ECG-signal multi-classification model based on squeeze-and-excitation residual neural networks

Junsang Park

Huinno Co., Ltd.

Jinkook Kim

Huinno Co., Ltd.

Sunghoon Jung

Huinno Co., Ltd.

Yeongjoon Gil

Huinno Co., Ltd.

Jong-Il Choi

Korea University Anam Hospital

HoSung Son (✉ hssonmd@korea.ac.kr)

Korea University Anam Hospital <https://orcid.org/0000-0003-2766-1186>

Research article

Keywords: ECG signal multi-classification, deep learning, convolutional neural network, arrhythmia

Posted Date: June 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-34360/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Accurate electrocardiogram (ECG) interpretation is crucial in the clinical ECG workflow because it is most likely associated with a disease that can cause major problems in the body. In this study, we developed an ECG-signal multi-classification model using deep learning.

Methods

We used a squeeze-and-excitation residual network (SE-ResNet), which is used for efficient image classification. Experiments were performed for seven different types of lead-II ECG data obtained from the Korea University Anam Hospital in South Korea. These seven types are normal sinus rhythm, atrial fibrillation, atrial flutter, sinus bradycardia, sinus tachycardia, premature ventricular contraction and first degree atrioventricular block. We compared SE-ResNet with a residual network (ResNet), as a baseline model.

Results

The SE-ResNet classifier with 152 layers achieved F1 scores of 97.05% for the seven-class classifications. Our model surpassed the baseline model, ResNet, by + 1.40% for the seven-class classifications.

Conclusion

For ECG-signal multi-classification, considering the F1 scores SE-ResNet might be better than the ResNet baseline model.

Background

The human heart has an electric transmission system that voluntarily generates regular electrical signals and transmits these signals to the entire heart. Heart disease takes the lives of many all over the world [1, 2, 3]. However, arrhythmia, i.e., irregular heartbeats caused by changes or dysfunctions of this system, has been unfamiliar to the general public.

Arrhythmia can generally be diagnosed using a measured electrocardiogram (ECG), which is a record of the electrical activity in the heart, obtained through electrodes located on the skin of the chest and limbs. An ECG usually refers to a 12-lead ECG, which gathers 12 different types of information from the heart. To precisely classify a 12-lead ECG signal, doctors examine the ECG data and diagnose specific arrhythmias based on their medical knowledge and extensive experience. Unfortunately, judgment errors are likely to occur during this process. Even an experienced specialist requires considerable time to

analyze the signals, and the accuracy may not be high [4, 5]. In addition, in the case of a Holter monitor, the cardiologist cannot see the entire signal, which is usually recorded over several days. Thus, many scholars have attempted to classify 12-lead ECG signals automatically and accurately.

Thus far, rule-based algorithms for ECG signal classification have been unsuitable for use in practice, owing to their poor performance. In addition, this classification has been approached using various machine-learning methods, e.g., logistic regression [6], support vector machines (SVMs) [7], random forests [8], and K-nearest neighbors [9, 10]. The deep-learning model is as a closed model for use in real hospitals because it exhibits much better ECG signal-classification performance than conventional classification algorithms and rule-based algorithms.

The most natural deep-learning research using ECG data involves creating a deep-learning model using 12-lead ECG information measured in a hospital. Smith et al. [11] found that the accuracy of a new deep-learning network using 12-lead ECG data was higher than that of a conventional algorithm, with 13 convolutional layers and 3 fully connected layers. However, in most cases, rather than using all the ECG information, scholars have approached ECG signal classification using the information from one specific lead; e.g., lead I or lead II. Lee et al. [12] used a residual network (ResNet) with six residual blocks and an Alex network to classify atrial fibrillation (Normal / Atrial Fibrillation), which provided accuracies of 99.9% and 99.7%, respectively.

