Intelligent Electrocardiogram Acquisition Via Ubiquitous Photoplethysmography Monitoring

Zhangdaihong Liu [ID], Tingting Zhu [ID], Lei Lu [ID], Yuan-ting Zhang [ID], *Fellow, IEEE*, and David A. Clifton [ID]

Abstract—Recent advances in machine learning, particularly deep neural network architectures, have shown substantial promise in classifying and predicting cardiac abnormalities from electrocardiogram (ECG) data. Such data are rich in information content, typically in morphology and timing, due to the close correlation between cardiac function and the ECG. However, the ECG is usually not measured ubiquitously in a passive manner from consumer devices, and generally requires 'active' sampling whereby the user prompts a device to take an ECG measurement. Conversely, photoplethysmography (PPG) data are typically measured passively by consumer devices, and therefore available for long-period monitoring and suitable in duration for identifying transient cardiac events. However, classifying or predicting cardiac abnormalities from the PPG is very difficult, because it is a peripherally-measured signal. Hence, the use of the PPG for predictive inference is often limited to deriving physiological parameters (heart rate, breathing rate, etc.) or for obvious abnormalities in cardiac timing, such as atrial fibrillation/flutter ("palpitations"). This work aims to combine the best of both worlds: using continuously-monitored, near-ubiquitous PPG to identify periods of sufficient abnormality in the PPG such that prompting the user to take an ECG would be informative of cardiac risk. We propose a dual-convolutional-attention network (DCA-Net) to achieve this ECG-based PPG classification. With DCA-Net, we prove the plausibility of this concept on MIMIC Waveform Database with high performance level (AUROC > 0.9 and AUPRC > 0.7) and receive satisfactory result when testing the model on an independent dataset (AUROC > 0.7 and AUPRC > 0.6) which it is not perfectly-matched to the MIMIC dataset.

Index Terms—PPG, ECG, continuous monitoring, DCA-Net, neural network with attention mechanisms.

I. INTRODUCTION

ECG is a widely used medical test in predicting/classifying cardiac abnormalities, such as myocardial infarction, ventricular hypertrophy, heart failure, etc. [1], [11], [22], [28]. The 12-lead ECG is a standard approach to performing the test in primary care, which is measured by placing 10 electrodes on the skin surface of the chest and limbs to record the electrical activity of the heart. The waveform of ECG consists of a QRS complex representing ventricular depolarisation, a P wave representing atrial depolarisation, and a T wave representing ventricular repolarisation [27]. Such waveform is an informative and reliable measure to reflect the function of the heart, therefore, is extensively used in clinical practices for cardiac disease diagnosis [5], [9], [15], [21], [31]. 12-lead ECG devices are generally bulky with electrodes, a central unit, and accessories such as a monitor and keyboard. In comparison, portable ECG devices such as smartwatches and fitness trackers are smaller but are only able to measure one ECG lead. Moreover, they would require some user-initiated actions, for example, holding a sensor to close a conductive circuit such that a measurement might be taken. Therefore, ECG is difficult to acquire continuously with portable devices. Furthermore, cardiac abnormalities such as arrhythmia have the nature of being paroxysmal. It is challenging even for clinical experts to choose the timing for measuring ECG. It would be valuable to determine, from other passively-acquired, near-ubiquitous data, whether or not a patient is at sufficient "risk" (broadly defined) that it would be helpful to request that the "active" ECG measurement be taken.

PPG is an optical technique to measure the cardiac cycle by detecting blood volume changes. On the contrary to ECG, PPG is typically measured passively by pulse oximeters, and is often embedded as a function in portable devices, such as fitness wristbands and smartwatches. The waveform of PPG consists of a systolic wave and a diastolic wave, which is much simpler and smoother than the morphology of ECG. Previous research has shown that it is possible to use PPG to detect obvious abnormalities in cardiac timing [30], [37]. However, the nature of the PPG makes it very difficult to identify (let alone predict) other types of cardiac condition, particularly,

Zhangdaihong Liu and David A. Clifton are with the Department of Engineering Science, University of Oxford, OX1 3PJ Oxford, U.K., and also with the Oxford-Suzhou Centre for Advanced Research (OSCAR), University of Oxford, Suzhou 215123, China (e-mail: jessie.liu@oxford-oscar.cn; david.clifton@eng.ox.ac.uk).

Tingting Zhu and Lei Lu are with the Department of Engineering Science, University of Oxford, OX1 3PJ Oxford, U.K.

Yuan-ting Zhang is with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong SAR, China.

due to the following reasons; Firstly, PPG is a peripherally-measured signal that indirectly monitors the operation of the heart. Secondly, its relatively smooth morphology reflects less cardiac information compared with ECG. Thirdly, the waveform of PPG is prone to shape changes between different subjects, locations where the oximeter is attached, etc. [16], [18]. Hence, the PPG is mostly used to measure physiological parameters such as pulse rate, respiratory rate, blood pressure, etc. [6], [57]. However, it carried more physiological information beyond just heart/respiratory rate, and thus may be possible to construct models that identify when the PPG corresponds to intervals in which the ECG would yield positive identification of cardiac abnormality. Such a system could then duly prompt the user, in an "active learning" sense, to take an ECG measurement. It is anticipated that, in addition to being of value for consumer-based applications, such a system would also be extremely beneficial for using consumer devices in healthcare settings where existing monitoring can otherwise be either infrequent or entirely absent.

The goal of this work is to construct a PPG-based model that aims to predict when the ECG-based model would make a non-normal diagnosis. We emphasise that we are not aiming to predict specific cardiac abnormalities from the PPG, given the aforementioned difficulties in doing so-we are aiming to identify from the PPG when the existing ECG-based models would state that the user has a non-normal cardiac function.

To achieve such an objective, we propose an attention-based deep neural network model, namely a dual-convolutional-attention network (DCA-Net). DCA-Net is ResNet-backboned and contains two convolutional attention blocks, one attending the convolutional channels and one attending the temporal domain. We train DCA-Net on MIMIC Waveform Database and evaluate it on the dataset provided in the PhysioNet 2015 Challenge [8]. We further compare DCA-Net with a set of traditional and state-of-the-art machine learning models to show its superiority.

This task imposes several difficulties including the training data and models. To our knowledge, such application of PPG and ECG has never been explored. This work offers a potential solution to a significant drawback of ECG, which necessitates active acquisition. It also addresses a fundamental challenge in the context of cardiovascular diseases, which is their paroxysmal nature, making them challenging to capture through active measurements. Therefore, the contributions of this work are in three folds: 1) we are the first to undertake the proposed novel application of PPG to perform active sampling of the ECG; 2) we further propose a novel PPG classification model namely DCA-Net to alarm abnormal ECG. Inspired by applications in computer vision and 2D image classification models. DCA-Net is the first to utilize a dual-convolutional attention module to attend 1D time-series from both temporal and spatial domains; 3) our proposed model achieves the state-of-the-art performance with a high AUROC of 0.9 and AUPRC of 0.7 on the test set of MIMIC-III WDB. We further tested our model on an independent test dataset and received a satisfying AUROC of 0.7 and AUPRC of 0.6, under the condition that the independent dataset is not perfectly matched to the experimental setting.

