducted to confirm the effectiveness of Health-Radio. However, this approach has not been generalized to work over different signal modalities such as PPG, ECG, and PPG.

Radar technology is one of the most promising possibilities for contactless, non-invasive monitoring of vital signs such as cardiac signals. Impulse radio ultra-wideband (IR UWB) radar was used to monitor vital signs \cite{24, 25}. While monitoring, this IR UWB-based technique exploits signal properties completely disregarding any object qualities with which this signal interacts. Body-coupled antennas and UWB pulsed radar
in-body monitoring of lungs and heart motion are used in this scenario. The radar has to be installed at a specific location. Moreover, sudden oscillations cause abrupt phase fluctuations, which impact HR predictions. Many methods for calculating HR from UWB radar data have recently been developed. Their performance, however, is still insufficient for practical applications. Frequency-modulated continuous wave (FMCW) radar [24], continuous wave (CW) doppler radar for HR monitoring, and respiration monitoring has been developed significantly during the past few years [28-30]. Because CW radar or doppler radar does not capture the target’s range or distance, FMCW radar is designed to overcome this issue due to its range and radial velocity measurement. Since FMCW radar operates at lower transmit power, the received signal will not only be distorted by the environment but also could be weak. Although beamforming and range-gating approaches can separate the signal of interest from the noise, several difficulties, such as random body motions, must be addressed before radar-based non-contact measurements can be implemented in real-world applications. Sakamoto, Takuya, et al. [31] addressed the use of a UWB radar system for estimating the human HR precisely. The performance of the proposed approach was demonstrated through measurements. Target classification and type recognition are achievable with UWB radar because the received signal contains information not just regarding the target as a whole but also about its individual component elements [32]. The suggested approach estimates the HR efficiently and correctly by using the feature points of a radar signal. Nonetheless, as heartbeat waveforms vary, even within the beat-to-beat interval, Fourier and periodicity-based approaches are ineffective for estimating instantaneous HR in real time. Consequently, a Wavelet-Transform (WT) based method [33] has been proposed for faster HR detection using doppler radar because of the insufficient frequency resolution of the Fourier transform (FT). Unfortunately, the WT approach requires extensive signal processing in order to identify HR.

In 2010, researchers started applying machine learning (ML) in the field of ballistocardiography signals. Bruser et al. [34] brought an idea of an unsupervised modified K-means clustering training algorithm to estimate parameters for the BCG signal. Then, the parameters followed by the heartbeat detection and refinement process combine three indicators for localizing the heartbeat. Collectively, the method called BEAT has to be re-trained when facing a significant change in the BCG signal. Later, a real-time approach for identifying individual heartbeats without requiring intensive signal processing was studied in [35]. It is an Artificial Neural Network (ANN) based Heartbeat Detection Technique using a state-of-the-art mm-wave radar sensor. The ANN is trained on the raw radar signal, offering computational simplicity while estimating IPIs (inter-pulse intervals) with relatively low accuracy. Hence, a signal of high HRV is not suitable for this study. In addition, this shallow ANN model has a 2% chance of missing a pulse. Nowadays, adopting deep learning methods enables the proliferation of automatic, portable, non-invasive data-driven HR monitoring. In 2019, Dwaipayan et al. [36] proposed a temporal model CorNET (CNN + LSTM) method to predict HR and biometric identification. This was evaluated on two subjects and consequently lacked subject bias. This model needs longer training in order to adapt to the HR variability in an ambulant environment. [37] Xiangmao et al. [37] proposed a method to estimate HR after acquiring clean input PPG data from a denoising CNN (DCNN). Although DCNN adds more robustness to real-life artifacts, it adds overhead in time complexity. On a different note, the first triumphant attempt to predict pulse rate (PR) from facial video data after the surge of computer vision was the proposition of PRnet [38]. This method leverages the synergy of 3D convolution and LSTM to predict pulse rate from spatio-temporal features with less error. It requires a minimum of 2 seconds (60 frames) to effectively measure PR, which limits the method from being a real-time pragmatic solution.

3. Methodology

In this section, we describe the working principle of MIBINET. Since it is a DNN based-technique, it needs to be trained and validated on labeled data. First, we describe how the labeled data is collected.

3.1. Data Acquisition

Figure 2 is a schematic representation of the data collecting and processing procedure following the work presented in [39]. The RF front end shown in this figure consists of a Six-Port utilized as a quadrature interferometer for radar application. The six-port has two input signals and four output signals, where two input signals consist of a reference signal at a defined frequency and a received signal reflected from the target. The antenna direction was set perpendicular to the test subject’s thorax surface to maximize the signal quality. The baseband board back end (BB Back End) receives radar signals from the RF front end and digitizes the signals. It is equipped with an ECG and respiration sensor (RS) for simultaneous sampling. The digitized raw signals were subsequently received by a computer equipped with a PCG that served as a reference sensor. PCG signals are stored after re-sampling and synchronization.

