DSD-R: Deep Learning Based Segmentation for the Detection of R peaks in ECG Signals

Eyüpcan Çelik, Furkan Eren, Murat Şahin, Deniz Kılınç
Kalybe AI
İzmir, Türkiye
{eyupcan.celik, furkan.eren, murat.sahin, deniz.kilinc}@kalybeai.com

Fatih Erdemir
Amatis
İzmir, Türkiye
f.erdemir@amatis.nl

Robert Den Engelsman
Fysiologic ECG Service
Amsterdam, Netherlands
r.engelsman@fysiologic.nl

Mehmet Kemal Güllü
Electrical and Electronics Engineering
İzmir Bakırçay University
İzmir, Türkiye
kemal.gullu@bakircay.edu.tr

***Abstract*— Electrocardiography (ECG) is the recording of the electrical activity of the heart through electrodes. ECG signals are crucial in the early diagnosis of numerous cardiac diseases. Therefore, it is very important to read and analyze these signals using state-of-the-art technologies. The regular wave shapes in ECG data are frequently disturbed when certain heart diseases occur and these changes in signals help for detecting the disease. Signal processing and machine learning-based methods are widely used for this purpose. In recent years, deep learning-based methods have become widespread, and they offer promising results. This study aims to segmentation-based detection of R-peak locations in ECG signals. First, the ECG signal is transformed into a Continuous Wavelet Transform (CWT) based scalogram image, and then U-Net-based deep learning architectures are utilized for the segmentation. The comparisons are carried out on MIT-BIH Arrhythmia Database (MIT-DB). Whereas all methods provide promising results, U-Net 3+ model achieves 0.99 in Precision, 0.98 in Recall, 0.99 in F1 score, and 0.98 in Accuracy with the lowest parameter size.**

Keywords—ECG; Scalogram; Segmentation.

I. Introduction

Electrocardiography (ECG), which is used in the diagnosis of cardiovascular diseases, is the recording of electrical activity that occurs during the contraction of the heart using electrodes. An electrical activity occurs during the contraction and relaxation of the atria and ventricles, also known as the heartbeat. Thanks to the measurement of the electrical activity created by the heart at each contraction, detailed information is obtained about many topics such as heart rhythm, frequency, and propagation. The ECG is recorded with electrodes glued on the chest, arm, and leg areas. An ECG signal consists of P, Q, R, S, and T waves. The graphical values generated by these waves provide the physician with information about the patient's heart health.

Cardiovascular diseases are common types of heart disease, and they can occur for a variety of reasons. Examples of these diseases include arrhythmias, which are caused by problems in the electrical conduction system of the heart, cardiomyopathies, which are caused by problems related to the functions of the heart muscle, congenital heart diseases, coronary artery diseases expressing problems in the vessels responsible for feeding the heart and oxygen support, and infection in the heart tissue. Since cardiovascular diseases are one of the diseases that cause the deadliest in the world, early diagnosis is crucial. A disease that cannot be detected its early stages, might have disastrous effects. On the other hand, accurate diagnosis, along with early diagnosis, is crucial and and requires trained specialist. The training of specialist is also a time-consuming and expensive endeavor. To reduce this cost, state-of-the-art machine learning approaches trained with data sets verified by experts can be utilized for the diagnosis of diseases.

Many studies have been conducted in the literature on the determination of R-peaks. Vijayarangan et al. obtained an F1 score of 0.9837 using the INcResU-Net architecture with a 1-dimensional ECG signal [1]. Zhou et al. obtained 90.61% accuracy using 1-dimensional convolutional neural networks (CNN) and long-short-term memory (LSTM) [2]. Zahid et al. obtained 99.83% F1 score, 99.85% Recall, and 99.82 Precision in MIT-DB using 1-dimensional CNN [3]. Merah et al. obtained 99.84% Sensitivity and 99.88% Predictivity in MIT-DB with a Stationary Wavelet Transform (SWT) based study [4]. Qin et al. obtained 99.39% sensitivity, 99.49% predictivity, and 99.89% accuracy in MIT-DB with an adaptive and time-efficient R-peak detection algorithm [5].