II. RELATED WORK

Although ECG and PPG are fundamentally different physiological signal acquisition approaches (one electrical and one hemodynamic), they are closely related functionally, physiologically and morphologically [14], [25], [26], [39], [40]. Both can be used to monitor irregular heartbeats which is a symptom of many arrhythmias [14], [23]. Due to the various advantages of ECG and its wide application in clinical settings, reconstructing ECG from PPG has emerged as a popular topic in recent years [46], [55], [56]. Such a close relationship between PPG and ECG also prompts their joint applications of them for clinical tasks. However, most joint applications are based on the extracted features of the signals rather than the direct application of the waveforms because of the high dimensionality and noise level in the waveform. Reference [26] used features derived from ECG and PPG for hypertension assessment; [41] also extracted features from ECG and PPG for blood pressure estimation. For the very few works that did integrate the two signals together, physiological parameter estimation, especially respiration rate, is still the primary task [10], [34]. In the application of cardiovascular diseases, [7] was piloting on using both ECG and PPG to alert atrial fibrillation (AF), however, in an experimental setting using smart wristbands.

The vast majority of studies are single-modality based, mostly using ECG only and much less commonly using PPG [39], [57]. Among the studies focusing on PPG only, again, physiological parameters estimation is the main application direction. However, there are studies showing the strengths of PPG in indicating cardiovascular diseases using a wide range of methodologies, from traditional machine learning approaches to deep neural network models [4], [12], [33], [37], [47]. Reference [33] classified AF using extracted features from PPG and a logistic model. The sample size in this study is only 46 with 15 AF patients. [12] used several extracted features from PPG and a random forest model to classify AF/atrial flutter in a cohort of 40 patients. Reference [47] applied the neural network model (long short-term memory (LSTM) recurrent layer + convolutional layer) to two extracted features of PPG to classify AF using thousands of subjects. Reference [4] also employed a neural network approach (convolutional-recurrent neural network) but to the raw PPG signal to detect AF in 51 subjects. Most relevantly, [42] proposed an attention-based recurrent neural network (RNN) to detect AF. The model was first trained on ECG data and then transferred to PPG data to improve the performance. All the above works presented high performance/accuracy in detecting AF which shows that PPG is indicative of some cardiac dysfunction such as AF. However, in these works, ECG was needed for confirmation/pre-training. Moreover, we found no research using large-scale PPG data to target general arrhythmia.

Among the recent PPG classification works, mostly employed an architecture of convolutional neural network (CNN), RNN or a combination of the two. Reference [38], [45] applied 1D CNN to the PPG segment (an interval of the waveform) classification. References [3], [24], [38], [42] all applied a hybrid version of CNN and RNN to classify PPG segments, among which [42] also added attention. Since the advances of attention mechanisms,
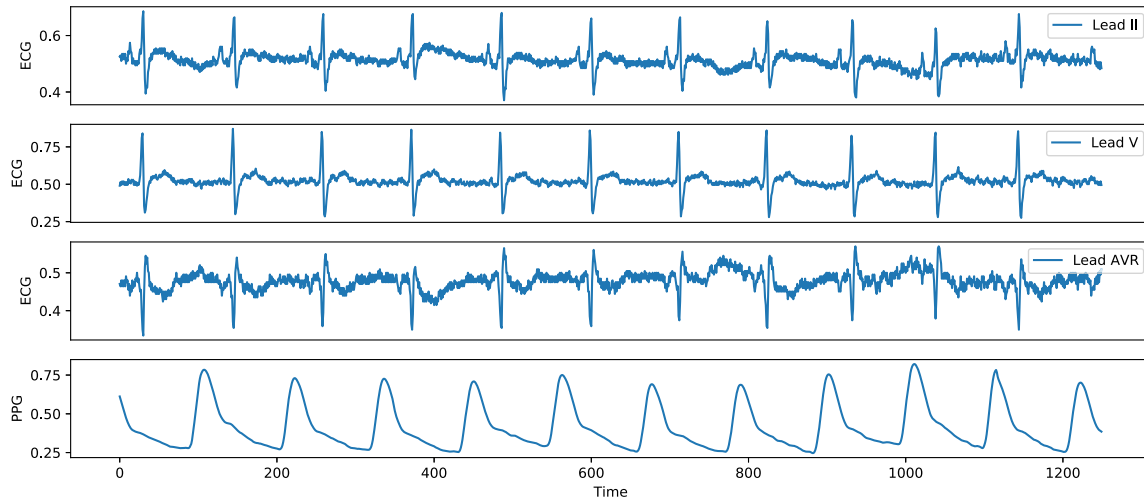
Fig. 1. Example of the training data in MIMIC-III WDB. The top three panels are the three ECG leads and the bottom panel is the matched PPG signal. All signals are 10 s long.

especially its successes in computer vision (CV) and natural language processing (NLP), studies with novel attention architectures bloomed in these areas (CV and NLP) [19], [50], [51], [52]. SENet (Squeeze-and-Excitation Networks [19]) turned a multi-layer perceptron (MLP) into an attention module to attend the convolutional channels. This addition of attention to the CNN improved the image classification performance. CBAM (Convolutional Block Attention Module [52]) improved SENet by adding a convolutional spatial attention module and showed superior performance in image classification as well. The more recent work, ECA-Net (Efficient Channel Attention Networks [51]) further improved the efficiency of the channel attention by turning the MLP into a convolutional-based attention module but not implementing any spatial attention. ECA-Net showed that this modification outperforms both SENet and CBAM. The adaption and transfer from 2D image to 1D time-series are intuitive and straightforward. Moreover, with the success of attention in CV, its adaption to the 1D physiological signal started to rise in recent two years. However, the applications are still limited to quality assessment and blood pressure monitoring [2], [53]. Little research was found on the application of cardiac disease-related tasks.

III. DATA

We used MIMIC-III Waveform Database (WDB) Matched Subset [20], [32] as our training dataset. We chose the Matched Subset due to its linkage to the clinical records – all patients in the Matched Subset are identified and matched to the medical record system. MIMIC-III WDB includes multi-lead ECG signals and their paired PPG signals. The Matched Subset contains waveform records for 10,282 unique ICU patients. For each patient, different lengths of the waveform data were sampled at different time points during their hospital stays. Since different ECG records have different numbers and kinds of leads, we selected the 3 most common and representative ECG leads for analysis, leads II, V and AVR. Therefore, records with these three leads of ECG and PPG signals were included in the analysis. Notably,

these signals were sampled at 125 Hz. To reduce the noise level, we only considered records that are longer than two minutes of the signals. Considering the applicable usage in wearable devices, we set the length of the analysing signal to be 10 seconds and took the second 10 seconds out of the whole signal to avoid the initial unsteadiness. An example of the raw signal segments is shown in Fig. 1.