Unlike [39], we tackled variation with augmentation and rolling window averaging techniques.

Figure 2: Data collection scheme - The figure depicts how the data was collected.
Let us now take a brief glance at the processed data, which consists of the raw radar data along with the filtered data from the respiration sensor, the ECG leads, the PCG data, and the Radar heart sounds. Figure 3 provides a breakdown of the collected data. As we can see, 11 subjects participated in this measurement process. Although the total data consist of 13376 s of recordings, only the default scenarios in [39] are of interest to us. In the following sub-section, we describe our novel rolling window-based pre-processing scheme, which allows us to seamlessly meta-morph the collected dataset in a form compatible with deep neural networks.

![Figure 3: Total signal duration of each subject in the dataset.](image)

### 3.2. Dataset Distribution

Following standard practice, the entire dataset was permuted into 11 folds, with each subject’s data appearing only once in the test set of each fold. Each fold consists of separate training, validation, and test sets. The folds are numbered according to the ID of the subject in the test set. For instance, fold one indicates that the data of the first subject is in the test set. The training and validation sets consist of the data of the remaining ten users, with eight in training and two in validation.

The distribution of the IBI values from the data of different test subjects reveals the typical variations usually found in clinical trials. For example, as shown in Figure 3(a), the median values of the participant test subjects vary from 0.6 to 0.8 seconds (s), with subjects 3 and 10 having comparatively lower IBI values while subjects 7, 8, and 9 exhibit relatively high values.

Analyzing the distribution of each fold shows some disparities between the range of IBI values in the training and test sets. In particular, folds 3 and 10 display significant variations, making it quite challenging to train deep learning-based models. This leads us to propose a novel data augmentation approach to training the deep models, which will be elaborated on next.

### 3.3. Synthetic IBI Generation Augmentation

Due to the scarcity of lower IBI values in the training sets of folds 3 and 10, the model struggles to perform well on the test sets, which contain a relatively lower range of IBI values. We resolve this issue by generating synthetic radar-HS signals to inject smaller IBI values into the training set. To synthesize radar-HS signals characterized by smaller IBI values, specifically in the range of 0.5 – 0.6 s, we first start with signals that have IBI values in the higher range, preferably in the region of 0.9 – 1.1 s. We then left-shift those signals in the time domain by 0.4 – 0.5 s and add them with their shifted versions as depicted in Figure 5a. Mathematically, it can be expressed as follows

\[
x_{\text{aug}}(t) = x_{o}(t) + x_{o}(t + \alpha),
\]

where \(x_{\text{aug}}(t)\) and \(x_{o}(t)\) denote the augmented and original signals respectively. Here, \(\alpha\) is a uniform random variable between 450 ms and 550 ms. Now the R-peaks of the augmented signal are the union of the R-peaks from the original signal and the left-shifted signal, as shown in Figure 5b. Due to such augmentation, the distribution of the resulting training set becomes much more consistent, as can clearly be observed in Figure 4.

In order to further facilitate our neural network-based approach, we supplement the processed signals with an additional rolling window-based pre-processing scheme, which will be discussed next.

### 3.4. Rolling Window-based Pre-processing and Augmentation

IBI is the time interval between two consecutive R-peaks, as shown in Figure 6(a). Here, it can be noticed that the rolling window extends over 8 consecutive R-peaks of the processed signals, each containing 7 IBIs; see Figure 6(b). Although the specific size of the rolling window is not based on exact theoretical calculation, our empirical results demonstrate that the proposed size is near optimal for this study to trade-off between computational time and performance. It is well known that the dimensions of the inputs and outputs of neural networks must be predefined. However, the available radar data does not adhere to those constraints. Hence, we need to restrain the dimensions of the inputs and outputs to a predefined value. Here, by using a rolling window of size 8, we are able to restrict the size of the output to a vector of size 7. In our dataset, each input consists of a windowed version of the original data containing 7 IBIs, and the maximum windowed input length is found to be 4885 samples. In order to restrict the dimension of the input data, we zero-pad all the inputs to a size of 4910 samples. These zeroes are inserted randomly at the start and end of each window to remove any prospective bias during training. The starting point of every successive rolling window is randomly selected between two consecutive R-peaks to remove any regional bias.