This study aims to successfully determine the R-peak locations using the input scalogram image obtained by taking CWT, and segmentation-based deep learning architectures. First, the 1-dimensional signal is downsampled from length 1440 to length 512. Then the CWT of the signal is computed. Then, the 2-D scalogram matrix is obtained by taking the magnitude of the complex CWT coefficients. Finally, the scalogram image is normalized and applied as input to various U-Net-based deep segmentation methods. The detection is carried on the segmentation map. MIT-DB is used in the experiments. The block diagram of the proposed method is shown in Fig. 1.

978-1-6654-5432-2/22/$31.00 ©2022 IEEE

2022 Medical Technologies Congress (TIPTEKNO) | 978-1-6654-5432-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/TIPTEKNO56568.2022.9960181

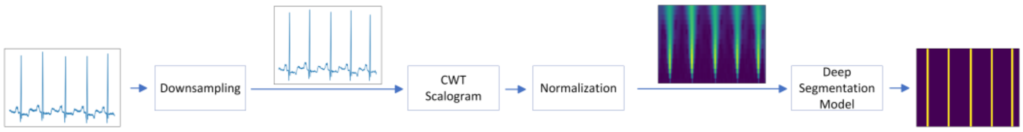


Fig. 1. Block diagram of the proposed method

II. Materials And Methods

A. MIT-BIH Arrhythmia Database

The MIT-DB was used as the data set in this study. The MIT-DB consists of 48 ECGs with a sampling frequency of 360 Hz and a duration of approximately 30 minutes each. These 48 ECGs belong to 47 patients [6]. In this study, 109.494 R-peaks from MIT-DB were used.

B. Data Preprocessing

In this section, the ECGs contained in the MIT BIH data set have been made desirable. First of all, each ECG is taken and divided into 4-second windows. each 4-second window corresponds to a length of 1440. A total of 21648 windows with a length of 1440 were obtained here. Later, these 1440-length windows were subsampled to 512 lengths. This subsampling was performed by the Fourier Method.

C. Continuous Wavelet Transform

CWT is used for the time-frequency analysis of a signal. It improves the localization property of the Short-Time Fourier Transform (STFT). CWT provides high time resolution and low-frequency resolution at high frequencies by adjusting the scale and position parameters. At low frequencies, it provides low time resolution and high-frequency resolution [7]. The CWT formula is given in Eq. 1.

$$T(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\psi^{*}\frac{(t-b)}{a}dt \tag{1}$$

Where a is the scale parameter, b is the position of the wavelet, ψ is the wavelet function, and x is the signal to be transformed.

In this study, the Morlet wavelets are used. The Morlet wavelet is the product of the complex exponential and the Gaussian window. This type of wavelet has been preferred because it makes R-peaks evident in ECG signals. As a result of CWT, a 2D matrix in the range of 0.68 Hz - 47.13 Hz was obtained. Since the R-peak is more prominent here, the range of 16.66 Hz - 47.13 Hz was chosen. Therefore, a 2D matrix of 16x512 was obtained.

D. Wavelet Scalogram

By taking the CWT of ECG signal using Morlet wavelets, a 2-D complex matrix is obtained. The 2-D complex matrix is converted into decimal numbers by taking the magnitude as given in Eq. 2. Finally, a 3-channel pseudo color image is obtained by applying Viridis color map.

$$magnitude = |a+bi| = \sqrt{a^2+(b)^2} \tag{2}$$

E. Labeling

The positions of the R-peaks are provided with the dataset. We created a label image of the same size as the scalogram image, which contains columns filled with ones around R-peak locations and zeros elsewhere. Column width is set to 10 in the experiments.

