

# Prediction of Atrial Fibrillation Cases: Convolutional Neural Networks Using the Output Texts of Electrocardiography

Tak-Sung Heo,<sup>1,2</sup> Chulho Kim,<sup>3</sup> Jong-Dae Kim,<sup>1,2</sup>  
Chan-Young Park,<sup>1,2</sup> and Yu-Seop Kim<sup>1,2\*</sup>

<sup>1</sup>Department of Convergence Software, Hallym University, Chuncheon-si, Gangwon-do 24252, Republic of Korea

<sup>2</sup>BIT Research Center, Hallym University, Chuncheon-si, Gangwon-do 24252, Republic of Korea

<sup>3</sup>Department of Neurology, Hallym University College of Medicine,  
Chuncheon-si, Gangwon-do 24253, Republic of Korea

(Received June 30, 2020; accepted November 24, 2020)

**Keywords:** atrial fibrillation, electrocardiogram, FastText, convolutional neural networks, prediction

Atrial fibrillation (AF) is the most common arrhythmia. Since AF can cause strokes if it lasts for a long time, it is important to detect AF in advance and receive treatment. Electrocardiography is usually used for AF diagnosis. Electrocardiography records the electrical activity of the patient's heart to obtain an electrocardiogram (ECG), which usually consists of waves and a commentary on them. The onset of AF occurrence or its likelihood is judged by a comprehensive analysis of an ECG, which requires considerable prior knowledge and clinical experience. In this study, to make this process simpler, the output text of ECGs is analyzed by deep learning to predict the possibility of future AF. The proposed model represents words as vectors using FastText and extracts features using one-dimensional convolutional neural networks (CNNs). The model also combines features using global average pooling (GAP) and is trained to calculate the probability of developing AF. In an experiment, the model showed 85.03% accuracy in predicting the presence or absence of AF. We thus demonstrated the possibility of predicting the occurrence of AF in advance using only text analysis without prior knowledge and clinical experience of AF.

## 1. Introduction

Atrial fibrillation (AF) is the most common persistent arrhythmia in which the heart beats irregularly.<sup>(1)</sup> The prevalence of AF is less than 1% for those under the age of 60, but it is known to increase rapidly after the age of 60.<sup>(2)</sup> Recently, the average age of AF patients has been gradually increasing as general health has improved. In addition, the treatment-related mortality rate has been reduced as a result of advances in treatment, although the prevalence of AF is expected to increase in the future.<sup>(3)</sup> The factors that induce AF are not only age but also hypertension, obesity, diabetes, excessive drinking, and smoking.<sup>(4)</sup>

Electrocardiography is usually used for AF diagnosis. An electrocardiograph is a device that records the electrical activity of a patient's heart.<sup>(5)</sup> The electrocardiograph obtains the electrical activity state of a patient's heart through analog waveform conversion and mechanically

---

\*Corresponding author: e-mail: yskim01@hallym.ac.kr  
<https://doi.org/10.18494/SAM.2021.3023>

analyzes these waveforms to automatically obtain an output text reading. The record extracted here is called an electrocardiogram (ECG). An ECG is composed of waves and an output text. The sensor of the electrocardiograph detects P, Q, R, S, and T waves. In particular, Q, R, and S waves (QRS) are the most characteristic waves, with higher amplitudes than P and T waves. In some studies, waves other than QRS are treated as noise and deleted, and the presence or absence of AF is determined by using QRS.<sup>(6)</sup> The output text of an ECG is the result of automatic analysis of these waves by the machine. The ECG's output text includes not only the contents of the P, Q, R, S, and T waves, but also the patient's condition through their interpretation. The clinician interprets the waves and output text of an ECG comprehensively to determine whether AF has occurred or the likelihood of its onset. However, in order to interpret the waves and the output text of the ECG, much prior knowledge and clinical experience of AF are required, and the interpretation requires a lot of time.

To avoid these difficulties, in this study, we attempt to predict the probability of AF in the future by automatically analyzing the output text of ECGs by a deep learning model. First, words appearing in the output text of ECGs are transformed into the input of a deep learning model using FastText, with words represented as vectors.<sup>(7)</sup> In order to represent the combined meaning using neighboring words, one-dimensional convolutional neural networks (CNNs), which show good performance in text processing, are used.<sup>(8,9)</sup> Then, global average pooling (GAP) is used to extract representative values of the combined features through the CNNs.<sup>(10,11)</sup> Finally, the probability of AF is predicted through a softmax layer. Through this process, the proposed model can predict the probability of AF in the future by analyzing only the output text of the ECG without prior knowledge of AF.