### 3.5. Architecture of the MIBINET

The primary goal of our neural network-based approach is to minimize computational time as well as achieve high prediction accuracy. With that in mind, we propose the MIBINET architecture with a comparatively smaller number of parameters than other contemporary networks and estimate IBI values with higher accuracy. The MIBINET is a lightweight convolutional neural network that relies on both standard spatial convolutional layers and depth-wise separable convolutional layers to extract distinguishable features from the one-dimensional input.
data. The inputs are normalized before being fed to the initial batch normalization layer that adjusts the mean and variance of the input data for faster convergence during training. Next, we have a sequence of convolutional and pooling layers with different kernel numbers and sizes. The number of kernels and their sizes are determined through extensive trial and error. Finally, all the extracted features are fed into a fully connected network of dense layers. The last dense layer predicts the output, which is an array of length seven consisting of consecutive IBI values. The swish activation function was used in the convolutional layers as it performs better than ReLU after convolutional layers, and ReLU was used in the fully connected dense layers. A schematic diagram of the entire model architecture is shown in Figure 7. Multiple 1D depth-wise separable convolutional layers reduce the number of parameters in MIBINET to make it even more computationally efficient.

3.6. Custom Loss Function

The loss function plays a vital role in supervised learning algorithms such as feed-forward networks. It determines how the predictions approach the target labels. Since our goal is to measure the IBI values consistently with high precision, we not only require the error between the predictions and true values to be small but also want there to be a high correlation between the predicted values and the true values. To achieve this, we design a novel weighted loss function that takes into account the aforementioned considerations. The designed loss function is a weighted sum of three well-known loss functions for regression tasks, namely, the mean squared error (MSE) loss, the Huber loss (HL), and the mean absolute error (MAE) loss. While the MSE is too sensitive to outliers, the MAE weights all the errors equally, disregarding outliers completely. The HL provides a good balance between both the MSE and the MAE. Finally, to
ensure a high correlation between the predicted outputs and the actual value, we add another component to our weighted loss in the form of correlation coefficient loss. The resultant loss function can be expressed as

\[ L_{corr} = w_1 \times (1 - r^2) + w_2 \times L_{HL} + w_3 \times \epsilon^2 + w_4 \times |\epsilon|, \]

where \( L_{HL} \) denotes the well-known Huber-loss value, \( r \) represents the Pearson correlation coefficient, \( \epsilon \) is the error in the prediction, and \( w_1, w_2, w_3 \), and \( w_4 \) are the weights applied to the correlation loss, Huber-loss, MSE, and MAE, respectively. The exact values used for \( w_1, w_2, w_3 \), and \( w_4 \) are determined through extensive trial and error to best suit the required task. It is also noteworthy that none of the well-known regression losses (i.e., MSE, MAE, mean absolute percentage error, etc.) except HL solely performed well in our experiments.

### 3.7. Post-Processing

We first process the outputs by a rolling window averaging scheme to smooth out the predictions. This exploits our rolling window-based pre-processing scheme, where there is considerable overlapping between the samples of neighboring windows. More precisely, each window retains seven R-peaks from the previous window and hence has exactly six same IBIs. This is utilized during the post-processing, where all predictions on the same initial IBI value are averaged, and hence the errors are further reduced. Following the post-processing schemes of [39], we also use a median filter of length 5 and a moving average filter of length 6, which boosts the reliability of the model’s predictions.

As described earlier, each rolling window consists of seven consecutive IBIs; see Figure 6. Consequently, this leads to the overlapping of successive windows, and the intersection of each of seven such successive windows contains a single common IBI as shown in Figure 8. Furthermore, our model predicts the IBI values from each window separately. Thus, it gives seven distinct predictions for each IBI value. Averaging those IBI values yields a more accurate prediction. Mathematically, this can be expressed as

\[ z(k) = \frac{\sum_{i=1}^{7} y_{k-7+i}(8-i)}{7}; \quad k = 7, 8, \ldots \]

where \( z(k) \) denotes the \( k \)th IBI value of a sample and \( y_j \) contains seven consecutive IBI predictions from the \( j \)th rolling window.

### 4. Experimental Results

In this section, we discuss the experimental data and analysis used to evaluate the performance of our proposed MIBINET. We assess the efficacy of our proposed system using the Pearson correlation coefficient, and root means square error (RMSE). These metrics are chosen due to their suitability for regression tasks. In addition, the established state-of-the-art in [39] is used as the baseline methodology.
Figure 7: Proposed MIBINET architecture.

Figure 8: Rolling window-based pre-processing.

4.1. Evaluation Metrics

The expression of the Pearson correlation coefficient is

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}},$$  \hspace{1cm} (4)

where $x_i$ and $y_i$ are the $i^{th}$ predicted IBI and the ground truth IBI, respectively. Their sample means are denoted by $\bar{x}$ and $\bar{y}$. The term $n$ denotes the total number of IBIs in the sample. The other metric, RMSE, can be defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}}.$$  \hspace{1cm} (5)