Rajpurkar et al. [4] and Hannun et al. [5] showed that a deep-learning model exceeded average cardiologists in terms of ECG discrimination ability of 12 output rhythm classes (10 arrhythmias/Normal/Noise), using a 34-layer ResNet model. It is important to note that the data they collected were large-scale, obtained from patients in actual hospitals. Their model used 91,232 modified lead-II ECG records from 53,549 patients, recorded using Zio cardiac monitors.

Recently, beyond simple convolutional neural network (CNN) structures, attempts have been made to find a better ECG signal-classification structure by using structures that produce good results for image classification. Kim et al. [13] used the visual DenseNet architecture with 34 layers for two classifications (Normal / Abnormal), with lead-II ECG data measured in a hospital. This structure achieved an overall accuracy of 98.89% and an F1 score of 99.09%. Their results showed that a single-lead ECG, rather than the 12-lead ECGs measured in a hospital, was sufficient to distinguish between normal and abnormal.

In contrast to the methods mentioned thus far, some scholars have used short-term Fourier and wavelet transforms to convert ECG data into two-dimensional (frequency, time) data and used them as input for a deep neural network. Salem et al. [14] used the transformation “spectrogram” from a one-dimensional (1D) ECG signal from the MIT-BIH dataset and the European ST-T dataset to make 2D images. They also used a 161-layer DenseNet, pre-trained on millions of images, to extract abstract information and then applied an SVM for four-class classification (Normal Sinus/Atrial Fibrillation and Flutter/Ventricular Fibrillation/ST Segment Change). Their model’s accuracy and F1 score were 97.23% and 97.35%, respectively.

Rajput et al. [15] constructed an ECG-based heartbeat-classification model that consisted of preprocessing (filtering and segmentation), feature extraction (Morlet wavelet transform and short-term Fourier transform), and a densely connected network. Their model's F1 score was 83.4% (Normal Sinus / Atrial Fibrillation / Sinus Tachycardia / Sinus Bradycardia / Ventricular Bigeminy / Ventricular Trigeminy / Ventricular Tachycardia / Paroxysmal Supraventricular Tachycardia (PSVT) / Noise / Ventricular Ectopic Beats (VEB)).

Thus far, deep learning's approach to ECG classification is similar to that of image classification, with deep learning layer deepening and a complex structure. In this study, we also followed this trend to find a suitable structure, among those structures that were deeper and more complex but provided good results for existing image classification, for ECG signal multi-classification.

Methods

ECG dataset description

We constructed a large ECG-signal dataset that includes 28,308 lead-II ECGs collected from the Korea University Anam hospital in South Korea. The collected data are meaningful in that they are not refined, but they include various types of actual data measured in hospitals. The data consist of the following 7 categories:

- Normal sinus rhythm (Normal)
- Atrial fibrillation (AF)
- Atrial flutter (AFL)
- First degree atrioventricular block(FAB)
- Sinus bradycardia (SB)
- Sinus tachycardia (ST)
- Premature ventricular contraction (PVC)

Cardiologists in the Korea University Anam Hospital in South Korea annotated the labels for these 7 classes. In this study, our model was designed to classify seven rhythm classes (Normal / AF / AFL / FAB / SB / ST / PVC) from raw single-lead ECG data. The data ratios for each sector were 34.48% (Normal), 33.86% (AF), 6.17%(AFL), 6.90%(FAB), 6.87% (SB), 6.20% (ST), and 5.53% (PVC).

The ECG data were measured for 10 s at a frequency of 200 Hz. The data we used were based on lead-II ECG data taken from 12-lead ECG data. In addition, the range of data values was adjusted to enable the smooth learning of deep-learning models with min-max normalization.

In this study, we used a squeeze-and-excitation residual network (SE-ResNet), which is a ResNet with an added squeeze-and-excitation (SE) block, to create a model for ECG-signal multi-classification. SE-ResNet is considered to be one of the most popular of the many CNN architectures because of its high

performance on ImageNet for image classification. In addition, an SE network is easy to apply because it simply adds an SE block without changing the shape of the existing model. We used ResNet as the baseline model for our ECG-signal multi-classification model; ResNet is known as one of the best models for ECG-signal multi-classification [4, 5].