Section 2 explains text classification by deep learning, which is performed in this study. Section 3 gives the background knowledge of AF, Sect. 4 explains the output text of ECGs and the preprocessing, and Sect. 5 describes the structure of the proposed model. Section 6 outlines the parameters and environments used in the experiment, shows the results of the experiment, and discusses the results. Finally, Sect. 7 provides a summary of this study and the direction of future works.

## 2. Related Works

The prediction of AF occurrence using ECG texts can be regarded as a text classification problem; such problems have attracted much attention in the field of natural language processing. Models based on recurrent neural networks (RNNs), such as long short-term memory (LSTM) networks, which show good performance for sequential data, have mainly been used in the text classification problem. However, models based on CNNs, which are much faster and have high accuracy, have recently been widely used.<sup>(8,9,12)</sup>

Kim<sup>(8)</sup> used multichannel CNNs for text classification. Words appearing in sentences were represented as vectors using pretrained Word2Vec from Google News. Word2Vec is a word embedding technique that learns by using information from nearby words to represent the meaning of a word as a vector.<sup>(13,14)</sup> These vectors are input to CNNs with filters of various sizes. Global max pooling (GMP) is used to extract the largest value among the output of the

features from each CNN. Three features extracted from GMP are concatenated to one feature, are input to the fully connected layer by applying a dropout, and are finally input to the softmax layer.

Zhou *et al.*<sup>(12)</sup> used a CNN and LSTM for text classification. Words appearing in sentences were represented as vectors using pretrained Word2Vec from Google News. Vectors converted through Word2Vec were input to a single CNN with filter size 3, and features extracted from the CNN were input to the LSTM. Zhou *et al.* did not use pooling for features extracted from a CNN. This is because when applying pooling, the sequential information of features is lost. Among the values obtained from LSTM, the last hidden state with all the information of the sentence is finally input to the softmax layer.

### 3. AF

AF is the most common arrhythmia, with the higher the age, the higher its prevalence. AF is classified by the European Society of Cardiology into the following five types, depending on the duration and pattern of expression: first diagnosed AF, paroxysmal AF, persistent AF, long-standing persistent AF, and permanent AF.<sup>(15)</sup> First diagnosed AF refers to the first case of AF diagnosis, regardless of AF-related symptoms, duration, or severity. Paroxysmal AF refers to a case of conversion to a normal heartbeat without special treatment within 48 h. Paroxysmal AF can sometimes last up to 7 days. However, this is rarely converted to a normal heartbeat without treatment after 48 h. Persistent AF refers to a case that lasts more than 7 days. This may be converted to a normal heartbeat by a drug or direct current cardioversion. Long-standing persistent AF refers to AF that lasts more than a year when the use of a rhythm-control strategy is decided. Permanent AF refers to cases where the patient's condition does not improve despite treatment and management. AF initially exhibits the form of paroxysmal AF, but it progresses to persistent AF and permanent AF over time, and AF is known to be difficult to cure if it has progressed for a long time.<sup>(15,16)</sup>

Symptoms of AF include cardiac hyperactivity, fainting, dizziness, shortness of breath, and chest pain caused by an irregular pulse, but some patients may not have these symptoms.<sup>(17)</sup> AF can cause systemic embolism due to atrial expansion and atrial thrombi, and stroke can occur, especially if a thrombus blocks the brain blood vessels.<sup>(18)</sup> The risk of stroke is 4–5 times higher in AF patients than in normal people.<sup>(19)</sup> The closing of blood vessels in the brain due to blood clots is called ischemic stroke. The five major Trial of Org 10172 in Acute Stroke Treatment (TOAST) classifications are used when considering the cause of ischemic stroke.<sup>(20)</sup> These are large-artery atherosclerosis, small vessel occlusion, cardioembolism, other determined etiology, and undetermined etiology. In particular, cardioembolism due to AF is one of the causes of ischemic stroke, which makes it very difficult to treat when cerebral infarction occurs. Therefore, it is very important to detect AF in advance to reduce the above risk.