A. PPG Pre-Processing

Since PPG signals are generally noisy, we carefully designed the following processing steps:

1) *Raw signal flatness detection:* if the value of 60 consecutive time points (about 0.5 seconds) does not change over a threshold (1e-5), we treat it as a flat signal and remove the whole segment. This removed about 7% of the signals.
2) *Normalisation:* normalise each signal to zero-mean and unit-variance.
3) *Filtering:* 3 rd order band-pass Butterworth filter was applied with the low band cut being 0.5 Hz and the high band cut being 8 Hz, a method that was proposed in [13].
4) *Peak detection:* we applied the Python toolbox 'HeartpPy' [48], [49] to detect valid peaks of the signal. If the 10 s signal has less than 5 peaks (corresponding to lower than 30 bpm), we removed the signal from the study. This removed another 7% of PPG signals from the dataset.
5) *Skewness SQI:* skewness was calculated in a sliding window fashion. It is calculated for every 250 samples (2 seconds) with a stride being 125 samples (1 s). If the majority (over 50%) of the calculated skewness is negative, we consider it as a poor-quality signal. This further removed 14% signals.
6) *Outlier sample replacement:* lastly, we applied the Hampel filter to detect outliers. For every 10 samples, MAD (median absolute deviation) is calculated. Then segment standard deviation is estimated in terms of the MAD value assuming a normal distribution. Outliers are detected if the sample value is 3 std away from the segment median,
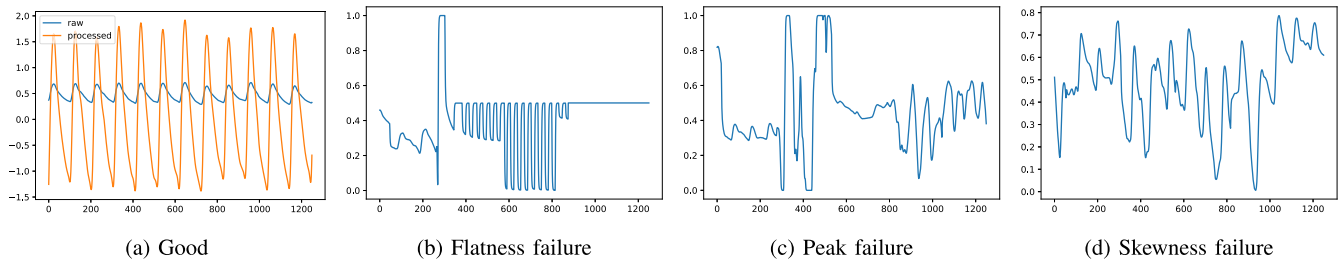
Fig. 2.   Examples of PPG segments with different qualities. (a) Shows a segment that passed the quality control with the raw signal shown in blue and the processed signal shown in orange. (b)–(d) are segments that failed due to different quality control criteria listed in the pre-processing steps.
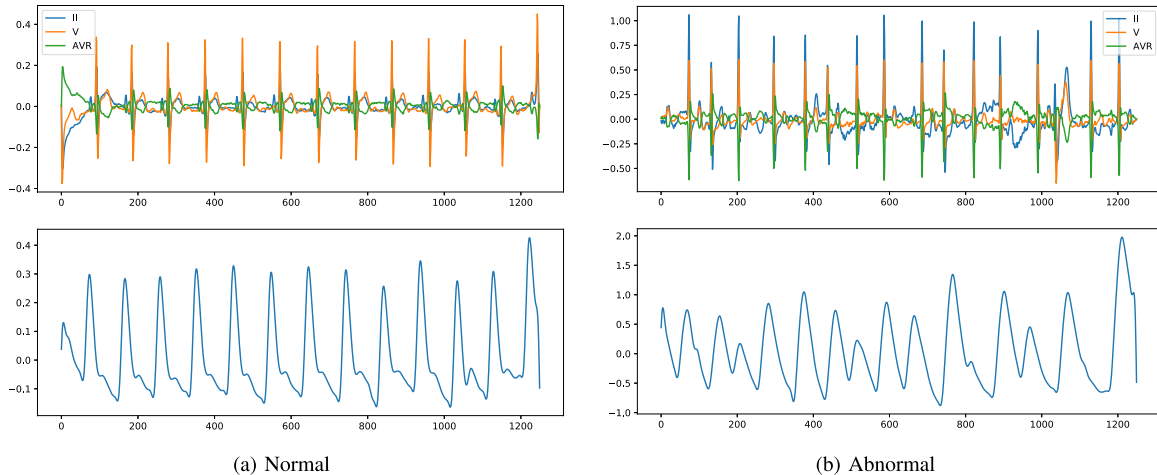

Fig. 3.   Examples of normal (a) and abnormal (b) ECG signals labelled by AutoNet and their paired PPG signal. Signals are shown after pre-processing.

and the outliers are replaced by the segment median. This mainly replaced the start and end of the signals.

Step 1) can be considered as a quality screening step. Steps 2) and 3) are cleaning and denoising and Steps 4)-6) are quality controlling. After processing, we had 38,320 10s-long paired ECG and PPG training segments in total. A snapshot of the good and poor PPG segments is shown in Fig. 2.

Since there are no ECG labels given in the MIMIC-III WDB, we used a highly reliable in-house ECG classification model, AutoNet to label the ECG [43], [44]. For each ECG lead, we applied the 5th order high-pass Butterworth filter at 0.5 Hz, followed by powerline filtering with the powerline frequency being 50 Hz. Then we assessed the signal quality using the method proposed in [54]. This assessment outputs one of the three categories of quality – unacceptable, barely acceptable or excellent. If any of the ECG leads has 'excellent' quality, we accept the segment. All ECG pre-processing was implemented using the Python package NeuroKit2 [29].

Fig. 3 gives an example of the normal and abnormal ECG labelled by AutoNet and their paired PPG.

convolutional neural networks with attention in the computer vision area [19], [51], [52], we employed a 1D dual-convolutional-attention (DCA) module added to the ResNet-34 backbone to classify the 1D PPG. Notably, we replaced the 2D convolutional layers in the original ResNet with 1D convolutional layers to suit the 1D PPG signal classification. The model architecture of 1D DCA-Net is illustrated in Fig. 4. We added a DCA module right before adding the residual in a normal ResNet block to form the new ResNet+DCA block. The rest of the model architecture stays the same with ResNet-34.

We reviewed several attention mechanisms including the aforementioned SENet, CBAM and ECA-Net which all adopted a single MLP/convolutional attention module or a combination of them to learn the attention weights. We further conducted empirical experiments to analyse the efficacy of different attention blocks and designed the DCA module trying to maximise the advantage of the attention block to fit into the 1D scenario. The input of DCA-Net is the 10-second long processed 1D PPG signal and the output of the model is a binary label indicating whether this segment of the signal is considered to be abnormal by the model.

