ECG Fiducial Points Localization Using a Deep Learning Model

1st Murtadha D. Hssayeni 2nd Arash Andalib 3rd Rishabh Singh 4th Diego Pava 5th Kan Li 6th Kaustubh Kale

AventuSoft LLC Boca Raton, FL ORCID | *AventuSoft LLC* Boca Raton, FL ORCID | *AventuSoft LLC* Boca Raton, FL ORCID | *AventuSoft LLC* Boca Raton, FL EMAIL | *AventuSoft LLC* Boca Raton, FL ORCID | *AventuSoft LLC* Boca Raton, FL ORCID

***Abstract*—ECG signals are essential in diagnosing cardiovascular diseases (CVD). Automatic localization of ECG fiducial points helps in the end-point detection and tracking of CVD. Nowadays, collecting ECG signals is more accessible due to the availability of wearable devices. We develop an algorithm to estimate the peaks of P and T waves and the onset and offset of the QRS complex. We evaluate it using ECG signals collected using a wearable device named HEMOTAG. The algorithm combines a rule-based method for heartbeat detection and a deep CNN for fiducial points localization. Three datasets are used to train and evaluate the proposed algorithm. The first and second datasets are QT and Lobachevsky University Electrocardiography Database (LUDB), which are used in ten-fold cross-validation. The third dataset is collected using HEMOTAG, which is used as a held-out set. A percentage of error (PoE) less than 1.75% is achieved based on the cross-validation, and PoE less than 2.42% is achieved based on the held-out set.**

***Index Terms*—Wearable devices, electrocardiography, ECG fiducial points, deep learning, cardiovascular disease.**

I. INTRODUCTION

Wearable devices have been developed to collect surface electrocardiogram (ECG), increasing its ease of use and accessibility [1]. These devices are employed for the early detection and long-term monitoring of cardiovascular diseases (CVD) [2]. Heartbeat detection and fiducial points localization in ECG signals are the key steps for end-point detection of CVD. The fiducial points include the onsets, offsets, and peaks of P, QRS, and T waves. These points are used to measure different time intervals in a heartbeat and help detecting and interpreting abnormalities, if any.

Heartbeat detection in ECG signals is well-established in the field with a high accuracy using traditional rule-based methods [3], [4]. However, these methods are prone to error when used for detecting fiducial points and rule adaptation is required when tested in new domains [5]. Recently, complex and deep learning algorithms are proposed to increase the accuracy of fiducial points' localization or delineation by utilizing their data-driven feature extraction ability [5]–[8]. In the work of Kalyakulina *et al.* and Abrishami *et al.*, the localization is formulated as a regression problem where the models output continuous values [6]–[8]. Jimenez-Perez *et al.* [5] formulated it as a binary classification problem where the models detect the presence of specific waveforms [5]. These algorithms are not evaluated on ECG signals collected using wearable devices, and their performance on ECG signals with different CVD is not assessed.

The collected ECG signals using wearable devices are impacted by noise at a higher rate than signals collected using traditional devices used in clinical settings. Deep learning models with the proper data augmentation methods to synthesize the noise during training can generalize better than rule-based methods. Therefore, we develop an algorithm to localize ECG fiducial points that combines a rule-based heartbeat detection method and a deep convolutional neural network (CNN). We use the Pan-Tompkins method [3], [4] as the rule-based method to localize R points. As a standardization method, the R points are used to segment ECG signals into multiple windows where each window contains a single heartbeat. The deep CNN localizes the peaks of P and T waves and the onset and offset of the QRS complex. Our contributions in this work are

1) We utilized and introduced new transformation-based data augmentation methods.
2) We analyzed the algorithm's performance for different groups of patients based on CVD. The analysis is essential since the intention is to use the algorithm in a clinical setting for people with possible CVD.
3) We evaluated the developed algorithm on ECG signals collected using a wearable device.

Being trained on an expanded dataset of multiple training sets, and due to different regularization techniques used, we could successfully train seven convolutional layers, which makes the deep CNN able to localize even the onset and offset features of the QRS complex, which is more challenging when compared with localizing a peak feature in P and T waves, as targeted by [8].

The rest of the paper is structured as follows: the datasets and preprocessing section describes three datasets utilized for the training and evaluation of the proposed algorithm and preprocessing steps; the methods section describes the proposed algorithm; the results section reports the data splits and

Research reported in this paper was supported by the National Heart Lung and Blood Institute of the National Institutes of Health under award number R44HL149561, to Aventusoft.

(corresponding author: Arash Andalib)

A. Andalib, D. Pava, K. Li, and K. Kale are with Aventusoft LLC. M. Hssayeni is currently with the University of Technology, Baghdad, Iraq. R. Singh is currently with the University of Florida, FL, USA

metrics used for evaluation, and the algorithm performance. We compare the results of the proposed algorithm with other models proposed in the literature and report its limitations in the discussion section. The last section is the conclusion.

II. DATASETS AND PREPROCESSING

Two retrospective datasets are used in this work which are QT [9] and Lobachevsky University Electrocardiography Database (LUDB) [6], [10]. Both datasets are retrieved from PhysioNet, a research resource for physiologic signals [11]. In addition, an in-house, prospective dataset that is collected using a wearable device is used. The device is HEMOTAG, shown in Figure 1, that captures cardiac vibrations and ECG.

A. QT dataset

QT dataset [9] contains two-channel ECG signals recorded for 15 minutes at 250Hz sampling frequency. The fiducial points (i.e., onset, peak, and offset for P, QRS, T, and U waves) in these signals were measured using an automatic algorithm and then partially corrected by two cardiologists. The automatic annotations were performed on the whole recordings and for each channel separately. The manual annotations were performed for few seconds in two passes considering both channels (i.e., single annotations were provided from both channels and not separate as in the automatic annotations). The annotated segments have at least 30 beats. One cardiologist annotated the signals for all the subjects, and the other annotated the signals for eleven subjects only.