### 4. Dataset

The ECG dataset used in this study was collected by Hallym University Chuncheon Sacred Heart Hospital<sup>(21)</sup> from February 2010 to October 2019. This dataset mainly consists of a record

of patients with a history of hospitalization for acute ischemic stroke. The number of output texts of the ECG dataset is 10359, with 5846 AF cases and 4513 normal cases. To predict the probability of AF in the future, AF patients are assigned a label of 0 regardless of whether or not AF has actually occurred, and a label of 1 is given to normal patients.

Figure 1 shows an ECG generated by electrocardiography. The waves in Fig. 1 show the electrical activity of the patient's heart as graphs, and the output text above shows the result of interpreting the waves automatically by electrocardiography. If AF appears explicitly in the output text, it can be concluded that the patient has AF, but there are cases where AF does not appear in the output text despite the onset of AF in the patient.

#### 4.1 Data preprocessing

ECG output text is generated by the electrocardiography, so there are many duplicated expressions that can cause overfitting during the training. Therefore, in this study, we eliminate all but one of the duplicate expressions, convert uppercase letters to lowercase letters, and place spaces before and after special characters such as punctuation marks. Also, the words 'Atrial', 'Fibrillation', and 'Flutter', which can determine AF reliably, are removed from the sentences in order to make it possible to learn AF even when the words do not appear. Lastly, 'Abnormal electrocardiography' and 'Normal electrocardiography' appearing in the last part of the output text of the ECG are common to all texts, so we remove them.

Table 1 shows examples of data preprocessing. The first example shows the result of not only deleting 'Atrial fibrillation' and 'Abnormal electrocardiography' but also converting uppercase letters

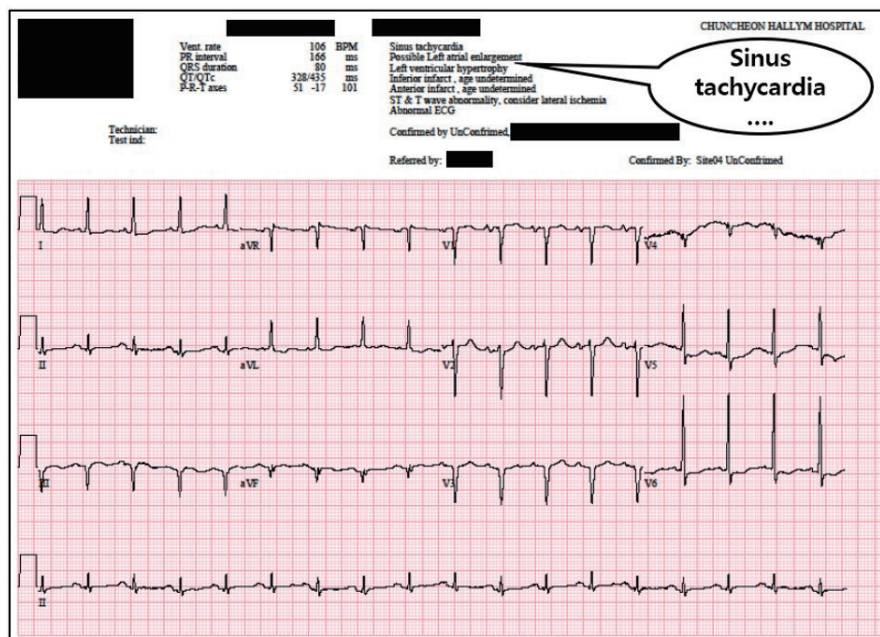


Fig. 1. (Color online) Example of wave and output text in ECG.

Table 1  
Example of data preprocessing.

Preprocessing	Name	AF	Text
Before	A	0	Atrial fibrillation Low Voltage QRS Right bundle branch block
After	A	0	Abnormal electrocardiography low voltage qrs right bundle branch block
Before	B	1	Sinus tachycardia with frequent premature ventricular complexes Left ventricular hypertrophy with repolarization abnormality Abnormal electrocardiography
After	B	1	sinus tachycardia with frequent premature ventricular complexes left ventricular hypertrophy with repolarization abnormality

letters to lowercase letters. The second example shows the result of removing ‘Abnormal electrocardiography’ and converting uppercase letters to lowercase letters. After removing duplicates, there are a total of 3543 texts, of which 2055 have label 0 and 1488 have label 1.