F. Normalization

Values in each input scalogram image are linearly scaled to the range of 0 and 1 as given in Eq. 3.

$$x' = \frac{x-\min(x)}{\max(x)-\min(x)} \tag{3}$$

G. Deep Learning Architectures

In this study, U-Net and its derivatives R2U-Net, Attention U-Net (Att U-Net), U-Net++, and U-Net 3+, which are segmentation-based architectures, are used with the proposed architecture.

1) U-Net

It is a type of enhanced convolutional neural network used for image segmentation. It was published in 2015 by Ronneberger et al. With the U-Net architecture, it is aimed to increase the data by using a limited number of labeled data. U-Net architecture has a network structure that first narrows and then expands [8].

2) R2U-Net

R2U-Net architecture is a U-Net architecture based on the Recurrent Residual Convolutional Neural Network (RRCNN). The residual network uses the U-Net and RCNN infrastructure. The residual unit in the R2U-Net architecture helps training during the training of the deep network architecture. Recurrent residual convolutional layers provide more powerful feature extraction for segmentation [9].

3) Attention U-Net

Attention gate is a method used for medical imaging that learns to automatically focus on structures of different shapes and sizes. Attention gate-trained models thoroughly emphasize salient features for a given structure. It also suppresses irrelevant regions in the input image. Attention gates improve model precision and accuracy. In addition, it can be easily integrated into CNN architectures such as U-Net with minimum computational load [10].

4) U-Net++

The U-Net++ architecture is a deeply supervised network in which the encoder and decoder sub-networks are connected by a series of nested, dense hopping paths. The network paths redesigned for this network aim to reduce the semantic gap between the feature maps of the sub-networks between encoder and decoder. Here, since the features in the decoder and encoder networks are semantically similar, the network learns what to do more easily [11].

5) U-Net 3+

The U-Net 3+ architecture utilizes full-scale jump links and deep inspections. Full-scale jump links combine high- and low-level details from feature maps at different scales. Deep inspections use the full-scale merged feature maps to learn hierarchical representations from there [12].

III. Experimental Results

In this study, 48 ECG signals in the MIT-DB were divided into 4-sec windows (1440 samples). Each window was then downsampled to 512 samples. Then, CWT of the obtained signals was taken and 16x512 2D matrices were obtained. A scalogram was obtained by taking the amplitude of this 2D matrix. The scalogram was normalized with a minimum-maximum scaler. Finally, segmentation-based models were trained with the generated input and labels. All segmentation models were trained with the same training set and tested with the same test set. For training, 75% of the data, i.e. #16237 images were used. For testing, 25% of the data, i.e. #5411 images, was used.

The Keras API of the Tensorflow library of the Python programming language was used to implement the segmentation models. The training of the segmentation models was carried out in the Google Colab environment. The tests were performed on a computer with Intel Core i7-10510U CPU and 16 GB Ram.

In the encoder layer of the segmentation models, there are convolution blocks with 32-64-128-256-512 filters respectively. Adam optimizer is used with 1e-4 learning rate. Dice loss was taken as loss function. Trainings were limited to 50 epochs. Deep supervision was also used in U-Net++ and U-Net 3++ models.

Accuracy, Precision, Recall, and F1 score metrics are used for comparison. Each R-peak occupies 10 pixels width rectangular shape in the label image. The overlapping ratio between the segmentation map and the label image is used for the detection decision. Detection performances for different overlapping ratios are evaluated, and metric comparisons of the segmentation models are given in Fig. 2. for overlapping ratios from 10% to 100%.