IV. Dual-Convolutional Attention Network

The backbone of the Dual-Convolutional Attention Network (DCA-Net) is a standard ResNet [17]. Inspired by the deep

A. The Dual-Convolutional Attention Module

DCA-Net follows the overall architecture of ResNet. The difference exists in each of the ResNet blocks. The DCA module
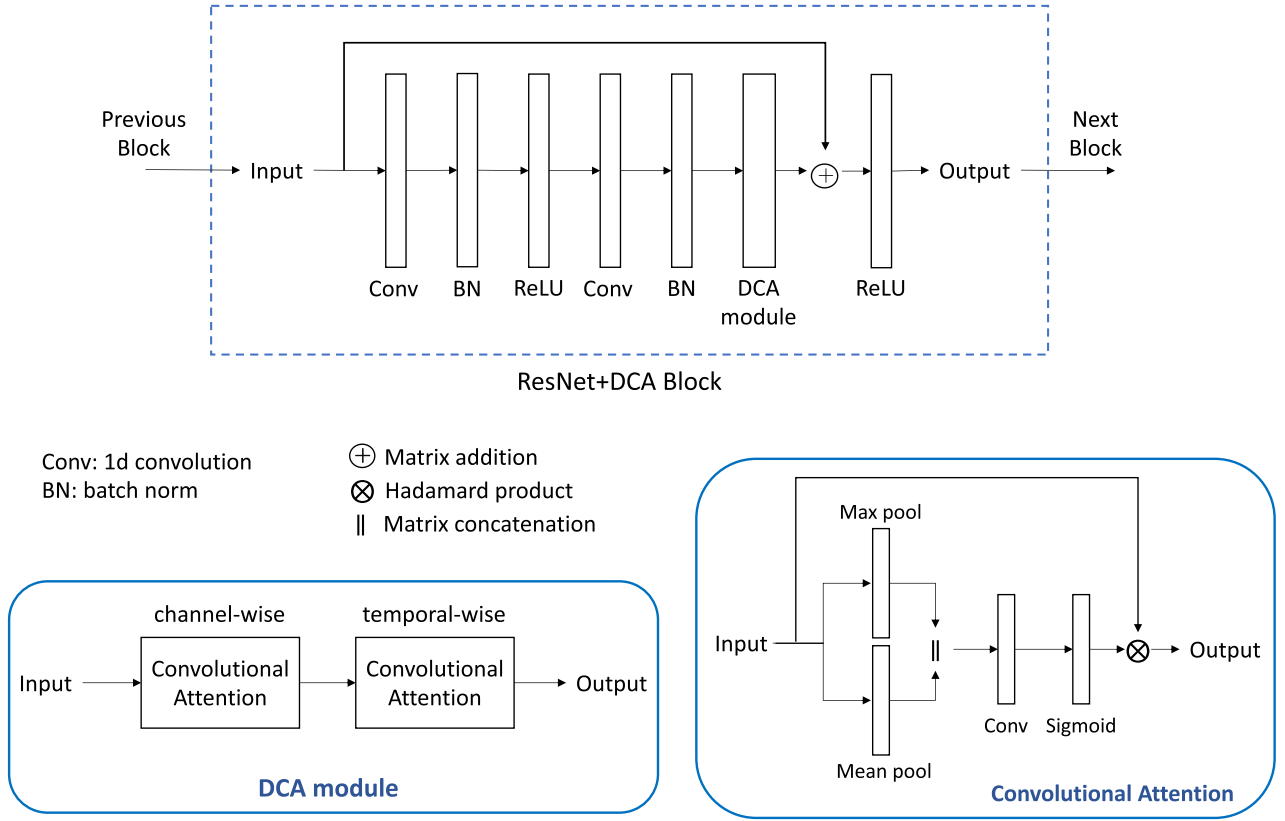
Fig. 4. DCA-Net overview. The top figure illustrates the ResNet+DCA module block which replaces the normal ResNet block in the backbone model. The convolutional attention convolves over the channel dimension for the channel-wise attention and over the temporal dimension for the temporal-wise attention.

consists of two convolutional attention blocks arranged in sequential order as shown in the 'DCA module' box in Fig. 4. The convolutional attention relies on applying the convolution operation to a certain dimension of the data to learn the interactions between data points along that dimension. Compared with the MLP attention module, convolutional attention is computationally more efficient since it requires no parameters in a convolutional layer. Moreover, [51] suggested that avoiding dimensionality reduction, i.e. using fully-connected layers in a MLP style is helpful towards effective channel attention learning.

The first convolutional attention is applied to the convolution channels and the second is applied in the temporal domain. Both the channel-wise and temporal-wise attentions adopt the architecture illustrated in the 'Convolutional Attention' box in Fig. 4 with the pooling and convolutional layers applied to the corresponding directions. We used the same channel attention introduced in [52] and adapted it to 1D scenario. Specifically, let $X \in \mathbb{R}^{N \times C \times D}$ be the input of the channel-wise convolutional attention (also the input of the DCA module), where N is the sample/batch size, C the number of convolution channels output from the original ResNet layers and D is the length of the signal (denoted as the temporal dimension). The output of the channel-wise attention $\widetilde{X}$ is

$$\widetilde{X} = X \otimes \sigma(\text{conv}_C^{(7)}(g(X))), \tag{1}$$

$$\text{where} \quad g(X) = [p_1(X) \parallel p_2(X)]_D, \tag{2}$$

$$\text{where} \quad p_1(X) = \text{max\_pool}_D(X), \tag{3}$$

$$\text{and} \quad p_2(X) = \text{mean\_pool}_D(X). \tag{4}$$

In 1, $\otimes$ indicates the Hadamard product (element-wise matrix multiplication); $\sigma(\cdot)$ represents the Sigmoid activation function and $\text{conv}_C^{(7)}$ denotes the 1D convolutional layer with kernel size 7 and convolves over the channel direction. The $[\cdot \parallel \cdot]_D$ in (2) denotes the matrix concatenation along the temporal direction, and the subscript D in (3) and (4) show that the pooling layers are applied in the temporal dimension as well. The function $\sigma(\cdot) \in \mathbb{R}^{N \times C \times 1}$ outputs the channel-wise attention weights which are further expanded to the same shape of X to scale X. The scaled X, denoted as $\widetilde{X}$, is then served as input to the temporal-wise attention.