## 5. Methodology

Figure 2 is a schematic of the methodology proposed in this paper, which is conducted in the order of data collection, preprocessing, FastText, CNN, GAP, and softmax. Words are represented as vectors through FastText after preprocessing the output text of the ECG. These word vectors are input to the CNN and several feature maps are generated. After that, GAP is used for various feature maps. Finally, the probability of AF in the future is calculated in the softmax layer.

### 5.1 FastText

In the field of natural language processing, word embedding is used to represent the meanings of words. Word embedding converts a word into a meaningful dense vector using surrounding words so that the computer can understand the language of human beings.<sup>(13)</sup> This is based on the assumption that words appearing in a similar context have similar meanings, which is called the distributional hypothesis. Through word embedding, words can be vectorized to measure the similarity between them. In this paper, FastText, which is an extended model of Word2Vec, is used as the word embedding method.<sup>(7)</sup>

Word2Vec maximizes the inner product value of the vector corresponding to the central word and the vectors corresponding to the surrounding words while sliding the corpus by increments equal to the window size. At the same time, the Word2Vec model is continuously updated to minimize the inner product value with the vectors corresponding to words that do not belong to the window slide.<sup>(13,14)</sup> Word2Vec has good performance, but it has limitations because it involves word-based learning. The limitations of this method are that it does not reflect the morphological characteristics of the language, cannot solve the out-of-vocabulary problem, and it is difficult to represent meaning-appropriate embedding values for rare words.<sup>(7)</sup>

FastText is a word embedding technique designed to partially address these limitations.<sup>(7)</sup> FastText has a similar mechanism to Word2Vec, but Word2Vec learns a word unit, while

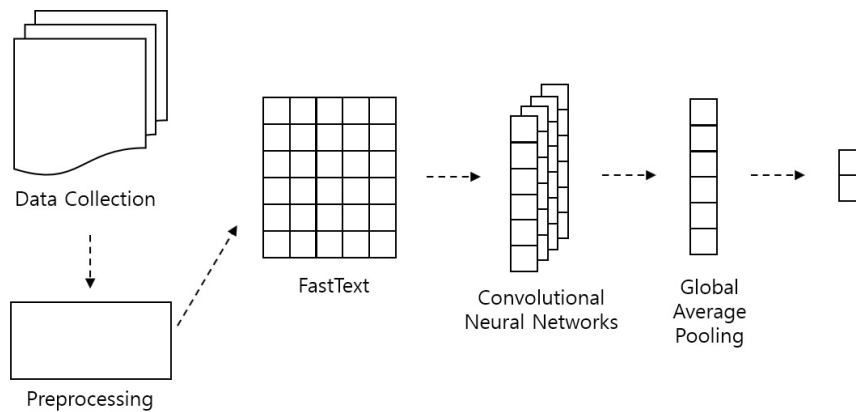


Fig. 2. Schematization of research methodology.

FastText learns by considering a word and its subwords. A subword in FastText is the  $n$ -gram of letters in one word, and the number of separate subwords depends on the value of  $n$  that is set. For example, if  $n$  is set to 5 for the word ‘speaking’, FastText learns by separating it into ‘speak’, ‘peak’, ‘eakin’, and ‘aking’. In addition, by learning not only subwords but also word units, the word ‘speaking’ is represented in vectors similar to subwords. Considering this point, morphologically similar words are represented as similar vectors, which enables FastText to cope better with out-of-vocabulary problems than Word2Vec. In addition, FastText has a large number of words that can be referenced by considering subwords, so embedding values of rare words are represented more appropriately with Word2Vec.<sup>(7)</sup>

In this study, we use a pretrained model of FastText called BioWordVec.<sup>(22)</sup> BioWordVec is a model that learns the texts of biomedical databases PubMed and MIMIC-III with FastText.