4.2. Training Setup

All simulations are conducted using Google Colaboratory, a Python development environment that runs in the browser using Google Cloud and provides free access to powerful graphical processing units (GPU). Our proposed MIBINET and peripherals are implemented in Python 3.7 utilizing TensorFlow 2.9 and a Tesla T4 GPU provided by Google Collaboratory. All codes are executed with the same setup to enable an accurate and fair comparison of different processes. We have trained our network by utilizing a batch size of 1024 for 200 epochs. The learning rate is tuned using an exponential decay scheduler as shown in Figure 9. The initial learning rate is set to 0.007. It is decreased by half every 40 epochs to 0.0018 after 80 epochs. Then, it is set to 0.00007 and is decreased by 10 times every 40 epochs through the next 120 epochs. All experiments used Adam [41] optimizer along with momentum with a decay of 0.9. In addition, the best weights were saved based on the validation-weighted metric. The weighted metric can be expressed as

$$M_{\text{weighted}} = \alpha_1 \times (1 - r^2) + \alpha_2 \times \epsilon^2 + \alpha_3 \times |\epsilon|,$$  \hspace{1cm} (6)

where the weights $\alpha_1$, $\alpha_2$, and $\alpha_3$ are chosen to be 10, 0.1, and 0.1, respectively, to ensure equal importance on correlation coefficient and RMSE. Final model predictions are compared after the post-processing as mentioned in sub-section 3.7. It is to be noted that, for every fold, the model was trained with the same
parameters. In all our simulations, we have set $w_1 = 0.002$, $w_2 = 1.0032$, $w_3 = 0.0096$ and $w_4 = 0.002$ in (3) as those are found to be near optimal.

### 4.3. MIBINET vs. HSMM Method

The Radar-HS dataset in [39] has recordings sampled at multiple frequencies. To maintain consistency, all the recordings are resampled to 500 Hz. We then prepare the dataset for MIBINET as described in Section 3.2. Following the aforementioned pre-processing and post-processing steps, the final results are generated, and the fold-wise results are given in Table 1. The fold-wise results of the HSMM method, the existing state-of-the-art, are also provided here for ease of comparison.

![Learning rate used for MIBINET training with the number of epochs.](image)

**Figure 9**

Table 1: Fold-wise performance

| Test User No | Correlation Coefficient, r (%) | RMSE (ms) |
|--------------|-------------------------------|-----------|
|              | HSMM Based [42]               | MIBINET   | HSMM Based [42]               | MIBINET   |
| 1            | 78.41                         | 97.08     | 59.58                         | 19.77     |
| 2            | 91.93                         | 98.99     | 25.67                         | 8.45      |
| 3            | 94.09                         | 92.90     | 5.59                          | 11.88     |
| 4            | 93.34                         | 97.56     | 19.59                         | 16.99     |
| 5            | 89.32                         | 96.04     | 38.51                         | 24.10     |
| 6            | 61.69                         | 88.91     | 89.58                         | 24.34     |
| 7            | 87.92                         | 97.01     | 53.76                         | 24.43     |
| 8            | 98.75                         | 99.18     | 7.32                          | 5.70      |
| 9            | 85.68                         | 98.80     | 77.63                         | 17.21     |
| 10           | 96.91                         | 93.19     | 13.81                         | 39.19     |
| 11           | 81.40                         | 98.58     | 41.75                         | 11.54     |

In Table 1 it can be noticed that apart from folds 3 and 10, our proposed MIBINET significantly outperforms the state-of-the-art HSMM in both metrics. To substantiate the robustness and efficacy of our proposed approach compared to its counterpart, we test their performance over various datasets of different modalities. These particular ones are chosen because of their availability and authenticity. In the following subsections, we present a brief account of these datasets.

### 4.3.1. PCG Dataset

Reference [39] provides the phonocardiograph (PCG) signals corresponding to the Radar-HS data from the previously considered 11 test subjects. With the similar split, recording number, sampling rate, and ground truth of Radar-HS, it just differs in the signal modality.

### 4.3.2. PTB-XL ECG Dataset

A 12-lead electrocardiography [43] dataset comprising of 21837 records from 18885 patients is used for comparison. The sampling rate is 100 Hz in the given recordings. However, we resample the dataset to 500 Hz to maintain conformity with the Radar-HS dataset. As it has a massive number of unique subjects under test, it can test the subject dependency of an algorithm. Hence, this dataset is used to demonstrate the universality of MIBINET. For dataset validity, it is split sequentially into train, validation, and test sets with a ratio of 60:20:20 based on the number of subjects. For the sake of simplicity, only the signal from lead 2 is considered, as the QRS complex here is more prominent compared to the ones in the other leads. It is to be noted that, ECG signals are annotated using BioSPPy [44] as suggested in [43].