Similarly, let the output of the temporal-wise attention be $\widetilde{\widetilde{X}}$.

$$\widetilde{\widetilde{X}} = \widetilde{X} \otimes \sigma(\text{conv}_D^{(7)}(g(\widetilde{X}))), \tag{5}$$

$$\text{where} \quad g(\widetilde{X}) = [p_1(\widetilde{X}) \parallel p_2(\widetilde{X})]_C, \tag{6}$$

$$\text{where} \quad p_1(\widetilde{X}) = \text{max\_pool}_C(\widetilde{X}), \tag{7}$$

$$\text{and} \quad p_2(\widetilde{X}) = \text{mean\_pool}_C(\widetilde{X}). \tag{8}$$

Equations (7), (8) indicate that we first operate max pooling and mean pooling in the channel dimension, therefore, both $p_1(\widetilde{X})$,

$p_2(\widetilde{X}) \in \mathbb{R}^{N \times 1 \times D}$. Then the pooled matrices are then concatenated along the channel dimension and $g(\widetilde{X}) \in \mathbb{R}^{N \times 2 \times D}$ (6). The 1D convolutional layer convolves over the temporal dimension and outputs a single channel such that $\text{conv}_D^{(7)}(g(\widetilde{X})) \in \mathbb{R}^{N \times 1 \times D}$. Finally, this convolved matrix is activated by Sigmoid and used to scale the temporal domain of the input $\widetilde{X}$.

We set the kernel size in the convolutional attention block to 7 to allow a wide coverage of the interactive learning. This choice was also validated by heuristic experiments using kernel sizes of 3, 5 and 7. Furthermore, the concatenation of the two pooling layers was tested against using either pooling layer as well as applying a convolutional layer to each pooling layer and concatenating them afterwards. These experiments were inferior to the current setting. There is published work showing learning channel-wise attention first gives slightly better performance than the parallel or channel-last order [52]. Therefore, by linking the two convolutional attention blocks with channel-first order, the DCA module first learns the inter-channel interaction and then learns the temporal-wise interaction based on the channel-attended output.

B. Model Assessment and Implementation Details

We applied DCA-Net to the MIMIC-III WDB processed PPG using ECG labels for binary classification. The 38320 PPG segments are split into training, validation and test sets with a ratio of 64%:16%:20%. The model parameters are learnt using the training set and validated on the validation set to avoid over-fitting. The test set is only used to report the results. We evaluated the model performance using the specificity, sensitivity/recall, precision, accuracy, AUROC (area under the receiver operating characteristic curve) and AUPRC (area under the precision-recall curve). We regard AUPRC more valuable than AUROC in our case since it is more sensitive towards imbalanced data. We further ran the model for 10 random repetitions and took the mean and standard deviation of the above assessment measures.

The DCA-Net was implemented using PyTorch [35] version 1.11.0 with CUDA version 11.7. The loss function is binary cross entropy. We further experimented with adding loss weights to adjust the class imbalance. The model was optimised by Adam optimiser with the default hyper-parameter setting and trained using a batch size of 64. We applied a step-wise learning rate scheduler to reduce the learning rate to 0.1 times every 10 epochs. Early-stopping was also applied – if the validation loss does not reduce compared with the current lowest loss for 5 consecutive epochs, the training terminates. We always save the model with the lowest loss and apply it to the test set for assessment.

C. Computation Complexity

The number of trainable parameters in our proposed DCA-Net is 7.23 millions, which is the same with its backbone 1D ResNet-34 and the state-of-the-art model ECA-Net. However, ResNet-34 does not contain any attention mechanism and ECA-Net in 1D setting only attends one dimension of the time-series (rather than both temporal and spatial dimensions). The convolutional

attention block does not add any number of trainable parameters to its backbone model, therefore, achieves higher computational efficiency compared with the MLP attention mechanism such as employed by CBAM and SENet. The number of trainable parameters with MLP attention block in an equivalent architecture would be 7.39 millions.

D. Baseline Models

To set the baseline performance level, we considered a set of non-neural network models including Logistic Regression (LR), Random Forest Classifier (RFC), Support Vector Classifier (SMV) and XGBoost Classifier (XGB). To choose the backbone model, we evaluated ResNet-18, ResNet-34 and ResNet-50, among which ResNet-34 gave the best performance and therefore, was selected as the backbone model for DCA-Net. To further compare the performance of our proposed DCA-Net, we implemented the current state-of-the-art model ECA-Net which was shown to have outperformed other related CNN with attentions. We adapted ECA-Net to 1D space and use the same backbone model with our proposed DCA-Net. The non-neural network model was implemented using scikit-learn version 1.0.2 [36] with a training ratio of 80% and test ratio of 20%. We applied class weights to adjust to the balanced classes. For the neural network baselines, they were trained exactly the same with DCA-Net as stated in Section IV-B. Notably, the test sets remain the same for all models implemented.

V. EVALUATION ON PHYSIONET 2015 CHALLENGE DATASET

There are very limited public waveform datasets that have paired ECG and PPG signals. The 2015 PhysioNet Challenge [8] provided one such dataset for the task of Reducing False Arrhythmia Alarms in the ICU, where the ECG signals were labelled. The PhysioNet dataset consists of 750 waveform records for training and each recording contains PPG and two ECG leads. All ECG in the dataset triggered bedside alarms in ICUs and they were taken from roughly 5 minutes before the onset of the alarm to the onset of the alarm. Therefore, all records meet the 2 minutes long requirement applied in the training dataset. These recordings were further examined by a team of expert annotators to identify the false alarms. We regarded the false alarms as normal signals and true alarms as abnormal ones.

The PPG signals went through the same processing pipeline as the training dataset. However, these signals were resampled to 250 Hz, rather than 125 Hz which was used in the MIMIC-III WDB. We added a resampling step to the pre-processing after the signal cleaning and before quality control (between Steps 3 and 4) in the PPG pre-processing pipeline). We applied the same quality assessment to remove the ECG with poor qualities and maximise the pre-processing consistency with MIMIC WDB. This left us with 108 signals for validation. Table I summarises the datasets used in this study.

VI. RESULTS

A. PPG Classification Using MIMIC

The best benchmark AUROC we obtained from the non-neural network baseline models is around 0.88 and AUPRC

TABLE I
DATASET OVERVIEW

Dataset	Sample size	Sampling frequency	Signals	Train/Validation/Test	Positive labels
MIMIC-III WDB	38320	125 Hz	PPG and ECG leads II, V and AVR	64% / 16% / 20%	17.66%
PhysioNet 2015	108	250 Hz	PPG and ECG leads II and V	0 / 0 / 100%	38.89%

TABLE II
NON-NEURAL NETWORK BASELINE MODELS

Model	Specificity (%)	Sensitivity/Recall (%)	Precision (%)	Accuracy (%)	AUROC (%)	AUPRC (%)
LR	80.40	22.91	20.03	70.25	51.56	28.28
RFC	97.99	27.72	74.70	85.58	**88.40**	58.50
SVC	87.75	70.43	55.21	84.69	87.52	**65.43**
XGB	95.23	45.08	66.96	**86.38**	88.07	60.87

The numbers under each column are percentages. The highest Accuracy, AUROC and AUPRC are highlighted by bold font. All assessments are point estimations.