## 5.2 CNNs

CNNs perform well not only in image processing but also in text processing.<sup>(8,9)</sup> When applying a CNN to text, the CNN’s inputs use values in which each word is represented as a vector. That is, if the number of words in one sentence is  $N$  and the number of dimensions in which each word is represented as a vector is  $M$ , the size of the CNN’s input value is  $N \times M$ . A one-dimensional CNN uses all the vector values in each row and can create combined meanings of different numbers of words by varying the filter size. At this time, the filter size is used together with the stride that serves to slide the filter. If the stride is 1 and the filter size is  $n$ , the combined meaning is extracted by an  $n$ -gram method defined in computational linguistics.<sup>(9)</sup> Thus, the feature extracted from the CNN is called the feature map, and the size of the feature map is as follows.

$$feature\ map = \left\lfloor \frac{Number\ of\ words - filter\ size}{Stride} - 1 \right\rfloor \quad (1)$$

As in Eq. (1), the size of the feature map depends on the filter size and stride. However, in the case of using a CNN in this way, information at the edge of the input value is lost. To solve this problem, there is a technique called padding, which fills the edges with zeros, enabling the information at the edge to be preserved. The size of the feature map including padding is as follows.

$$feature\ map_{padding} = \left\lfloor \frac{Number\ of\ words - filter\ size + 2 \times P}{Stride} - 1 \right\rfloor \quad (2)$$

In Eq. (2),  $P$  is the width of the padding, and the size of the feature map is equal to the number of words regardless of the filter size when the stride is 1. Through this process, one feature map is extracted. However, using only one feature map does not help significantly in improving the performance because the amount of learning is insufficient. To solve this problem, we use multiple filters, which increases the amount of learning using multiple CNNs. As a result, the final feature map has a size of  $feature\ map_{padding} \times number\ of\ filters$ .

Figure 3 shows an example of a one-dimensional CNN to which padding, a filter size of three, a stride of one, and four filters are applied. The parameters of the CNN used in this study will be covered in detail in Sect. 6.

### 5.3 Prediction of AF

In general, the values extracted from a CNN derive its output through a fully connected layer to optimize the parameters. However, if a fully connected layer is used, a long time is required for learning, which can cause overfitting.<sup>(10,11)</sup> To address this problem, we use GAP instead of a fully connected layer. GAP is a technique that derives the average value from all feature maps extracted from a CNN. GAP can avoid overfitting because there is no need to optimize the parameters owing to its simple calculation method.<sup>(10,11)</sup> The features extracted through GAP are finally entered into the softmax layer to predict the probability of AF occurrence. The

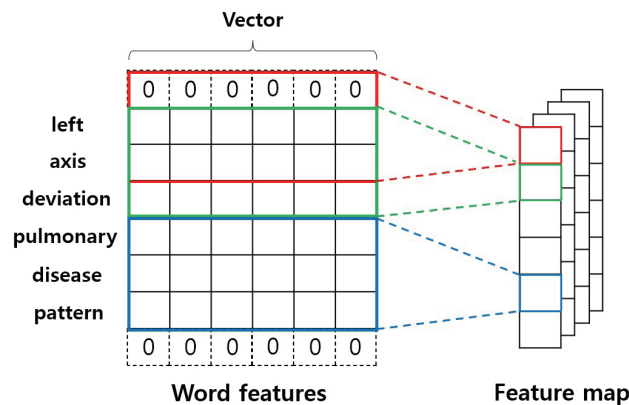


Fig. 3. (Color online) Example of one-dimensional CNNs.

softmax is a layer that normalizes the input value to a value between 0 and 1 using the following formula.

$$P_i = \frac{\exp(z_i)}{\sum_{j=1}^n \exp(z_j)}. \quad (3)$$

In Eq. (3),  $z_i$  is the value of the corresponding class and  $n$  is the total number of classes. The number of values output by softmax is equal to the number of classes to be classified, and the sum of these output values is always 1. In this study, since there are two classes to classify, the output values of softmax are the probability of AF occurring and the probability of no AF occurring. That is, according to the purpose of the study, the probability of AF occurring can be predicted through the softmax layer.

## 6. Experiment

In this study, an ECG dataset is collected for the experiment, and preprocessing is performed on the data. Patients with AF are given a label of 0 regardless of whether or not AF has actually occurred. In addition, the text of the applicants who showed no AF occurring is given a label of 1. The entire data is divided in the ratio of training set:validation set:test set = 8:1:1 to conduct the experiment.

### 6.1 Parameters of CNN

Table 2 shows the hyperparameters of the CNN used in this study. By setting the filter size to 3 and the stride to 1, the combined meaning of three words is extracted by the n-gram method. A total of 256 filters are used to increase the amount of learning, and padding is used to prevent the loss of edge information. In addition, ReLU is used as the activation function of the one-dimensional CNN.