There are 105 subjects in the QT dataset divided into seven subsets chosen from among existing and high-variable ECG databases. The subsets are the MIT-BIH Arrhythmia, the MIT-BIH ST Change, the MIT-BIH Supraventricular Arrhythmia, the MIT-BIH Normal Sinus Rhythm, the European ST-T, sudden death patients from the BIH, and the MIT-BIH Long-Term ECG.

B. LUDB dataset

LUDB dataset [6], [10] contains twelve-channel ECG signals recorded for 10 seconds at 500Hz sampling frequency. The signals were collected from 2017 to 2018. The waveforms' fiducial points in these signals were annotated manually by cardiologists. The manual annotations were performed on the whole duration of the recordings for each of the twelve channels separately. There are 200 subjects in LUDB dataset. One hundred forty-three of them were healthy, the others had various cardiovascular diseases, and some had pacemakers.

C. HEMOTAG dataset

HEMOTAG dataset, among other signals, contains one-channel ECG recorded for 30 seconds at 488Hz sampling frequency using HEMOTAG, a small portable device that uses micro-sensors to capture cardiac vibrations and electrocardiogram, transduced via thoracic electrodes [12]–[15]. The data was collected in 2021, and manually annotated by an specialist for the peaks of P and T waves and the onset and offset of the QRS complex. The manual annotations were performed for the whole duration of the recordings. There were 11 subjects in the HEMOTAG dataset. Seven of them were male, and four were female. The subjects had various CVD such as atrial fibrillation, ischemic cardiomyopathy, and previous myocardial infarction.

The research material from human participants recorded under IRB Protocol Number HT-17-VV-010. HEMOTAG is a registered trademark by Aventusoft, Florida. Patent No: US 10165985 B2, Patent No: US 8475396 B2, US 2017/0199962 A1, US 10531839 B2, US 2020/0170527 A1. Patents pending.

D. Preprocessing

In this work, we preprocessed the signals in both the QT and the LUDB datasets for noise removal and data standardization to train and evaluate the CNN model.

For the QT dataset, the first step was filtering the signals using a band-pass filter (0.5-40 Hz) to help remove the bias and high-frequency noise [16]. The second step was using the second-pass manual annotations done by the first cardiologist as they are more accurate than the first-pass annotation, as used in *et al.* [5]. Also, we used the automatic heartbeat detection (i.e., R points). The third step is selecting the beats with annotations for all fiducial points to be used for training and validation. The main reason for this step is that some of the heartbeats have missing manual annotations.

For LUDB, the first step was downsampling the signals to a 250 Hz sampling frequency. The second step was selecting only channels I and II and the associated annotations. We chose these channels since they show similar waveforms to the collected signals using HEMOTAG. The third step was filtering the signals using a band-pass filter (0.5-40 Hz).

The last step was segmenting the signals as described in subsection III-B. We used the automatic annotations of R points in the QT dataset and the manual annotations of R points in the LUDB dataset for segmentation. The Pan-Tompkins model described in subsection III-A was used in the HEMOTAG dataset. The combination of both LUDB and QT datasets resulted in a set of 305 subjects. This set had 5.5k segments with full manual annotations from channel I of the QT dataset and channels I and II from the LUDB dataset. Also, it had 96k segments with automatic annotations from the QT dataset (93.2k segments from channels I and 2.8k segments from channel II).

III. METHODS

The proposed algorithm for the localization of fiducial points in 1-lead ECG signals has three main components, as shown in Figure 1. The first component is the Pan-Tompkins model for R-point localization, the second component is a segmentation of heartbeats, and the third component is a CNN model for fiducial points localization.

A. Pan-Tompkins model for R-point localization

The localizing of R points in ECG recording is well-established in the field; thus, we utilized previous Pan-Tompkins models to perform this operation [3], [4]. The

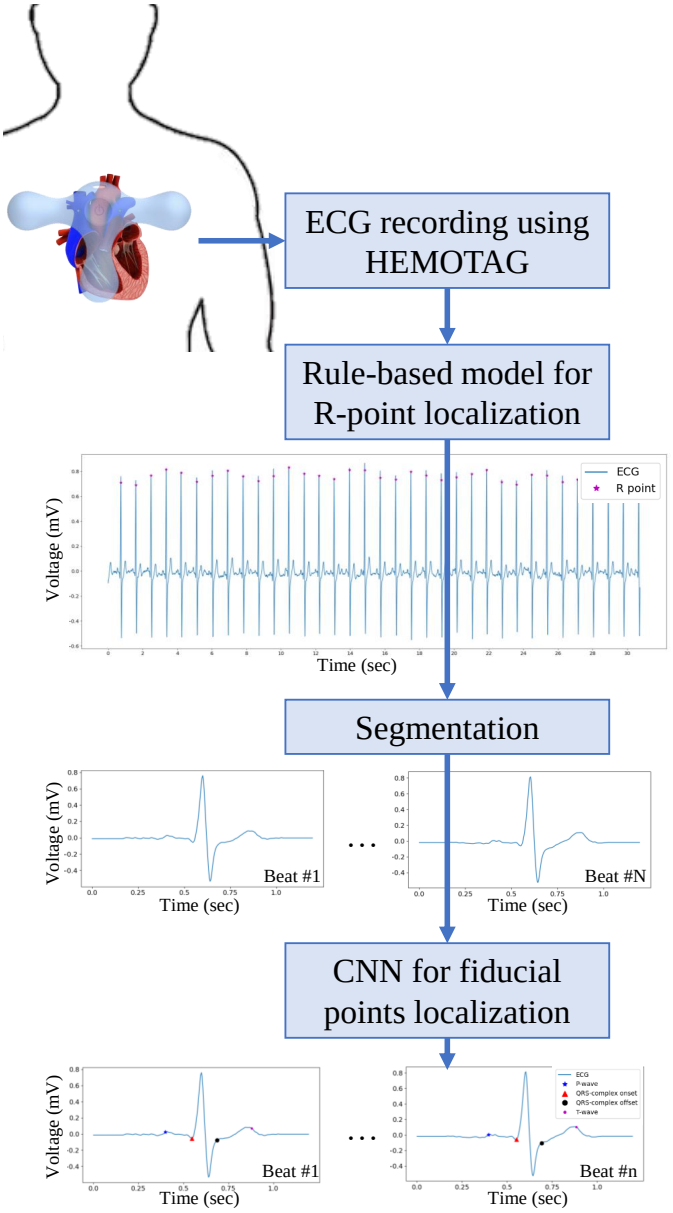

Fig. 1. The architecture of the proposed algorithm for the localization of fiducial points in 1-lead ECG signals. 30-sec ECG signal collected using HEMOTAG is shown as an example for the output of each component.

output from this component is the location of the R points ($\{t_r^i\}_{i=1}^{N_{beats}}$) for each heartbeat as shown in Figure 1.