Specifically, we used a modified ResNet [16], which uses pre-activated weight layers, instead of the original ResNet [17], which uses post-activated, because the modified ResNet has better performance than the original. Through these processes, we confirmed that SE-ResNet was an efficient ECG signal classifier and minimized human intervention. In addition, we conducted experiments to observe the changes in the model's classification performance considering the number of layers, and we found an optimal model for classifying ECG-based heartbeats.

Classifier Model Architecture And Experiment

We developed an ECG-signal multi-classification model using SE-ResNet. SE-ResNet focuses on the interdependencies between the channels of its convolutional features, instead of investigating the spatial information. The SE block consists of a squeeze operation, which summarizes the overall information about each feature map, and an excitation operation, which scales the importance of each feature map. That is, the squeeze operation extracts only the important information from each channel using global average pooling, and the excitation operation computes the inter-channel dependencies using a fully connected layer and a nonlinear function. The main difference between the proposed network and the original SE-ResNet on ImageNet is that the proposed network uses 1D convolutions instead of 2D convolutions. We modified the original SE-ResNet by changing the input and output from $224 \times 224 \times 3$ and 1000 to $1 \times 2000 \times 1$ and seven (classes). To verify the performance of our model using SE-ResNet, we chose the ResNet used in [4, 5] as the baseline model.

We evaluated our model on the lead-II ECG signal dataset measured in the Korea University Anam Hospital; it consists of seven classes. We considered 28,308 10-s ECG signals. We first split our lead-II ECG dataset into two, a training dataset and a test dataset, in the ratio of 8:2. After this, we set aside the test dataset and chose 80% of the training dataset to be the actual training dataset and the remaining 20% to be the validation dataset. For dividing the dataset, we randomly selected each dataset; however, we fixed the random seed to compare the results. Thus, 64% of our total dataset was used to train the network, 16% to validate the model, and the remaining 20% to test the model. The number of training samples was 18,116. For the validation and test of the ECG signals, 4,530 and 5,662 samples were used, respectively.

As an optimizer, we selected the Adam optimizer presented by Kingma et al. [18] with an initial learning rate of 0.0001. The training process continued until the validation loss did not decrease for a certain step. Similar to other deep-learning classification models, we used categorical cross entropy as a loss function.

Under the above training setting, we investigated SE-ResNet and ResNet with 18/34/50/101/152 layers for seven-class classification. The test set was evaluated using the parameters in the validation set that exhibited the best performance. For the model evaluation, the accuracy and F1 score were used. The F1 score is the harmonic mean of the precision and recall. It is a more efficient criterion for model evaluation than accuracy, if the ratio between the data sectors is very different, e.g., the ECG signal dataset we cover in this study.

Results

Tables 1 and 2 summarize the performance of the seven-class classification models on the testing data using SE-ResNet and ResNet, respectively. For all sectors, our model has a higher F1 score than the baseline model, ResNet. The best result for the seven-class classification model using SE-ResNet was the 152-layer model with a 97.05% F1 score (Table 1). The best result for the seven-class classification model using ResNet was the 152-layer model with a 95.65% F1 score (Table 2).

From Tables 1, 2, we can see that our methodology using SE-ResNet outperforms the baseline model. Specifically, the F1 scores of the best SE-ResNet models for the seven-class ECG signal classifications were + 1.40% (difference between 97.05% for the 152-layer SE-ResNet and 95.65% for the 152-layer ResNet) higher than the baseline model.

When the data were analyzed, the point to note about Tables 1 is that the F1 scores of our model for AFL, PVC, SB and FAB were relatively lower than the F1 scores for Normal, AF and ST. Furthermore, to check the results in more detail, we selected the 152-layer SE-ResNet models with the highest F1 scores for the seven-class classifications. We calculated the confusion matrices of these models, and they are graphically shown in Fig. 1, which confirms the advantages and disadvantages of our models. As shown in Fig. 1, our model shows good overall performance for most sectors; however, it had difficulty distinguishing between AF and AFL, FAB and PVC, and FAB and SB, which explains the lower F1 scores for AFL, PVC, SB and FAB.