### 4.3.3. Wrist PPG

Jarchi et al. [45] released a PPG dataset collected using smartwatches from 8 participants. Here, all the recordings are sampled at a 256 Hz sampling rate, and 19 recordings are recorded in four different conditions, i.e., running, walking, fast-easy bike riding, and slow-difficult bike riding. We include this dataset as it has heavy motion artifacts, which is typical in real-life scenarios. Here, the R-peaks are determined from the given synchronized ECG signal. It also contains high HRV cases due to sudden initiation of motion and abrupt stoppages. Utilizing the 19 recordings, 19-fold cross-validation results are reported where fifteen, three, and one folds are in training, validation, and testing, respectively.

Now, we present the cumulative performance of MIBINET and HSMM in the above four (Radar-HS, PCG, ECG, and PPG) datasets; see Table 2. Here, we observe the following:

1. For the Radar-HS dataset, MIBINET provides an overall performance improvement of more than 5% in terms of correlation and 27 ms in RMSE, compared to HSMM.

| Dataset     | Methodology | r (%) | RMSE (ms) |
|-------------|-------------|-------|-----------|
| Radar-HS [39] | HSMM        | 93.66 | 47.94     |
|             | MIBINET     | 98.73 | 20.69     |
| PCG [39]    | HSMM        | 87.02 | 53.52     |
|             | MIBINET     | 98.76 | 21.02     |
| PTB-XL ECG [43] | HSMM | 71.83 | 140.55    |
|             | MIBINET     | 99.60 | 11.93     |
| Wrist PPG [45] | HSMM | -5.55 | 2641.43   |
|             | MIBINET     | 56.01 | 182.28    |
2. In the PCG dataset, our algorithm performed even better than in Radar-HS. A higher correlation coefficient and lower RMSE portrays this performance improvement.

3. For the ECG dataset, the performance improvement of MIBINET over HSMM is even more prominent compared to the previous two datasets (correlation coefficient improvement > 27%).

4. Although the PPG signal is more challenging due to its miscellaneous artifacts and high noise content, MIBINET still outperforms the HSMM by a massive margin. The results in the final row of Table 2 show the robustness of our algorithm in challenging conditions induced by this PPG modality.

4.4. Performance of Proposed Loss Function

A major contribution of this work is the weighted custom loss introduced in (2). The efficacy of the proposed loss is tested on various modalities of the dataset. We compare the performance of the proposed weighted loss with that of the best-performing existing regression loss, namely Huber-loss; see Table 3. As can be observed, the proposed loss function assists MIBINET in improving its performance substantially in all datasets. This signifies that our proposed loss is not dataset-specific but performs equally well over other datasets.

Next, we conduct Bland Altman analysis to show the agreement between predicted and ground truth IBI; see Figure 10. In this analysis, we consider four different cases. The comparison between the true and predicted values is shown through the deviation plots and the Bland Altman plots of the IBI values and the corresponding BPM values. The plots in Figure 10 show the mean and standard deviation in radar heart sound data. The LoA (limits of agreement) is lower in our method than that of state-of-the-art, which is shown in Figure 10(b - c) and 10(k - l). The deviation plots also reveal a significant improvement over the HSMM method. We can clearly observe that there are some errors in the region of 0.1 s to 0.25 s for the HSMM deviation plot shown in 10(a). Figure 10(d - f) vs. Figure 10(j - l) shows the clear distinction between Huber loss and proposed loss, respectively. As we can observe, the regressor with Huber loss has many errors beyond 0.05 s in the deviation plot. Moreover, the LoA and the mean are better for the regressor with the proposed loss.

4.5. Effectiveness of Synthetic IBI Generation

We again perform the Bland-Altman analysis to demonstrate the potency of our synthetic IBI generation process. It can be clearly noticed from Figure 10(b - i) and 10(k - l) that, our proposed augmentation method decreases the error and gives a better LoA. In purely quantitative terms, the RMSE is decreased by almost 3 ms for the Radar-HS dataset. Similarly, the results obtained through PCG data also show the effectiveness of this novel synthetic IBI augmentation. In PPG modality, we observe a performance improvement of around 10% in correlation coefficient and 40 ms in RMSE, as displayed in Table 4.

4.6. Comparison with Related Models

Besides HSMM, we compare the performance of the MIBINET model against the current state-of-the-art lightweight models applicable to the Radar-HS dataset provided in [39]. To evaluate the performance of the neural network models alone, we test them without post-processing, and the results are presented in Table 5. It is to be noted that all the peripheral settings were kept exactly the same for all the models to ensure a fair comparison. We notice that MIBINET outperforms the current state-of-the-art lightweight deep learning models by a margin of at least 3% in the correlation coefficient metric. Importantly, MIBINET, having the least number of parameters among all these models, claims itself as the best-suited real-time solution for contactless vital signs monitoring.