TABLE III
RESULTS OF THE NEURAL NETWORK MODELS

Model	Specificity (%)	Sensitivity/Recall (%)	Precision (%)	Accuracy (%)	AUROC (%)	AUPRC (%)
ResNet-18	93.74 (0.34)	64.23 (2.73)	68.77 (0.81)	88.53 (0.34)	91.47 (0.46)	67.78 (1.44)
ResNet-34	93.17 (0.39)	67.12 (1.78)	67.85 (0.77)	88.57 (0.15)	91.59 (0.16)	67.89 (0.88)
ResNet-50	93.42 (0.35)	62.21 (2.03)	66.97 (0.73)	87.91 (0.20)	90.55 (0.26)	65.95 (1.04)
ECA-Net	92.38 (1.97)	71.09 (4.91)	67.10 (3.26)	88.62 (0.81)	91.85 (1.71)	69.04 (1.14)
DCA-Net (ours)	93.34 (0.46)	67.92 (1.45)	68.67 (1.10)	**88.86 (0.20)**	**91.88 (0.24)**	**70.12 (0.78)**

The numbers shown are the average score in percentage for 10 random repetitions with their standard deviation shown in brackets. The best average accuracy, AUROC and AUPRC are highlighted with bold font. DCA-Net is significantly better than the second best model based on AUPRC (p-value = 0.001).

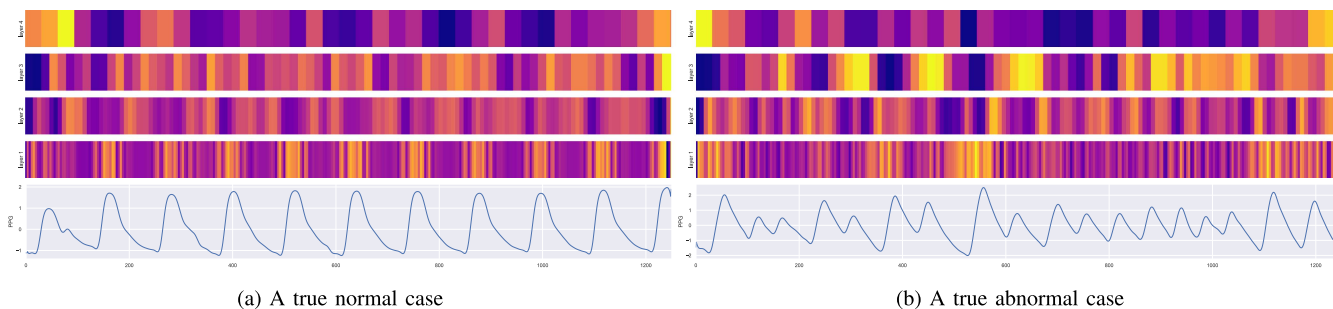(a) A true normal case   (b) A true abnormal case

Fig. 5. Examples of normal (a) and abnormal (b) PPG segments correctly classified by DCA-Net and their temporal attention weights in layers 1-4 (heat maps from the bottom up). The brighter colours indicate heavier weights.

is around 0.65 (Table II). The highest AUPRC, AUROC and accuracy scores are all from different models. None of the traditional machine learning models in Table II offered consistent performance across the assessment measures we considered. Although SVC has the highest AUPRC, it took over 10 times longer than the second-best (regarding the AUPRC score) model XGB. Apart from SVC, all non-neural network models performed poorly on the abnormal class with low sensitivity/recall. In particular, the worse performance is from the logistic regression classifier (LR) which may suggest the data contain non-linear relationships with the label.

For the backbone neural network models we tested (ResNet-18, 34 and 50), there is a clear performance drop from ResNet-34 to ResNet-50 (Table III). This may be a sign of the model being over-parameterised. Although there is no significant improvement from ResNet-18 to 34, due to the better performance of ResNet-34 on the abnormal class which can be seen from the sensitivity/recall scores in Table III, we selected ResNet-34 as our backbone model to add attention modules. The ECA-Net and our proposed DCA-Net were all based on the ResNet-34 backbone.

Comparing the bottom block with the top block in Table III, we observed significant improvement by adding the attention modules to the backbone model. Since AUROC is less sensitive to imbalanced datasets, we only provided it for reference and measured the performance mainly based on AUPRC.

Our proposed DCA-Net has the best average accuracy, AUROC and AUPRC scores. The AUPRC of DCA-Net is significantly better than the ECA-Net with a p-value=0.01 in a one-side T-test. Moreover, DCA-Net performs more stably than ECA-Net due to the smaller standard deviations across all assessment measures.

All results shown in Tables II and III are from the test set of MIMIC-II WDB which consists of 7664 segments, 20% of the whole dataset.

1) Temporal Attention Interpretation: We extracted the temporal attention from each layer in the DCA-Net. Fig. 5 shows an example of correctly classified normal and abnormal PPG segments and their temporal attention weights in each ResNet layer. It appears that the first layer focuses on the peaks of the waveform – the normal case Fig. 5(a) has its attention evenly distributed on the peaks while the attention for the abnormal

Model		Specificity (%)	Sensitivity/Recall (%)	Precision (%)	Accuracy (%)	AUROC (%)	AUPRC (%)
DCA-Net	w/ c/w	56.97 (2.46)	77.86 (3.70)	53.53 (1.37)	65.09 (1.55)	71.31 (1.83)	57.16 (3.10)
	w/o c/w	78.64 (4.03)	46.67 (3.72)	58.47 (4.00)	66.20 (2.20)	71.72 (1.65)	60.13 (3.46)

'w/ c/w' and 'w/o c/w' indicate during training, the loss function was with and without class weights adjustment, respectively. The numbers shown are the average score in percentage for 10 random repetitions with their standard deviation shown in brackets.

case (Fig. 5(b)) highlights the three larger peaks located in the beginning, the middle and the end of the signal. Layers 2 and 3 are attending the non-peak areas or smaller peaks. The deepest layer, layer 4 appears to be always focusing on the beginning and the end of the signals.

For the normal signal, the attention heat maps present more regular patterns than the abnormal heat maps, and the attention values are more evenly distributed, having smaller variances – the abnormal heat maps have brighter colour patches. This indicates the model pays more attention to certain areas in the abnormal case than in the normal case.

B. PhysioNet 2015

Finally, we tested the DCA-Net on an external dataset, PhysioNet 2015 challenge dataset. Due to the small sample size, the model is not re-trained or fine-tuned. It is a direct application of the model that was trained on MIMIC-III WDB. The results are shown in Table IV.

Although the performance level dropped on this external dataset which is not surprising due to their differences in data acquisition, sampling frequency, labelling mechanisms and class ratios, we still observed fair performance of the model, especially for the class weight adjusted setting – the recall on the minority class is 77.86. It suggests the model is still able to identify most of the abnormal signals.

VII. DISCUSSIONS AND CONCLUSIONS

In this article, we proposed a novel task – intelligent ECG acquisition via monitoring the PPG signals. This is the first attempt in applying ECG labels to guide the training of PPG where the alarms raised by PPG can be used to alert abnormal ECG so that informative signals can be collected at more 'favourable' times (when a cardiovascular abnormality happens). We used the large-scale public dataset MIMIC-III WDB which has rich paired ECG and PPG data to experiment with this proof-of-concept.