B. Segmentation

Each detected heartbeat is segmented into a single window of size 300 samples in this step. The use of a fixed window length is a requirement for CNN networks. Each window is centered around the R point. We follow a few steps to ensure each widow has only one heartbeat. The first step is calculating the average beat-to-beat interval (t_{bb} in samples) using the R points. The second step constructs a window by starting from the location $t_r^i - (t_{bb}/2)$ to $t_r^i + (t_{bb}/2)$. The third step is padding the begging and end of the window to 300 samples by repeating the sample $t_r^i - (t_{bb}/2)$ and $t_r^i + (t_{bb}/2)$, respectively. Figure 1 shows an example of this component's output.

C. Convolutional Neural Network for fiducial points localization

CNN is a data-driven model that is constructed of convolutional filters and pooling operations divided into several layers. These filters and pooling operations make the CNN shift-invariant. CNN networks are successfully applied to solve regression problems in biomedical applications [17], [18] such as estimating the heart rate using Photoplethysmography (PPG) sensors [19], estimating the severity of Parkinson's disease (PD) [20], and estimating positive and negative affects and human behavior using physiological signals from wearable sensors [18], [21]. The CNN network can extract features from the raw signals and then perform the classification or regression through the fully-connected layers.

1) Architecture: Figure 2 shows the proposed 1-D CNN that takes the raw ECG windows as input and estimates four outputs representing P-wave peak, onset and offset of QRS complex, and T-wave peak. The network consists of four convolutional blocks. The first block consists of two layers, each with 64 convolutional filters of size 1×3, that are followed by a max-pooling layer. The second and third blocks also consist of two layers but deeper by doubling the number of filters to 128. The fourth block has one convolutional layer with 64 filters followed by a max-pooling layer and a flatten layer to transfer the feature maps to a vector before feeding it to two fully-connected layers (fc). The first fc has 512 nodes, and the output fc has four nodes for each of the fiducial points.

2) Training procedure: In both the QT and LUDB datasets, a relatively small number of heartbeats were manually annotated by cardiologists to mark all fiducial points, but a large number of heartbeats were automatically annotated. Training by considering only the scarce manual annotations will lead to an overfitted model. Therefore, four types of regularization techniques were used during training to prevent the model from overfitting the training ECG windows and thus make it generalize better on unseen ECG windows [22].

The first technique was training the network on noisy labels (i.e. the automatic annotations in QT dataset) and then fine-tuning it using the accurate labels (i.e., the manual annotations in both QT and LUDB datasets). The first step was performed for 200 epochs, and the second step was performed for 100 epochs. An early stopping procedure was implemented in both steps with a patience parameter of 20. The second technique was adding dropout layers before each fully-connected layer during training [23]. The third technique was adding L2 norm as a regularization parameter to the loss function. The fourth technique was applying eight signal augmentation methods. Six of them were performed based on the methods proposed by Jimenez-Perez *et al.* [5], and we developed two novel methods. The augmentation methods are:

- White Gaussian noise with a signal-to-noise ratio (SNR) between 15 to 30 dB.

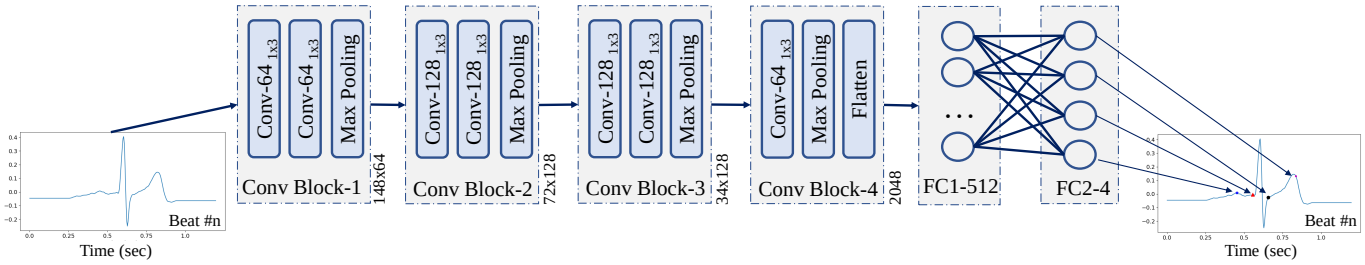

Fig. 2. The CNN network proposed in this work to localize fiducial points in ECG windows. The shape of the output of each convolutional block is shown after each block.

- Random periodic spikes with an SNR between 5 to 30 and a period between 50 to 150 samples.
- Powerline noise with an SNR between 15 to 30 and and a frequency between 37.5 to 62.5 Hz.
- Baseline wander with an SNR between 15 to 30 and and a frequency between 0.11 to 0.19 Hz.
- Pacemaker spikes with an SNR between 5 to 30.
- Amplifier saturation with a threshold representing percentage of the signal to be saturated in the range of 0 to 50%.
- Left/right shifting with a range 50 samples of left shift to 50 samples of right shift. This can help the model to estimate the fiducial points even if the segments were not centered correctly around R point.
- Time warping by stretching/shrinking in the temporal domain at a rate 0.5 to 1.5. This method is simulating higher and lower heart rates.

Figure 3 shows sample windows after applying the augmentation methods. At the beginning of each training iteration, one of the methods was selected randomly. Then the associated parameter (such as SNR and threshold) with this method was initialized randomly in the range given above in the list.