### 6.2 Trigram of ECG dataset

Table 3 shows the eight trigrams with the highest frequency among the collected data, where Freq means frequency and Rate means the relative frequency. In Table 3, ('left', 'ventricular', 'hypertrophy') appears with high frequency in both label 0 and label 1. Thus, some trigrams and

Table 2  
Parameters of CNN.

Description	Hyperparameter
Filter size	3
Stride	1
Number of filters	256
Padding	Used
Activation function	ReLU



Table 3  
Trigram of ECG dataset.

Label 0	Freq	Rate	Label 1	Freq	Rate
('t', 'wave', 'abnormality')	1012	0.028375	('t', 'wave', 'abnormality')	474	0.019169
('wave', 'abnormality', ',')	782	0.021926	('normal', 'sinus', 'rhythm')	444	0.017956
('abnormality', ',', 'consider')	704	0.019739	(';', 'age', 'undetermined')	443	0.017916
('st', 'and', 't')	624	0.017496	('left', 'ventricular', 'hypertrophy')	443	0.017916
('and', 't', 'wave')	609	0.017076	('infarct', ',', 'age')	442	0.017875
('left', 'ventricular', 'hypertrophy')	577	0.016178	('sinus', 'rhythm', 'with')	341	0.013791
('infarct', ',', 'age')	546	0.015309	('wave', 'abnormality', ',')	337	0.013629
(';', 'age', 'undetermined')	546	0.015309	('abnormality', ',', 'consider')	336	0.013588

words appear to a similar degree in both labels. On the other hand, the words 'sinus' and 'rhythm' appear frequently in the label 1 non-AF dataset. These words can be seen as characteristic words for expressing label 1. Trigrams that are concentrated only on one label can help with prediction, but the number of such trigrams is very limited.

### 6.3 Prediction of AF onset

Table 4 shows the performance evaluation results for the model predicting the onset of AF. The numbers of test data in this experiment are 215 and 139 for labels 0 and 1, respectively. The results of the experiment shown in Table 4 are the average values of 10 experiments. The precision and recall values of labels 0 and 1 show significant differences because they do not have equal data distributions. In this experiment, the macro-averaged precision is 0.8473, the macro-averaged recall is 0.8637, the macro-averaged F1-score is 0.848, and the model accuracy is 85.03%.

Table 5 is the confusion matrix of Table 4. In Table 5, N is the total number of data, Actual 0 and Actual 1 are the actual labels, and Predicted 0 and Predicted 1 are the labels predicted using the model.

### 6.4 Results and discussion

In this study, in addition to the simple judgment of the presence or absence of AF, an objective is to calculate the probability of AF in the future. To this end, we aim to determine the probability of AF onset using the output of the softmax layer.

The six texts shown in Table 6 are comments often generated in an ECG. Texts No. 1 and No. 2 are comments with 'Atrial fibrillation' deleted during data preprocessing, and the probability of AF is the probability that it occurs in the future. Text No. 3 is the text of a patient who was later identified as having AF, which was generated before AF appeared. Texts No. 4, No. 5, and No. 6 are the texts of patients determined to be normal.

Texts No. 1 and No. 2 have very high probabilities of AF of 99.75 and 97.3%, respectively, even though 'Atrial fibrillation' has been deleted. The 'premature ventricular or aberrantly conducted' and 'right bundle branch block' appearing in their texts are features that are highly likely to cause AF.<sup>(23–25)</sup> Text No. 3, which is a text generated before AF appeared, has a 50.91%

Table 4

Performance evaluation when judged by the presence or absence of AF.

Label	Number of test data	Precision	Recall	F1-score	Accuracy
0	215	0.9436	0.8014	0.8667	85.03%
1	139	0.7509	0.9259	0.8292	

Table 5

Confusion matrix for the presence or absence of AF.

$N = 354$	Predicted 0	Predicted 1
Actual 0	172.3	42.7
Actual 1	10.3	128.7

Table 6

Prediction results for probability of future AF occurrence.