We further proposed a PPG classification model, dual-convolutional-attention network (DCA-Net), with a novel attention module added to the backbone model ResNet and compared it with several traditional non-neural network machine learning models and neural network models. DCA-Net achieved the highest AUPRC score among all models considered. Due to the employment of dual convolutional attention, not only we increased the performance level, but also reduced the number of parameters overhead compared with the previous state-of-the-art models.

From the analysis results, we can see a clear improvement path from the non-neural network models to neural network models and to neural network models with attention, which indicates the

merit of deep neural network models and attention mechanisms. Moreover, with the help of attention, we are able to visualise the learning of deep neural networks. This application offers another step away from the conventional 'black-box' impression of the neural network models.

Although the AUROC of our models reached over 90%, the AUPRC is less competitive. The AUPRC is pulled down by the performance of the abnormal class which is a substantially smaller class compared with the normal class. We can rationalise this performance reduction by several factors. First of all, the dataset is very imbalanced. We applied class weights adjustment to the loss function which in turn boosted the recall level of the abnormal class. However, this sacrificed the precision level and overall AUPRC. The second reason which also affects the overall performance is the noise in the ECG labels. Ideally, we would have ECG labelled by cardiac experts. Given this is not the case in MIMIC WDB, we labelled ECG with a published model that has high accuracy and trained on millions of labelled ECG data. However, this would still introduce label biases due to the differences between the dataset used to train the ECG model and MIMIC WDB. Thirdly, PPG data are highly noisy. Compared with ECG, the waveform structure of PPG is much simpler and prone to noise. Although we spent a considerable amount of effort in designing the PPG pre-processing pipeline, we still noticed some noisy signals slipped through. Adopting more sophisticated PPG processing can be a valuable future step.

In this study, we used signals each was 10 seconds long to suit easier real-world applications. Moreover, since we are using the actual waveform rather than derived features as model input, our model should be able to capture the information well within 10 seconds long signals. Classification using longer signals might be easier, but it imposes fewer constraints to signal acquisition. Depending on the actual use case, tailoring the signal length can be beneficial and is worth exploring in future work.

We attempted the classification of PPG using the 2D power spectrum as model input and received poor performance. By eye-balling the spectrum, we observed more noises than distinctive patterns which might also be caused by the short length of signal adapted in this study.

Another main challenge in this task is model validation. One challenge lies in the dataset. It is difficult to find PPG with abnormality labels obtained from ECG or PPG with paired ECG that has labels. The dataset provided by the PhysioNet 2015 Challenge is the most similar to our training dataset which also has labels. The DCA-Net showed fair performance on this dataset, however, there is still a significant performance drop. The disparities between the two datasets including the sampling frequency, class ratio and the available ECG leads can be accounted for this performance drop. Most importantly, the labels provided by Physionet 2015 do not align perfectly with

our task as they are true or false ICU alarms for arrhythmia. We considered the false alarm as normal in ECG for our model validation despite that might not be the case. Moreover, the data were collected at different geospatial locations and by different machines which could lead to inherent differences in data distributions and noise levels.

We tried several transfer learning techniques on the PhysioNet dataset including fine-tuning the whole model and fine-tuning the classification layer only, but we did not achieve improvement due to the small sample size. One effective way to strengthen the model power is to have a larger and more reliable validation dataset with matched characteristics to the training data. Having some clinical experts-labelled data although would be labour-consuming, would benefit greatly from the model training which can be part of future work. Furthermore, applying data augmentation techniques to reliably labelled data to enlarge the data size can be an efficient way to improve the statistical power of the model.

Last but not least, this work is a proof-of-concept exploration of a novel application of PPG, using it to alarm abnormal ECG. We do not specify the type of cardiac abnormality. The goal is to utilize passively-measured PPG to initiate an ECG acquisition during periods of elevated cardiac risk. A valuable next step would be extending this application to specific cardiac diseases. Nevertheless, this task is expected to be more demanding due to the even greater imbalance in class ratios.

VII. DATASETS

The two datasets used in this work are both publicly available datasets. They can be downloaded with the following links (under authorised access):

1. MIMIC-III Waveform Database [32]:
https://physionet.org/content/mimic3wdb-matched/1.0/
2. PhysioNet 2015 Challenge dataset [8]:
https://physionet.org/content/challenge-2015/1.0.0/

REFERENCES

[1] U. R. Acharya et al., "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals," *Appl. Intell.*, vol. 49, no. 1, pp. 16–27, 2019.

[2] N. Aguirre, E. Grall-Maës, L. J. Cymberknop, and R. L. Armentano, "Blood pressure morphology assessment from photoplethysmogram and demographic information using deep learning with attention mechanism," *Sensors*, vol. 21, no. 6, 2021, Art. no. 2167.

[3] A. Aliamiri and Y. Shen, "Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inf.*, 2018, pp. 442–445.

[4] K. Aschbacher et al., "Atrial fibrillation detection from raw photoplethysmography waveforms: A deep learning application," *Heart rhythm O2*, vol. 1, no. 1 pp. 3–9, 2020.

[5] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," *Biomed. Signal Process. Control*, vol. 43, pp. 216–235, 2018.

[6] P. H. Charlton et al., "Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 2–20, 2018.

[7] E. Chen et al., "A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation," *Heart rhythm*, vol. 17, no. 5, pp. 847–853, 2020.

[8] G. D. Clifford et al., "The physionet/computing in cardiology challenge 2015: Reducing false arrhythmia alarms in the ICU," in *Proc. Comput. Cardiol. Conf.*, 2015, pp. 273–276.

[9] S. Dilmac and M. Korurek, "ECG heart beat classification method based on modified ABC algorithm," *Appl. Soft. Comput.*, vol. 36, pp. 641–655, 2015.

[10] X.-R. Ding, Y.-T. Zhang, H. K. Tsang, and W. Karlen, "A pulse transit time based fusion method for the noninvasive and continuous monitoring of respiratory rate," in *Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2016, pp. 4240–4243.

[11] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," *Expert Syst. Applications: X*, vol. 7, 2020, Art. no. 100033.

[12] L. M. Eerikäinen et al., "Detecting atrial fibrillation and atrial flutter in daily life using photoplethysmography data," *IEEE J. Biomed. Health Inf.*, vol. 24, no. 6, pp. 1610–1618, Jun. 2020.

[13] M. Elgendi, "Optimal signal quality index for photoplethysmogram signals," *Bioengineering*, vol. 3, no. 4, 2016, Art. no. 21.

[14] C. Fischer, B. Dömer, T. Wibmer, and T. Penzel, "An algorithm for real-time pulse waveform segmentation and artifact detection in photoplethysmograms," *IEEE J. Biomed. Health Inf.*, vol. 21, no. 2, pp. 372–381, Mar. 2017.

[15] K. G. Harmon, M. Zigman, and J. A. Drezner, "The effectiveness of screening history, physical exam, and ecg to detect potentially lethal cardiac disorders in athletes: A systematic review/meta-analysis," *J. Electrocardiol.*, vol. 48, no. 3, pp. 329–338, 2015.