The model was trained using Adam optimizer [24] to minimize a mean squared error loss function of a mini-batch of size 64. The learning rate was 1e-4, and the dropout rate was 0.1. The L2 regularization rate was 1e-5.

IV. RESULTS

The heartbeats have high intra-recording similarity for each specific subject, and high inter-channel similarity [5]. Therefore, shuffling the windows (i.e., heartbeats) randomly and splitting them into training and testing sets is not the right way to evaluate the model performance because of the leaked similarity from training to test sets that deep learning models easily memorize. We solved this issue by performing a 10-fold cross-validation based on subject-wise splits. Each fold had the windows of 30 subjects, except the last fold, which had only 35 subjects. The results reported in this section are averaged for all the subjects after training the model 10 times on different training and testing splits based on the cross-validation. The windows of 10 subjects from the training splits were used for validation to select the best model and tune the hyper-parameters. Finally, we evaluated the trained model on QT and LUDB datasets on the holdout HEMOTAG dataset.

The metrics used for evaluation were first the mean absolute error (MAE) between the gold-standard and estimated fiducial points ($fid \in \{$*P-wave peak, QRS-complex onset and offset, and T-wave peak*$\}$ as shown in Equation 1. The second metric is the standard deviation (STD) of the mean absolute error across subjects. The third metric is the percentage of error (PoE) which is the MAE divided by the beat-to-beat interval of each subject, as shown in Equation 2.

$$MAE_{fid} = \frac{1}{N_{sub}} \sum_{s=1}^{N_S} \frac{\sum_{i=1}^{N^s_{beats}} |t_{fid}[i] - \hat{t_{fid}}[i]|}{N^s_{beats}} \quad (1)$$

where N_{sub} is the total number of subjects (i.e 305), N^s_{beats} is the number of beats (i.e., windows) of subject s, and $t_{fid}[i]$ is the estimated fiducial point.

$$PoE_{fid} = \frac{1}{N_{sub}} \sum_{s=1}^{N_S} \frac{\sum_{i=1}^{N^s_{beats}} |t_{fid}[i] - \hat{t_{fid}}[i]|}{N^s_{beats} \times t^s_{bb}} \times 100\% \quad (2)$$

where t^s_{bb} is the median of the beat-to-beat intervals of subject s.

In addition, we calculated additional metrics to make a direct comparison with other algorithms in the literature. The relative error ($RelE$) was calculated as the average difference between the gold-standard and estimated fiducial points without taking the absolute value as described in [5]. The precision ($\frac{TP}{TP+FP}$) was also calculated where TP is the number of the true-positive windows with MAE of 150 ms at most, FP is the number of false-positive windows with MAE of 150ms or higher. Using 150ms interval was following related work. The recall was one since we considered a window for each heartbeat and the model always estimates the fiducial points for a given window.

A. Evaluation on QT and LUDB Datasets

The testing results of the proposed CNN model including MAE, PoE, $RelE$, and STD for each of them are shown in Table I. MAE ranged from 7.3 ms for QRS-complex onset to 16.42 ms for T-wave peak. PoE was below 2% for all the estimated fiducial points. $RelE$ was very low and ranged between -1.29 ms to 0.68 m, but its STD was higher and

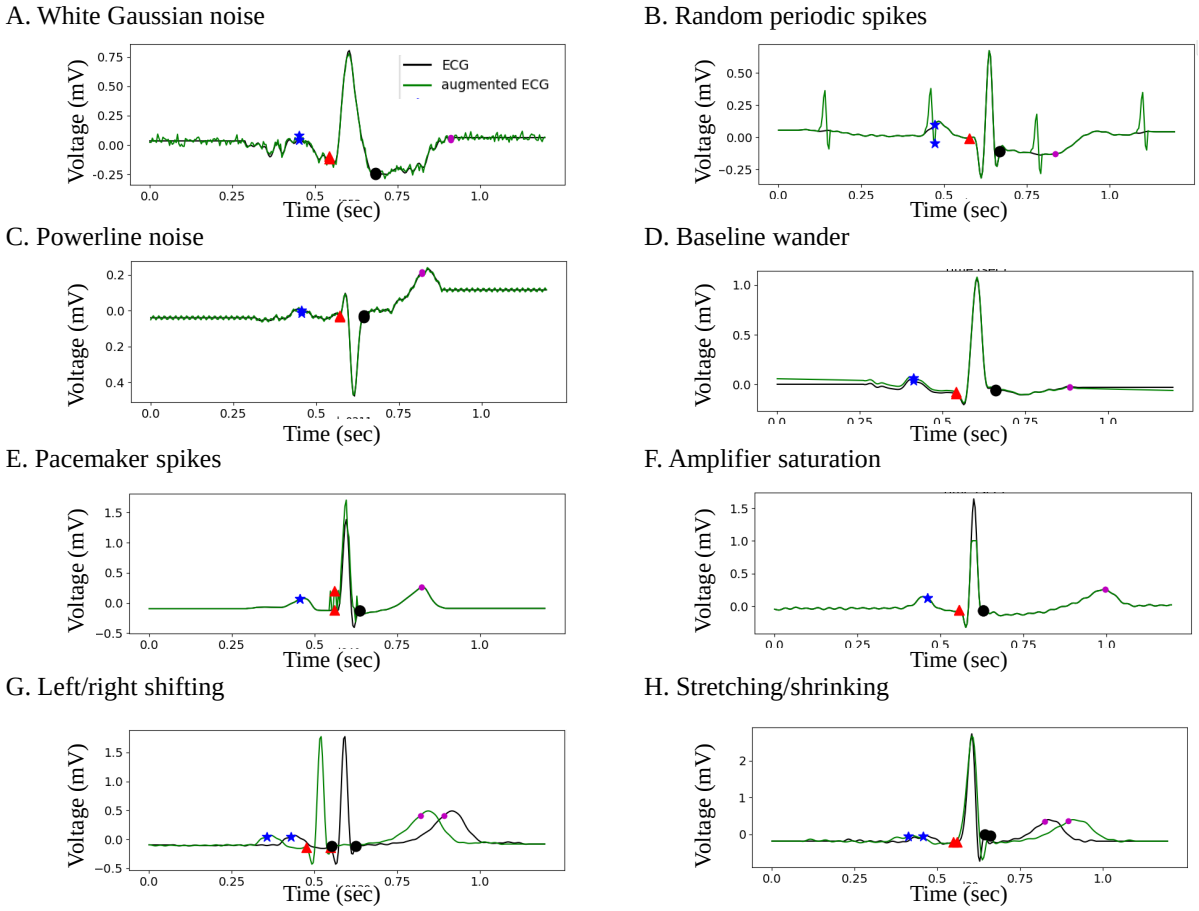

Fig. 3. Sample windows after applying each of the augmentation methods used in this work.

ranged between 11.41 ms for QRS-complex onset to 31.71 ms for T-wave peak.