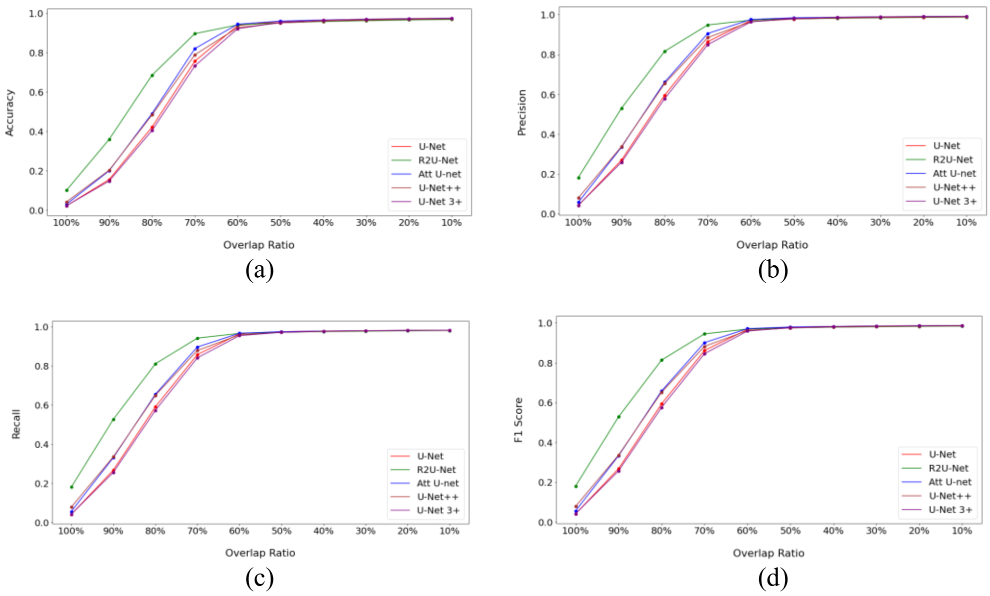


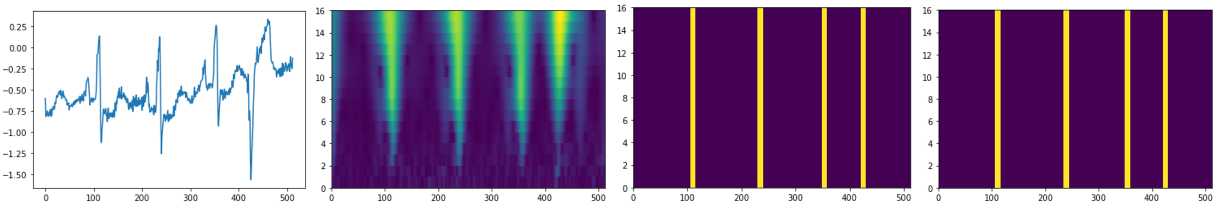
Fig. 2. Metric - Overlap Ratio graphs (a) Accuracy – Overlap Ratio (b) Precision – Overlap Ratio (c) Recall – Overlap Ratio (d) F1 Score – Overlap Ratio

When the results of Accuracy, Precision, Recall, and F1 score metrics are assessed, the R2U-Net model outperforms other models by a margin of 100% to 60% of overlap ratio. In the same overlap ratio ranges, The U-Net 3+ model gives the worst metric results. The success of the U-Net model is comparable to the U-Net 3+ in the same overlap range. The success of Att U-Net and U-Net++ models' performance is comparable to the U-Net 3+ and U-Net models at 100% overlap ratio and higher at a 90% overlap ratio. All models are approximately equally well, regardless of the overlap ratio, which ranges from 60% to 10%. The reason why the R2U-Net model has the highest performance at high overlap ratios may be due to the higher number of parameters compared to the other models. However, this is a drawback for the R2U-Net model in terms of prediction time. Since the best performances are obtained at 10% overlap ratio, the performance values for this ratio are only given in detail in Table 1. In Table 1, the U-Net 3+ model gives the best performance. Despite the Att U-Net model having the lowest prediction time is Att U-Net, the model with the highest prediction time is R2U-Net.