5. Conclusions

This work proposed a comprehensive DNN-based strategy (MIBINET) to surpass the present state-of-the-art (HSMM) in measuring IBI from mm-wave collected data. MIBINET is a unique lightweight architecture that features novel post-processing techniques. It achieved a cumulative 98.73% correlation coefficient and produced a meager 20.69 ms RMSE score.
Figure 10: Performance comparison based on error: (a) - (c) are plots for state-of-the-art HSMM-based algorithm. (d) - (f) are plots for Huber loss with augmentation. (g) - (i) are plots for proposed loss without augmentation. (j) - (l) are for proposed loss with augmentation. In the first column, error deviation plots; in the second column, Bland Altman plots of deviation between ground truth (Radar-HS) and predicted values; and in the third column, Bland Altman plots of deviation in the BPM unit are depicted.
over 11 different test subjects on Radar-HS data. This was a significant improvement over HSMM, which offered a 93.66% correlation coefficient and 47.94 ms RMSE. To demonstrate the versatility of our proposed approach, it was further evaluated on datasets containing ECG, PCG, and PPG signals. Its performance was found to be superior to HSMM in all of them.

A custom-weighted regression loss was developed to train MIBINET, and a unique weight-saving approach was used based on a weighted metric. The novel weighted loss function improved the correlation coefficient by more than 1.4% and lowered the RMSE by more than 21 ms in Radar-HS data compared to the commonly used Huber-loss function. The robustness of the proposed loss function was demonstrated on other signal modalities (PCG, ECG, and PPG). A novel IBI augmentation technique was also introduced to negate the effects of dataset imbalance. This augmentation enhanced the correlation coefficient by more than 10% in the PCG modality while reducing the RMSE by 40 ms and noticeably improved the performance in Radar-HS data as well.

Although the proposed synthetic IBI augmentation technique improves the performance in Radar-HS and PCG modality, it cannot be applied to the ECG and PPG data due to their non-bursty nature. The ECG and PPG data do not contain the symmetry required to perform such augmentation. The heavy artifacts induced in the PPG signal due to sporadic motion and the non-static nature of measurements require more training data to achieve performance comparable to Radar-HS, PCG, and ECG. When the HRV is too high, as observed in the PPG modality, the performance of MIBINET slightly degrades as it has been trained on minimal data. Like all other DL-based approaches, MIBINET suffers from the dataset imbalance bottleneck. If it is operated on an imbalanced dataset, the performance slightly degrades, as observed from the performance before augmentation. However, all these issues can be mitigated using a well-balanced large dataset which is relatively easy to obtain nowadays.

The results presented here indicate that MIBINET has great potential in all these diverse signal modalities. From a developmental perspective, this work has a vast scope. Using raw radar data to create an end-to-end model could provide another layer of functionality to the model. Multi-user scenarios can be considered to make the MIBINET design even more practicable. Lastly, a multi-modal approach of combining multiple signals or fusion of multiple streams instead of single channel data may open up a new dimension in the field of contactless vital signs monitoring.

Declaration of Competing Interest
The authors report no declarations of interest.

Conflict of interest
The authors declare no conflict of interest.