TABLE I
THE TESTING RESULTS AVERAGE OF THE PROPOSED CNN MODEL USING QT AND LUDB DATASETS INCLUDING THE MEAN ABSOLUTE ERROR (MAE), THE PERCENTAGE OF ERROR (PoE) AND THE RELATIVE ERROR ($RelE$).

	P-wave Peak	**QRS-complex Onset**	**QRS-complex Offset**	**T-wave Peak**
MAE **(ms) ± STD**	11.19 ± 6.28	7.30 ± 4.57	9.71 ± 5.64	16.42 ± 8.08
PoE **(%) ± STD**	1.22 ± 0.71	0.81 ± 0.54	1.08 ± 0.65	1.75 ± 0.90
$RelE$ **(ms) ± STD**	-1.29 ± 8.45	-1.11 ± 6.27	0.57 ± 7.84	0.68 ± 10.70

Subjects with cardiovascular diseases such as arrhythmias and ST segment abnormalities may have different heart waveforms than people with normal sinus rhythms, which might affect the localization performance of deep models. Therefore, we averaged the error percentage for the estimated fiducial points in the different sets found in the QT and LUDB dataset, as shown in Figure 4. All subsets had PoE lower than or about 3% for all the fiducial points except the subset of sudden-death patients with about 5% PoE for the T-wave peak. Inspecting the recordings of sudden-death patients showed that the T wave did not appear in some of them, and they were corrupted with noise more than other subsets.

B. Evaluation on HEMOTAG Dataset

The CNN model performance on the HEMOTAG Dataset that was unseen before and collected using the HEMOTAG is shown in Table II. The model showed a better performance in estimating the T-wave peak but slightly lower performance in estimating the other fiducial points than the performance on QT and LUDB datasets. However, the PoE was still low, with a maximum of 2.42% in estimating the P-wave peak. The reason for the lower performance could be the domain adaptation problem with CNN networks since the electrodes of the HEMOTAG device have different placement than the devices used in QT and LUDB datasets. Also, other noise distributions and levels are present. The different placement may affect the morphology of the ECG signal.

Table III shows the testing mean absolute error (MAE) (in samples) of the proposed CNN model for each subject in the prospective HEMOTAG dataset beside the notes from the specialist. The low performance in estimating P-wave peaks is explained in the specialist's notes where P waves did not appear or were noisy and fainted (subject #6, 8, and 10).

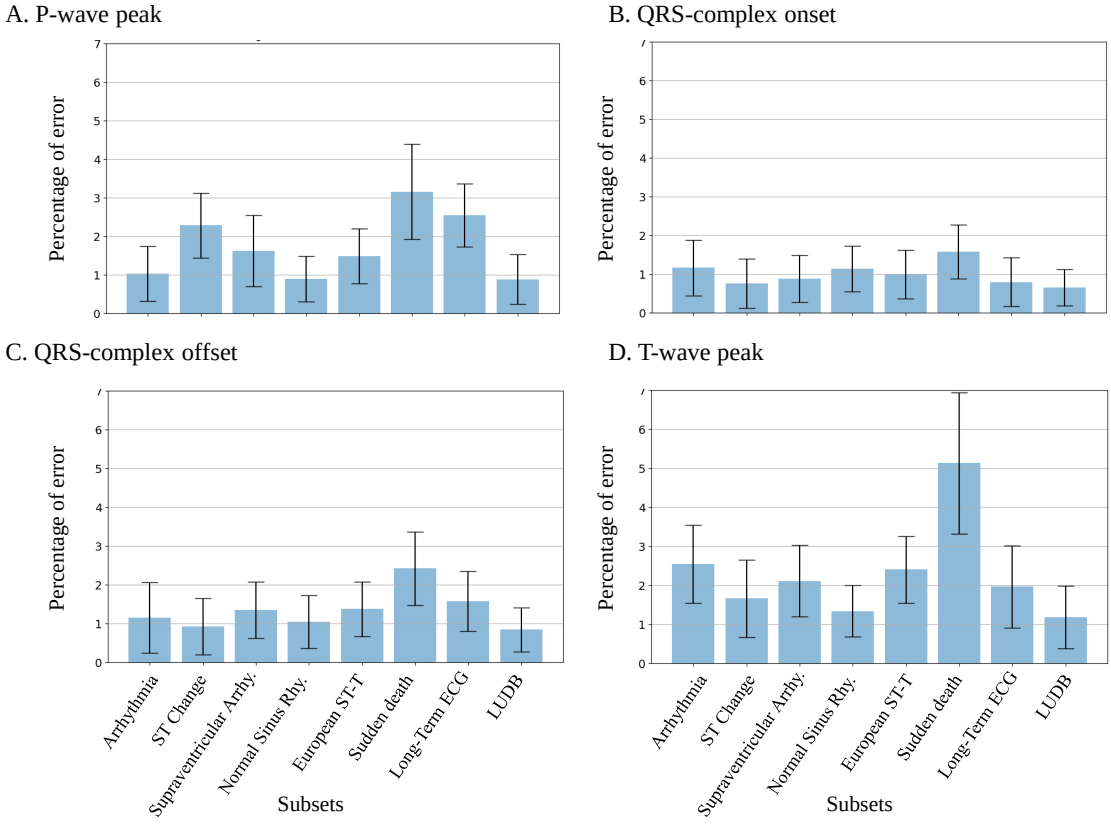

Fig. 4. The percentage of error for the estimated fiducial points in the different sets found in QT and LUDB dataset.