TABLE I. 10 % OVERLAP RATE METRIC RESULTS

Model	Accuracy	Precision	Recall	F1	Predict Time	#Params
U-Net	0.97	0.99	0.98	0.99	195 s	9.4 M
R2U-Net	0.97	0.99	0.98	0.98	526 s	26.3 M
Att U-Net	0.97	0.99	0.98	0.99	178 s	8.7 M
U-Net++	0.97	0.99	0.98	0.99	341 s	9.8 M
U-Net 3+	0.98	0.99	0.98	0.99	357 s	6.6 M

Fig 3. shows the input signal and outputs at different stages of the proposed method and the label image. Here (a) is an ECG signal of length 512. , (b) is the scalogram, (c) is the label image, and (d) is the segmentation result of the U-Net model.



(a) (b) (c) (d)

Fig. 3. Outputs of the proposed method (a) original signal, (b) scalogram, (c) label, (d) model output(prediction)

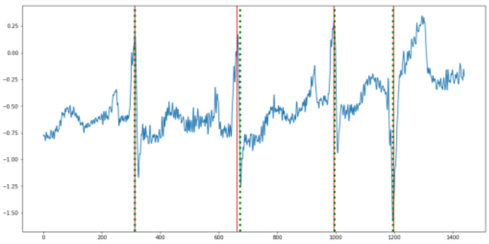


Fig. 4. R-peak detection output of the proposed architecture for the U-Net model (red line=ground truth, green dotted line = detected)

IV. CONCLUSIONS

This study aims to develop a segmentation-based method for the detection of R-peaks in ECG signals. For this purpose, the 1-dimensional signal was transformed into a 2-dimensional CWT matrix to extract the time and frequency characteristics of the R-peaks. For detection purpose, deep segmentation models were used. While all methods gave promising results in the tests, the most successful model was the U-Net 3+. It achieved 0.98 in Accuracy, 0.99 in Precision, 0.98 in Recall and 0.99 in F1 scores. Moreover, it has the least number of parameters with 6.6 M parameters. The Att U-Net model has the best prediction time with 178 s. This work can be easily integrated into other segmentation models in the future. And more successful models can be obtained. Additionally, higher performance scores can be achieved by increasing the amount of training data. This study can be used in hospitals in the future and can facilitate ECG analysis for those working in this field.

REFERENCES

[1] Vijayarangan, S., Vignesh, R., Murugesan, B., Preejith, S. P., Joseph, J., & Sivaprakasam, M. (2020, July). RPnet: A Deep Learning approach for robust R Peak detection in noisy ECG. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 345-348). IEEE.

[2] Zhou, P., Schwerin, B., Lauder, B., & So, S. (2020, December). Deep learning for real-time ecg r-peak prediction. In 2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS) (pp. 1-7). IEEE.

[3] Zahid, M. U., Kiranyaz, S., Ince, T., Devecioglu, O. C., Chowdhury, M. E., Khandakar, A., ... & Gabbouj, M. (2021). Robust R-Peak detection in low-quality holter ECGs using 1D convolutional neural network. IEEE Transactions on Biomedical Engineering, 69(1), 119-128.

[4] Merah, M., Abdelmalik, T. A., & Larbi, B. H. (2015). R-peaks detection based on stationary wavelet transform. Computer methods and programs in biomedicine, 121(3), 149-160.

[5] Qin, Q., Li, J., Yue, Y., & Liu, C. (2017). An adaptive and time-efficient ECG R-peak detection algorithm. Journal of Healthcare Engineering, 2017.

[6] Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine, 20(3), 45-50.

[7] Guo, M. F., Zeng, X. D., Chen, D. Y., & Yang, N. C. (2017). Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems. IEEE Sensors Journal, 18(3), 1291-1300.

[8] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[9] Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955.

[10] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[11] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 3-11). Springer, Cham.

[12] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., ... & Wu, J. (2020, May). Unet 3+: A full-scale connected unet for medical image segmentation. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1055-1059). IEEE.