References
[1] E. Thaulow, J. E. Eriksson, How important is heart rate?, Journal of Hypertension 9 (Supplement 7) (1991) S27–S30. doi:10.1097/00004872-199112007-00007
URL https://doi.org/10.1097/00004872-199112007-00007
[2] C. for Disease Control, Prevention, N. C. for Health Statistics, et al., About multiple cause of death, 1999–2019. cdc wonder online database website (2019).
[3] C. W. Tsao, A. W. Aday, Z. J. Almarzooq, A. Alonso, A. Z. Beaton, M. S. Bigno, J. K. Bohnmeier, A. E. Buxton, A. P. Carson, G. Commodore-Mensah, et al., Heart disease and stroke statistics—2022 update: a report from the american heart association, Circulation 145 (8) (2022) e153–e639.
[4] O. V. Volodina, https://pnsjournal.wordpress.com/2022/07/01/volodina-3/, Formation of future teachers’ worldview culture by means of foreign-language education, P Sci Edu 57 (3) (2022) 126–159.
[5] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, J. S. Suri, Heart rate variability: a review, Medical and biological engineering and computing 44 (12) (2006) 1031–1051.
[6] A. K. Schuster, J. E. Fischer, J. F. Thayer, D. Mauss, M. N. Jarzok, Decreased heart rate variability correlates to increased cardiovascular risk, International Journal of Cardiology 203 (2016) 728–730.
[7] F. Shafer, J. P. Ginsberg, An overview of heart rate variability metrics and norms, Frontiers in public health (2017) 258.
[8] M. Malik, Heart rate variability: Standards of measurement, physiological interpretation, and clinical use: Task force of the european society of cardiology and the north american society for pacing and electrophysiology, Annals of Noninvasive Electrocardiology 1 (2) (1996) 151–181.
[9] M. T. La Rovere, J. T. Bigger Jr, F. I. Marcus, A. Mortara, P. J. Schwartz, A. A. Tone, R. A. M. I. Investigators, et al., Baroreflex sensitivity and heart-rate variability in prediction of total cardiac mortality after myocardial infarction, The Lancet 351 (9101) (1998) 478–484.
[10] K.-c. Lan, P. Rakin, W.-F. Kao, J.-H. Huang, Toward hypertension prediction based on ppg-derived hrv signals: A feasibility study, Journal of medical systems 42 (6) (2018) 1–7.
[11] R. A. Harrigan, T. C. Chan, W. J. Brady, Electrocardiographic electrode misplacement, misconception, and artifact, The Journal of emergency medicine 43 (6) (2012) 1038–1044.
[12] R. R. Bond, D. Finlay, C. Nugent, C. Breen, D. Guldenring, M. Daly, The effects of electrode misplacement on clinicians’ interpretation of the standard 12-lead electrocardiogram, European journal of internal medicine 23 (7) (2012) 610–615.
[13] Z. Yang, P. H. Pathak, Y. Zeng, X. Liran, P. Mohapatra, Monitoring vital signs using millimeter wave, in: Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing, 2016, pp. 211–220.
[14] J. E. Johnson, O. Shuy, C. Kim, C. Liao, Wearable millimeter-wave device for contactless measurement of arterial pulses, IEEE transactions on biomedical circuits and systems 13 (6) (2019) 1525–1534.
[15] F. Topfer, S. Dudovor, J. Oberhammer. Millimeter-wave near-field probe designed for high-resolution skin cancer diagnosis, IEEE Transactions on Microwave Theory and Techniques 63 (6) (2015) 2050–2059.
[16] S. Di Meo, P. F. Espin-Lopez, A. Martellosio, M. Pasian, G. Matrone, M. Bozzi, G. Magenes, A. Mazzanti, L. Perregrini, F. Svelto, et al., On the feasibility of breast cancer imaging systems at millimeter-waves frequencies, IEEE Transactions on Microwave Theory and Techniques 65 (5) (2017) 1795–1806.
[17] S. Hoshi, Y. Nihawa, K. Kawai, S. Ebisu, Application of microwaves and millimeter waves for the characterization of teeth for dental diagnosis and treatment, IEEE Transactions on Microwave Theory and Techniques 46 (6) (1998) 834–838.
[18] F. Lin, W. Hu, A. Li, Millimeter-wave technology for medical applications, in: 2012 IEEE MTT-S International Microwave Workshop Series on Millimeter Wave Wireless Technology and Applications, 2012, pp. 1–1. doi:10.1109/IMWS2.2012.6338252
[19] G. Alizadeh, G. Shaker, J. C. M. D. Almeida, P. P. Morita, S. Safavi-Naeini, Remote monitoring of human vital signs using mm-wave fmcw radar, IEEE Access 7 (2019) 54958–54968. doi:10.1109/ACCESS.2019.2912956
[20] J. J. Bancifra, Tarlac State University, Tarlac City, Philippines, https://orcid.org/0000-0003-0641-1305, Supervisory practices of depart-
ment heads and teachers’ performance: Towards a proposed enhancement program, APIAEJ - Journal ay Asia Pacific Journal of Advanced Education and Technology (2022) 23–35.

[21] C.-C. Wang, S. B. Trivedi, F. Jin, S. Stepanov, Z. Chen, J. Khurpin, P. Rodrigue, N. S. Prasad, Human life signs detection using high-sensitivity pulsed laser vibrometer, IEEE Sensors Journal 7 (9) (2007) 1370–1376. doi:10.1109/JSEN.2007.9095041

[22] F. Wang, X. Zeng, C. Wu, B. Wang, K. R. Liu, mmhvr: Contactless heart rate variability monitoring using millimeter-wave radar, IEEE Internet of Things Journal 8 (22) (2021) 16623–16636.

[23] J. Zhang, Y. Chen, T. Chen, et al., Health-radio: towards contactless myocardial infarction detection using radio signals, IEEE Transactions on Mobile Computing (2020).

[24] A. Ahmad, J. C. Roh, D. Wang, A. Dubey, Vital signs monitoring of multiple people using a fmcw millimeter-wave sensor, in: 2018 IEEE Radar Conference (RadarConf), 2018, pp. 1450–1455. doi:10.1109/RADAR.2018.8377878

[25] E. Schires, P. Georgiou, T. S. Lande, Vital sign monitoring through the back using an uwb impulse radar with body coupled antennas, IEEE Transactions on Biomedical Circuits and Systems 12 (2) (2018) 292–302. doi:10.1109/TBCAS.2018.2799322

[26] S. Wu, T. Sakamoto, K. Oishi, T. Sato, K. Inoue, T. Fukuda, K. Mizutani, H. Sakai, Person-specific heart rate estimation with ultra-wideband radar using convolutional neural networks, IEEE Access 7 (2019) 168484–168494. doi:10.1109/ACCESS.2019.2952494