TABLE II
THE TESTING RESULTS OF THE PROPOSED CNN MODEL USING THE PROSPECTIVE HEMOTAG DATASET INCLUDING THE MEAN ABSOLUTE ERROR (MAE), THE PERCENTAGE OF ERROR (PoE) AND THE RELATIVE ERROR ($RelE$).

	P-wave Peak	**QRS-complex Onset**	**QRS-complex Offset**	**T-wave Peak**
MAE **(ms) ± STD**	22.86 ± 13.78	9.28 ± 5.66	16.46 ± 10.97	13.66 ± 14.00
PoE **(%) ± STD**	2.42 ± 1.50	1.06 ± 0.64	1.73 ± 1.14	1.44 ± 1.43
$RelE$ **(ms) ± STD**	4.63 ± 17.05	2.45 ± 7.95	-8.34 ± 12.44	2.66 ± 16.98

The issue in capturing the P wave is apparent in the ECG recordings in Figure 5 B that show five heartbeats for the subjects in the HEMOTAG dataset with the manual and CNN estimations of fiducial points. Subject 6 had an arrhythmia, and the P waves were absent, so the specialist and the CNN model were unsure of the peaks' locations.

C. Comparison between the Pan-Tompkins model and CNN model in localizing QRS-complex onset and offset

We calculated the MAE for the estimated onset and offset of the QRS complex using the proposed CNN and the Pan-Tompkins models for each of the three datasets, as shown in Table IV. CNN significantly outperformed the Pan-Tompkins model in estimating QRS-complex offset. The Pan-Tompkins

TABLE III
THE TESTING MEAN ABSOLUTE ERROR (MAE) (IN MS) OF THE PROPOSED CNN MODEL FOR EACH SUBJECT IN THE PROSPECTIVE HEMOTAG DATASET BESIDE THE NOTES FROM THE SPECIALIST. SAMPLING FREQUENCY IS 250 HZ.

Subject #	MAE **(ms) ± STD**				**Notes**
	P-wave Peak	**QRS-complex Onset**	**QRS-complex Offset**	**T-wave Peak**	
1	11.52 ± 10.48	15.16 ± 3.72	19.52 ± 10.56	10.84 ± 9.48	Fainted P waves
2	16 ± 10.6	11.16 ± 8.16	9.24 ± 8.68	18.24 ± 13.32	Noisy P waves
3	5.56 ± 5.56	14.92 ± 11.12	14.28 ± 7.44	10.4 ± 11.52	
4	3.72 ± 3.28	7.24 ± 2.32	8.56 ± 2.52	3.08 ± 2.4	
5	12.84 ± 14.32	8.72 ± 5.96	17.44 ± 17.16	44.44 ± 70	Noisy T waves
6	32 ± 24.52	6.48 ± 4.96	7.04 ± 4.44	6.96 ± 5.68	Fainted or missing P waves
7	28.28 ± 8.64	12.68 ± 7.68	9.88 ± 10.68	11.72 ± 7.44	
8	53.76 ± 35.32	4 ± 3.76	11.04 ± 17	14.56 ± 10.12	Fainted or missing P waves
9	46.16 ± 11.96	8.48 ± 4.2	15.52 ± 3.12	12 ± 8.88	Noisy ECG
10	28.12 ± 17.4	5.16 ± 4.12	10.52 ± 10.52	10.52 ± 9.08	Fainted or missing P waves
11	13.84 ± 9.48	8 ± 6.24	39.72 ± 23.96	7.56 ± 6.08	