Discussion

We gathered a large lead-II ECG dataset from the Korea University Anam Hospital in South Korea and found a suitable ECG-signal multi-classification model for this dataset. Our SE-ResNet model surpassed the F1 scores for the baseline model, ResNet, by + 1.40% for the seven-class ECG signal classifications. Thus, we confirmed that SE-ResNet is a good model for ECG-signal multi-classification with a high accuracy and F1 score. These results indicate that, with only one ECG signal, instead of the 12-lead ECG measured by the hospital, our SE-ResNet multi-classification model could classify ECG signals sufficiently correctly. In addition, our model is expected to be a tool to provide arrhythmia patients and the general public with information about arrhythmia, and to enable doctors to immediately perform the necessary treatment in the medical field.

However, our classifier has several limitations. First, the F1 scores of AFL, PVC, SB and FAB were much lower than those of the other sectors (Normal / AF / ST) in our model. This was because of the model's inability to distinguish between AF and AFL, FAB and PVC, and FAB and SB. This lower accuracy might be due to the insufficient data for AFL, PVC, SB and FAB compared with Normal, AF and ST. In the future, we plan to upgrade our model to increase the F1 scores of AFL, PVC, SB and FAB.

Second, owing to the lack of ECG data, we created a classification model for only some often-observed arrhythmias. To expand this model to other arrhythmia conditions, we must accumulate additional arrhythmia data; e.g., junction rhythm, SVT, VT, Wenckebach, etc. presented in Rajpurkar et al. [4] and Hannun et al. [5].

Third, we built the model and conducted various experiments to determine the best F1 score and accuracy, without considering the model's capacity. To achieve higher accuracy, we followed the deep-learning trend of making deeper and more complicated networks. However, many studies in the real world must be carried out on computationally limited platforms. In the future, we will consider the model's file size and computation speed, as well as its accuracy and F1 score.

Finally, we could not explain why our model arrived at a specific decision for ECG classification. We must be able to fully understand how these decisions are being made so that we can trust the model's decision. We will consider this direction as a future research topic.

Conclusions

For ECG-signal multi-classification, SE-ResNet might be better than the ResNet baseline model, considering the F1 scores.

Abbreviations

SE-ResNet

Squeeze-and-Excitation Residual Network

SENet

Squeeze-and-Excitation Network

ResNet

Residual Network

ECG

Electrocardiogram

CNN

Convolutional Neural Network

RNN

Recurrent Neural Network

Normal

Normal Sinus Rhythm

AF

Atrial Fibrillation

AFL

Atrial Flutter

PVC

Premature Ventricular Contraction

SB

Sinus Bradycardia

ST

Sinus Tachycardia

FAB

First degree Atrioventricular Block

Declarations

Ethics approval and consent to participate

This retrospective study got approved by the Institutional Review Board of Korea University Anam Hospital. (2018AN0037). Informed consent was waived by the IRB given that data were de-identified.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by a Korea University Grant.

Contributions

JP, JK, SJ, YG, JC and HSS (corresponding author) have directly participated in the planning, execution and analysis of the study. They read and approved the final manuscript.

Acknowledgements

Not applicable

Availability of data and materials

The datasets supporting the conclusions of this article are available from the corresponding author upon reasonable request.