[27] P. Wang, F. Qi, M. Liu, F. Liang, H. Xue, Y. Zhang, H. Lv, J. Wang, Noncontact heart rate measurement based on an improved convolutional sparse coding method using ir-uwb radar, IEEE Access 7 (2019) 158492–158502. doi:10.1109/ACCESS.2019.2950423

[28] G. Vinci, S. Lindner, F. Barbon, S. Mann, M. Hofmann, A. Duda, R. Weigel, A. Koelpin, Six-port radar sensor for remote respiration rate and heartbeat vital-sign monitoring, IEEE Transactions on Microwave Theory and Techniques 61 (2013) 2093–2100. doi:10.1109/TMTT.2013.2247055

[29] C. Li, J. Ling, J. Li, J. Lin, Accurate doppler radar noncontact vital sign detection using the relay algorithm, IEEE Transactions on Instrumentation and Measurement 59 (3) (2010) 687–695. doi:10.1109/TIM.2009.2025986

[30] V. L. Petrović, M. M. Janković, A. V. Lupšić, J. S. Popović-Božović, High-accuracy real-time monitoring of heart rate variability using 24 ghz continuous-wave doppler radar, IEEE Access 7 (2019) 74721–74733. doi:10.1109/ACCESS.2019.2921296

[31] T. Sakamoto, R. Imasaka, H. Taki, T. Sato, M. Yoshioka, K. Inoue, T. Fukuda, H. Sakai, Feature-based correlation and topological similarity for interbeat interval estimation using ultrawideband radar, IEEE Transactions on Biomedical Engineering 63 (4) (2016) 747–757. doi:10.1109/TBME.2016.2558368

[32] M. Li, J. Lin, Wavelet-transform based data-length variation technique for fast heart rate detection using 5.8-GHz cw doppler radar, IEEE Transactions on Microwave Theory and Techniques 66 (1) (2018) 568–576. doi:10.1109/TMTT.2017.2730162

[33] C. Brüser, K. stadthanner, A. Brauers, S. Leonhardt, Applying machine learning to detect individual heartbeats in ballistocardiograms, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 1926–1929.

[34] N. Malešević, V. Petrović, M. Belić, C. Antfolk, V. Mihajlović, M. Janković, Contactless real-time heartbeat detection via 24 GHz continuous-wave doppler radar using artificial neural networks, Sensors 20 (8) (2020). doi:10.3390/s20082351
URL https://www.mdpi.com/1424-8220/20/8/2351

[35] D. Biswas, L. Everson, M. Liu, M. Fanwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Achariya, C. Van Hoof, M. Konijnenburg, N. Van Helleputte, Cornet: Deep learning framework for ppg-based heart rate estimation and biometric identification in ambulant environment, IEEE Transactions on Biomedical Circuits and Systems 13 (2) (2019) 282–291. doi:10.1109/TBCAS.2018.2892297

[36] X. Chang, G. Li, G. Xing, K. Zhu, L. Tu, Deepheart: A deep learning approach for accurate heart rate estimation from ppg signals, ACM Transactions on Sensor Networks (TOSN) 17 (2) (2021) 1–18.

[37] B. Huang, C.-L. Lin, W. Chen, C.-F. Jang, X. Wu, A novel one-stage framework for visual pulse rate estimation using deep neural networks, Biomedical Signal Processing and Control 66 (2021) 102387.

[38] K. Shi, S. Schellenberger, C. Will, T. Steigleder, F. Michler, J. Fuchs, R. Weigel, C. Ostgathe, A. Koelpin, A dataset of radar-recorded heart sounds and vital signs including synchronised reference sensor signals, Scientific data 7 (1) (2020) 1–12.

[39] P. Ramachandran, B. Zoph, Q. Le, Swish: A Self-Gated Activation Function (2017).

[40] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[41] D. B. Springer, L. Tarassenko, G. D. Clifford, Logistic regression-hsmbased heart sound segmentation, IEEE Transactions on Biomedical Engineering 63 (4) (2015) 822–832.

[42] P. Wagner, N. Strothoff, R.-D. Boussejelot, D. Kreiseler, F. I. Lunze, W. Samek, T. Schaeffer, Ptb-xl, a large publicly available electrocardiography dataset, Scientific data 7 (1) (2020) 1–15.

[43] M. S. Milivojević, A. Gavrovski, I. S. Reljin, et al., Python based physiological signal processing for vital signs monitoring (2017).

[44] D. Jarchi, A. J. Casson, Description of a database containing wrist ppg signals recorded during physical exercise with both accelerometer and gyroscope measures of motion, Data 2 (1) (2016) 1.

[45] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017).

[46] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottleneck, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510–4520.

[47] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, et al., Searching for mobilenetv3, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1314–1324.

[48] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.