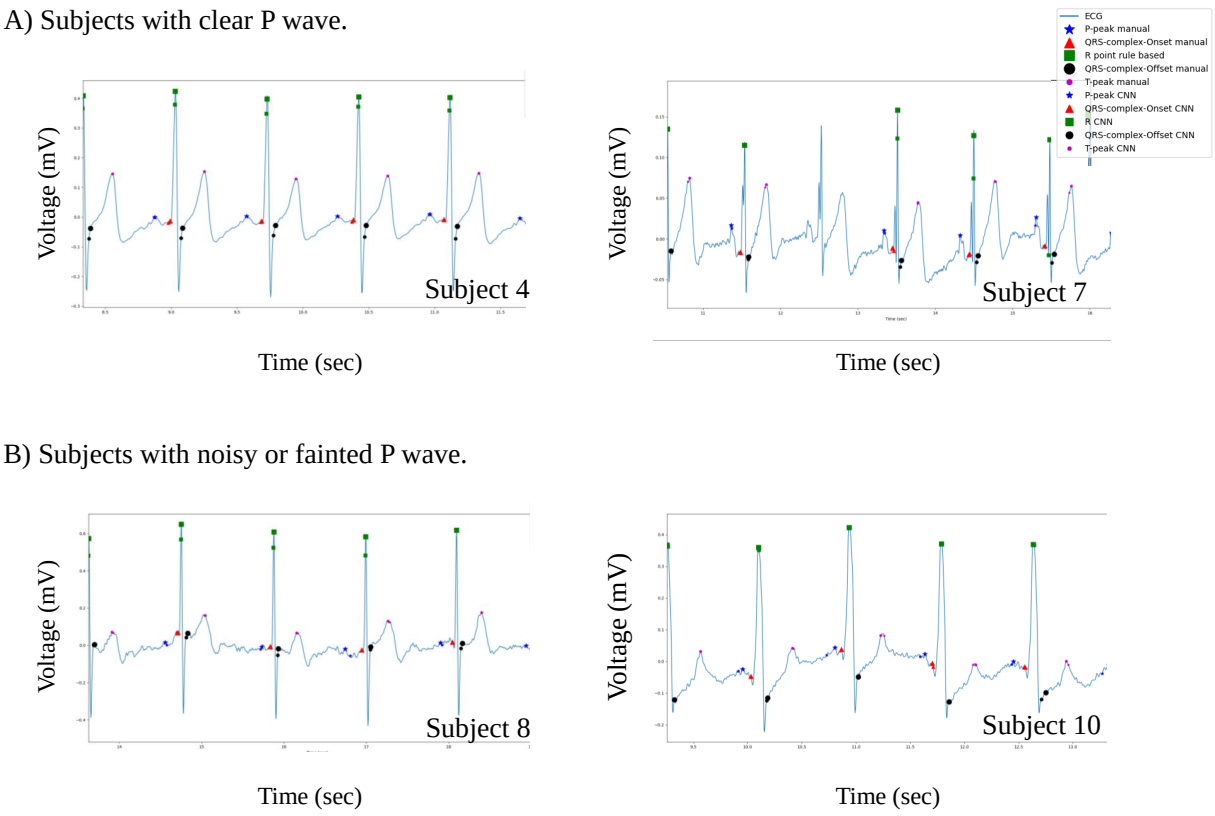

Fig. 5. ECG recording samples of 5 heartbeats for subjects from the HEMOTAG dataset. Clear P waves are shown in part A, and noisy or fainted P waves are shown in part B. The manual and CNN estimations of fiducial points are localized, and a small size marker is used for CNN estimations.

model had a high MAE and STD (23.56±17.12ms) in estimating the offset. Also, CNN had lower MAE in estimating the QRS-complex onset in all datasets. Also, the STD in estimating the onset by the CNN is lower than that of the Pan-Tompkins model because it estimated the onset of two subjects with about 30ms MAE. In general, the MAE in the estimated onset using the Pan-Tompkins model is increasing for noisy ECG and abnormal QRS wave-forms for unhealthy subjects, whereas the CNN is performing consistently better.

TABLE IV
TESTING ERROR OF THE PAN-TOMPKINS MODEL AND CNN MODEL IN ESTIMATING THE QRS-COMPLEX ONSET AND OFFSET.

Dataset	Evaluated metric	QRS-complex onset		QRS-complex offset	
		Pan-Tompkins	CNN	Pan-Tompkins	CNN
QT	MAE **(ms)** ± **STD**	13.98 ± 16.58	**9.76** ± **5.66**	15.35 ± 8.11	**13.14** ± **6.91**
LUDB	MAE **(ms)** ± **STD**	9.36 ± 6.47	**5.94** ± **3.98**	13.45 ± 9.06	**7.83** ± **4.94**
HEMOTAG	MAE **(ms)** ± **STD**	10.54 ± 8.56	**9.28** ± **5.66**	23.56 ± 17.12	**15.41** ± **11.24**

V. DISCUSSIONS

We are presenting a deep CNN with seven convolutional layers, utilizing a Pan-Tompkins R-point detection method, and performing a comprehensive analysis. Training a deeper CNN, when compared with existing works,was possible due to expanding the training set by including the LUDB dataset and applying four regularization techniques, including data augmentation methods that improved the performance of this work significantly.

A direct comparison with the proposed methods in the literature is summarized in Table V. Jimenez-Perez *et al.* [5] developed a fully-convolutional neural network (names U-Net) to delineate the onset and offset of P wave, T wave, and QRS complex. Kalyakulina *et al.* [6], [7] developed a model based on Wavelet analysis to localize the onset, offset and peaks of P wave, T wave, and QRS complex, except the onset of T wave. The proposed CNN model yielded the highest precision in both QT and LUDB datasets compared to the other methods. Also, the proposed CNN model yielded the lowest relative error for estimating three fiducial points and a comparable error for P peak estimation. Our algorithm's estimation error of T-wave fiducial points was significantly lower than other related work.

The largest error was reported in estimating T-wave peak, which was understandable given the complexity associated with T waves. Other researchers who reported the maximal error in estimating T-wave fiducial points [6], [25] have reported T waves complexity. T waves, for some subjects, are valleys instead of a hill shape. On the QT dataset, intra-observer bias in estimating T-wave onset was -51.96±105.88ms, and inter-observer bias for the same manual estimation was -9.52±44.85ms [5]. However, the CNN model achieved a significantly lower error (5.82±15.78).

TABLE V
THE PERFORMANCE MEASURES OF THE PROPOSED ALGORITHM IN COMPARISON WITH RELATED WORK IN LITERATURE. THE BEST METRICS FOR EACH OF THE DATASETS IS IN BOLD. N/A MEANS NOT REPORTED.

Method	Metric	P-wave Peak	QRS-complex Onset	QRS-complex Offset	T-wave Peak
Proposed CNN model QT dataset	**Precision**	**99.87**	**100.00**	**100.00**	**98.62**
	$RelE$ **(ms)**	**3.03**	-0.42	**0.20**	**3.94**
	± STD	**± 9.35**	± 7.55	**± 8.73**	**± 11.73**
Proposed CNN model LUDB dataset	**Precision**	**100.00**	**100.00**	**100.00**	**100.00**
	$RelE$ **(ms)**	-1.98	**-0.14**	**-0.08**	**0.51**
	± STD	± 8.48	**± 6.01**	**± 7.99**	**± 10.8**
Wavelet analysis by Kalyakulina *et al.* [6], [7] QT dataset	**Precision**	97.89	98.24	98.24	98.24
	Recall	97.50	98.42	98.42	98.24
	$RelE$ **(ms)**	4.3	-5.1	4.7	7.2
	± STD	± 10.0	± 6.6	± 9.5	± 13.0
Wavelet analysis by Kalyakulina *et al.* [6], [7] LUDB dataset	**Precision**	96.41	99.87	99.87	98.84
	Recall	98.46	99.61	99.61	99.03
	$RelE$ **(ms)**	**-0.3**	-8.1	3.8	4.0
	± STD	**± 6.2**	± 7.7	± 8.8	± 7.4
U-Net by Jimenez-Perez *et al.* [5] QT dataset (single-lead)	**Precision**	90.12	99.14		98.25
	Recall	98.73	99.94		99.88
	$RelE$ **(ms)**	N/A	**-0.07**	3.64	N/A
	± STD		**± 8.37**	± 12.55	
CNN by Abrishami *et al.* [8] QT dataset	**Recall**	99.63	N/A	N/A	98.06
	$RelE$ **(ms)**	17.03			16.05
	± STD	± 17.69	N/A	N/A	± 18.02