References

1. Roth GA, Huffman MD, Moran AE, Feigin V, Mensah GA, Naghavi M, Murray CJ. (2015). Global and regional patterns in cardiovascular mortality from 1990 to 2013. *Circulation*, 132(17), 1667–1678.
2. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, ... Forouzanfar MH. (2014). Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation*, 129(8), 837–847.
3. Korea S. (2017). Year 2016 statistics on causes of death in Korea. Sejong: Statistics Korea.
4. Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*.
5. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1), 65.
6. Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M. (2002). *Logistic regression*. New York: Springer-Verlag.
7. Cortes C, Vapnik V. (1995). Support-vector networks. *Machine learning*, 20(3), 273–297.
8. Breiman L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
9. Cover T, Hart P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21–27.
10. Escalona-Morán MA, Soriano MC, Fischer I, Mirasso CR. (2014). Electrocardiogram classification using reservoir computing with logistic regression. *IEEE Journal of Biomedical and Health Informatics*, 19(3), 892–898.
11. Smith SW, Rapin J, Li J, Fleureau Y, Fennell W, Walsh BM, ... Gardella C. (2019). A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation. *IJC Heart & Vasculture*, 25, 100423.
12. Lee KS, Jung S, Gil Y, Son HS. (2019). Atrial fibrillation classification based on convolutional neural networks. *BMC medical informatics and decision making*, 19(1), 206.
13. Kim J, Jung S, Park J, Han S, Arrhythmia detection model using modified DenseNet for comprehensible Grad-CAM visualization, submitted.
14. Salem M, Taheri S, Yuan JS. (2018, October). ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features. In 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS) (pp. 1–4). IEEE.

15. Rajput KS, Wibowo S, Hao C, Majmudar M. (2019). On Arrhythmia Detection by Deep Learning and Multidimensional Representation. arXiv preprint arXiv:1904.00138.
16. He K, Zhang X, Ren S, Sun J. (2016, October). Identity mappings in deep residual networks. In *European conference on computer vision* (pp. 630–645). Springer, Cham.
17. He K, Zhang X, Ren S, Sun J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
18. Kingma DP, Ba J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Tables

Table 1. Seven-class classification results for our model (SE-ResNet)

Layers in model	Accuracy	F1 Score							
		Total	Normal	AF	AFL	PVC	SB	ST	FAB
18	96.56%	96.57%	99.39%	99.04%	92.11%	88.67%	91.35%	92.31%	89.59%
34	96.50%	96.48%	99.58%	98.49%	88.95%	90.60%	91.37%	94.75%	89.66%
50	95.78%	95.78%	99.16%	97.97%	87.55%	89.04%	91.06%	93.51%	88.15%
101	96.68%	96.67%	99.39%	98.72%	90.98%	90.38%	92.20%	93.84%	90.35%
152	97.05%	97.05%	99.84%	99.12%	92.45%	90.57%	91.27%	94.66%	90.19%

Total : Frequency weighted average of F1 scores

Normal : Normal Sinus Rhythms

AF : Atrial Fibrillation

AFL : Atrial Flutter

PVC : Premature Ventricular Contraction

SB : Sinus Bradycardia

ST : Sinus Tachycardia

FAB : First degree atrioventricular block

Table 2. Seven-class classification results for the baseline model (ResNet)

Layers in model	Accuracy	F1 Score							
		Total	Normal	AF	AFL	PVC	SB	ST	FAB
18	95.21%	95.28%	99.39%	98.46%	85.49%	83.92%	89.18%	93.23%	85.32%
34	94.95%	95.05%	99.13%	99.08%	90.81%	80.25%	84.47%	91.85%	83.32%
50	95.23%	95.23%	99.29%	98.88%	86.32%	84.52%	88.83%	91.41%	83.69%
101	95.32%	95.35%	98.10%	98.32%	91.98%	84.79%	90.29%	91.96%	86.18%
152	95.73%	95.65%	98.88%	99.49%	91.75%	80.70%	88.03%	92.13%	86.19%

Total : Frequency weighted average of F1 scores

Normal : Normal Sinus Rhythms

AF : Atrial Fibrillation

AFL : Atrial Flutter

PVC : Premature Ventricular Contraction

SB : Sinus Bradycardia

ST : Sinus Tachycardia

FAB : First degree atrioventricular block

Figures

Figure 1

Seven-class classification confusion matrices using SE-ResNet152