[16] V. Hartmann, H. Liu, F. Chen, Q. Qiu, S. Hughes, and D. Zheng, "Quantitative comparison of photoplethysmographic waveform characteristics: Effect of measurement site," *Front. Physiol.*, vol. 10, 2019, Art. no. 198.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778 .

[18] M. Hickey, J. P. Phillips, and P. A. Kyriacou, "Investigation of peripheral photoplethysmographic morphology changes induced during a hand-elevation study," *J. Clin. Monit. Comput.*, vol. 30, no. 5, pp. 727–736, 2016.

[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141 .

[20] A. E. Johnson et al., "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, 2016.

[21] D. Kasper, A. Fauci, S. Hauser, D. Longo, J. Jameson, and J. Loscalzo, *Harrison's Principles of Internal Medicine, 19e*, vol. 1. New York, NY, USA: Mcgraw-hill, 2015.

[22] D. Kiyasseh, T. Zhu, and D. A. Clifton, "CLOCS: Contrastive learning of cardiac signals across space, time, and patients," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 5606–5615.

[23] A. G. Kléber and Y. Rudy, "Basic mechanisms of cardiac impulse propagation and associated arrhythmias," *Physiol. Rev.*, vol. 84, no. 2, pp. 431–488, 2004.

[24] S. Kwon et al., "Deep learning approaches to detect atrial fibrillation using photoplethysmographic signals: Algorithms development study," *JMIR mHealth uHealth*, vol. 7, no. 6, 2019, Art. no. e12770.

[25] Q. Li et al., "Transfer learning from ECG to PPG for improved sleep staging from wrist-worn wearables," *Physiol. Meas.*, vol. 42, no. 4, 2021, Art. no. 044004.

[26] Y. Liang, Z. Chen, R. Ward, and M. Elgendi, "Hypertension assessment via ECG and PPG signals: An evaluation using MIMIC database," *Diagnostics*, vol. 8, no. 3, 2018, Art. no. 65.

[27] L. S. Lilly, *Pathophysiology of Heart Disease: A Collaborative Project of Medical Students and Faculty*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2012.

[28] W. Liu et al., "Real-time multilead convolutional neural network for myocardial infarction detection," *IEEE J. Biomed. Health Inf.*, vol. 22, no. 5, pp. 1434–1444, Sep. 2018.

[29] D. Makowski et al., "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behav. Res. Methods*, vol. 53, no. 4, pp. 1689–1696, Feb. 2021.

[30] C. A. Millán, N. A. Girón, and D. M. Lopez, "Analysis of relevant features from photoplethysmographic signals for atrial fibrillation classification," *Int. J. Environ. Res. Public Health*, vol. 17, no. 2, 2020, Art. no. 498.

[31] B. Montague, J. Ouellette, and G. Buller, "Retrospective review of the frequency of ECG changes in hyperkalemia," *Clin. J. Amer. Soc. Nephrol.*, vol. 3, no. 2, pp. 324–330, 2008.

[32] A. E. W. Johnson et al., "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, 2016, Art. no. 160035, doi: 10.1038/sdata.2016.35.

[33] S. Nemati et al., "Monitoring and detecting atrial fibrillation using wearable technology," in *Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2016, pp. 3394–3397.

[34] C. Orphanidou, "Derivation of respiration rate from ambulatory ECG and PPG using ensemble empirical mode decomposition: Comparison and fusion," *Comput. Biol. Med.*, vol. 81, pp. 45–54, 2017.

[35] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.

[36] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[37] T. Pereira et al., "Photoplethysmography based atrial fibrillation detection: A review," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–12, 2020.

[38] M.-Z. Poh et al., "Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms," *Heart*, vol. 104, no. 23, pp. 1921–1928, 2018.

[39] H. Sardana et al., "Arrhythmia detection and classification using ECG and PPG techniques: A review," *Phys. Eng. Sci. Med.*, vol. 44, no. 4, pp. 1027–1048, 2021.

[40] M. Shabaan et al., "Survey: Smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis," *BMC Med. Inf. Decis. Mak.*, vol. 20, no. 1, pp. 1–16, 2020.

[41] I. Sharifi, S. Goudarzi, and M. B. Khodabakhshi, "A novel dynamical approach in continuous cuffless blood pressure estimation based on ECG and PPG signals," *Artif. Intell. Med.*, vol. 97, pp. 143–151, 2019.

[42] S. P. Shashikumar, A. J. Shah, G. D. Clifford, and S. Nemati, "Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 715–723 .

[43] Y. Shen, "ECG classification and the "heart age" prediction using machine learning," Ph.D. thesis, University of Oxford, Oxford, U.K., 2020.

[44] Y. Shen, R. Clarke, T. Zhu, and D. A. Clifton, "Generating neural network models, classifying physiological data, and classifying patients into clinical classifications," Worldwide patent WO2022189771A1, Sep. 2022.

[45] Y. Shen, M. Voisin, A. Aliamiri, A. Avati, A. Hannun, and A. Ng, "Ambulatory atrial fibrillation monitoring using wearable photoplethysmography with deep learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 1909–1916.

[46] X. Tian, Q. Zhu, Y. Li, and M. Wu, "Cross-domain joint dictionary learning for ECG reconstruction from PPG," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 936–940.

[47] G. H. Tison et al., "Passive detection of atrial fibrillation using a commercially available smartwatch," *JAMA Cardiol.*, vol. 3, no. 5, pp. 409–416, 2018.

[48] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project," *J. Open Res. Soft.*, vol. 7, no. 1, pp. 1–9, 2019.

[49] P. Van Gent, H. Farah, N. Van Nes, and B. Van Arem, "HeartPy: A novel heart rate algorithm for the analysis of noisy signals," *Transp. Res. Part F: Traffic Psychol. Behav.*, vol. 66, pp. 368–378, 2019.

[50] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[51] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11534–11542.

[52] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19 .

[53] M. Yu, Z. Huang, Y. Zhu, P. Zhou, and J. Zhu, "Attention-based residual improved u-net model for continuous blood pressure monitoring by using photoplethysmography signal," *Biomed. Signal Process. Control*, vol. 75, 2022, Art. no. 103581.

[54] Z. Zhao and Y. Zhang, "SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation," *Front. Physiol.*, vol. 9, 2018, Art. no. 727.

[55] Q. Zhu, X. Tian, C.-W. Wong, and M. Wu, "ECG reconstruction via PPG: A pilot study," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inf.*, 2019, pp. 1–4.

[56] Q. Zhu, X. Tian, C.-W. Wong, and M. Wu, "Learning your heart actions from pulse: ECG waveform reconstruction from PPG," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16734–16748, Dec. 2021.

[57] T. Zhu, P. Watkinson, and D. A. Clifton, "Smartwatch data help detect covid-19," *Nature Biomed. Eng.*, vol. 4, no. 12, pp. 1125–1127, 2020.