A. Limitations and Next Steps

The proposed CNN network estimates the P-wave peak even if it was not clear in the ECG signals. However, this model assumes that all fiducial points are present, which is not the case with various cardiovascular diseases. In case of missing P or T waves, this model may fail and not give a prediction probability. Therefore, our first future work is to develop a method to estimate a confidence value for CNN localization. One possible approach is visualizing the heat maps of the CNN. One advantage of a CNN network is that it can be easily upgraded to estimate new labels by adding new output nodes. Our second future work is modifying the output layer to have nine nodes to estimate nine fiducial points. The points were onset, peak, and offset of the P and T waves, and QRS complex.

VI. CONCLUSIONS

We developed an algorithm to localize four ECG fiducial points: P peak, onset and offest of QRS wave, and T peak. The algorithm utilized a rule-based model to detect heartbeats in ECG signals that were standardized, segmented, and fed to a CNN model. The algorithm achieved high performance on both the standard ECG datasets and the HEMOTAG dataset with PoE lower than 2.42% in estimating the fiducial points. The testing performance showed the model's generalizability to unseen ECG signals with different CVD.

REFERENCES

[1] S. Ramasamy and A. Balan, "Wearable sensors for ecg measurement: a review," *Sensor Review*, 2018.

[2] J. Lin, R. Fu, X. Zhong, P. Yu, G. Tan, W. Li, H. Zhang, Y. Li, L. Zhou, and C. Ning, "Wearable sensors and devices for real-time cardiovascular disease monitoring," *Cell Reports Physical Science*, vol. 2, no. 8, p. 100541, 2021.

[3] X. Hu, J. Liu, J. Wang, and Z. Xiao, "Detection of onset and offset of qrs complex based a modified triangle morphology," in *Frontier and Future Development of Information Technology in Medicine and Education*, pp. 2893–2901, Springer, 2014.

[4] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[5] G. Jimenez-Perez, A. Alcaine, and O. Camara, "Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.

[6] A. I. Kalyakulina, I. I. Yusipov, V. A. Moskalenko, A. V. Nikolskiy, K. A. Kosonogov, G. V. Osipov, N. Y. Zolotykh, and M. V. Ivanchenko, "Ludb: a new open-access validation tool for electrocardiogram delineation algorithms," *IEEE Access*, vol. 8, pp. 186181–186190, 2020.

[7] A. I. Kalyakulina, I. I. Yusipov, V. A. Moskalenko, A. V. Nikolskiy, A. A. Kozlov, N. Y. Zolotykh, and M. V. Ivanchenko, "Finding morphology points of electrocardiographic-signal waves using wavelet analysis," *Radiophysics and Quantum Electronics*, vol. 61, no. 8, pp. 689–703, 2019.

[8] H. Abrishami, *Deep Learning Based Electrocardiogram Delineation*. PhD thesis, University of Cincinnati, 2019.

[9] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg," in *Computers in cardiology 1997*, pp. 673–676, IEEE, 1997.

[10] A. Kalyakulina, I. Yusipov, V. Moskalenko, A. Nikolskiy, A. Kozlov, K. Kosonogov, N. Zolotykh, and M. Ivanchenko, "Lobachevsky university electrocardiography database (version 1.0. 0)," *PhysioNet*, 2020.

[11] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[12] K. Kale and A. Andalib, "System and method of marking cardiac time intervals from the heart valve signals," June 4 2020. US Patent App. 16/741,740.

[13] K. Kale and L. G. S. Giraldo, "System and method of extraction, identification, marking and display of heart valve signals," July 30 2019. US Patent 10,362,997.

[14] D. Pava, A. Andalib, K. Li, and K. Kale, "A novel non-invasive device for screening and optimized management to improve heart failure outcomes in patients with diabetes mellitus," *Journal of Cardiac Failure*, vol. 28, no. 5, p. S15, 2022.

[15] K. Li, D. Pava, A. Andalib, and K. Kale, "Non-invasive assessment of elevated left ventricular end diastolic pressure using a small portable device: Design for tele-monitoring," *Journal of Cardiac Failure*, vol. 28, no. 5, p. S13, 2022.

[16] A. Peimankar and S. Puthusserypady, "Dens-ecg: A deep learning approach for ecg signal delineation," *Expert Systems with Applications*, vol. 165, p. 113911, 2021.

[17] C. Cao, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo, and Z. Xie, "Deep learning and its applications in biomedicine," *Genomics, proteomics & bioinformatics*, vol. 16, no. 1, pp. 17–32, 2018.

[18] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in bioinformatics*, vol. 18, no. 5, pp. 851–869, 2017.

[19] D. Biswas, L. Everson, M. Liu, M. Panwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Acharyya, C. Van Hoof, M. Konijnenburg, *et al.*, "Cornet: Deep learning framework for ppg-based heart rate estimation and biometric identification in ambulant environment," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 2, pp. 282–291, 2019.

[20] M. D. Hssayeni, J. Jimenez-Shahed, M. A. Burack, and B. Ghoraani, "Ensemble deep model for continuous estimation of unified parkinson's disease rating scale iii," *Biomedical engineering online*, vol. 20, no. 1, pp. 1–20, 2021.

[21] M. D. Hssayeni and B. Ghoraani, "Multi-modal physiological data fusion for affect estimation using deep learning," *IEEE Access*, vol. 9, pp. 21642–21652, 2021.

[22] J. Kukačka, V. Golkov, and D. Cremers, "Regularization for deep learning: A taxonomy," *arXiv preprint arXiv:1710.10686*, 2017.

[23] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[25] S. Mehta and N. Lingayat, "Detection of p and t-waves in electrocardiogram," in *Proceedings of the world congress on Engineering and computer science*, pp. 22–24, Citeseer, 2008.