No.	Text	Label	Probability of AF (%)
1	with premature ventricular or aberrantly conducted complexes junctional st depression , probably normal	0	99.75
2	low voltage qrs right bundle branch block	0	97.3
3	marked sinus bradycardia voltage criteria for left ventricular hypertrophy nonspecific st and t wave abnormality	0	50.91
4	normal sinus rhythm voltage criteria for left ventricular hypertrophy t wave abnormality, consider lateral ischemia prolonged qt	1	41.98
5	normal sinus rhythm left axis deviation minimal voltage criteria for left ventricular hypertrophy, may be normal variant inferior infarct , age undetermined	1	7.7
6	sinus tachycardia with frequent premature ventricular complexes left ventricular hypertrophy with repolarization abnormality	1	16.95

probability of AF, which is lower than those of texts No. 1 and No. 2. In addition, it can be seen that texts No. 4, No. 5, and No. 6, which are determined as normal, have a lower probability of AF than those of No. 1, No. 2, and No. 3.

In this study, the probability of AF in the future is predicted through a CNN by a method that uses information from adjacent words. However, a shortcoming of this study exists. Text No. 4 was automatically generated as a result of the test for patients who did not have AF. However, the probability of AF occurrence of the text is not much different from the text of the patient with AF. This is because text No. 4 contains ('t', 'wave', 'abnormality'), ('wave', 'abnormality', ','), and ('abnormality', ',', 'consider'), which appear frequently in label 0 as shown in Sect. 6.2. This can be seen as a shortcoming caused by using only the information of adjacent words. In order to compensate for this shortcoming, it is necessary to find a method of determining AF by using not only the information of adjacent words but also the entire sentence.

## 7. Conclusions

In this study, we attempted to predict the probability of AF by using texts automatically generated by an ECG. For the purpose of learning, classified texts were collected on the basis of the occurrence of AF. The model proposed in this study is composed of FastText, CNN, GAP, and a softmax layer. This model can predict the presence or absence of AF with an

accuracy of up to 85.03%. At the same time, the probability of future AF is given. The predicted probabilities calculated using this model can automatically be transmitted to specialists to enable a more accurate diagnosis.

However, the output text of the ECG used in this study is automatically generated by a machine, so it lacks diversity. Therefore, in order to develop a more precise model, additional data that can be used together with existing data must be discovered. In addition, we hope to develop a model with higher performance by applying more diverse methodologies.

### Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2019R1A2C2006010 and NRF-2019R1G1A1097707).

### References

- 1 M. Rienstra, S. A. Lubitz, S. Mahida, J. W. Magnani, J. D. Fontes, M. F. Sinner, I. C. V. Gelder, P. T. Ellinor, and E. J. Benjamin: *Circulation* **125** (2012) 2933. <https://doi.org/10.1161/circulationaha.111.069450>
- 2 P. A. Wolf, R. D. Abbott, and W. B. Kannel: *Stroke* **22** (1991) 983. <https://doi.org/10.1161/01.str.22.8.983>
- 3 A. S. Go, E. M. Hylek, K. A. Phillips, Y. Chang, L. E. Henault, J. V. Selby, and D. E. Singer: *Jama* **285** (2001) 2370. <https://doi.org/10.1001/jama.285.18.2370>
- 4 C. J. Rodriguez, E. Z. Soliman, A. Alonso, K. Swett, P. M. Okin, D. C. Goff Jr, and S. R. Heckbert: *Ann. Epidemiol.* **25** (2015) 71. <https://doi.org/10.1016/j.annepidem.2014.11.024>
- 5 D. M. Mirvis and A. L. Goldberger: *Electrocardiography*, E. Braunwald, D. P. Zipes, and P. Libby, Eds. (W. B. Saunders, Philadelphia, 2001) Chap. 13.
- 6 B. Subramanian: *Biomed. Res.* **28** (2017) 3187. <https://www.biomedres.info/biomedical-research/ecg-signal-classification-and-parameter-estimation-using-multiwavelet-transform.html>
- 7 A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov: arXiv preprint (2016) arXiv:1612.03651. <https://arxiv.org/pdf/1612.03651.pdf>
- 8 Y. Kim: Proc. 2014 Empirical Methods in Natural Language Processing Conf. (EMNLP, 2014) 1746–1751. <https://doi.org/10.3115/v1/d14-1181>
- 9 N. Kalchbrenner, E. Grefenstette, and P. Blunsom: arXiv preprint (2014) arXiv:1404.2188. <https://arxiv.org/pdf/1404.2188.pdf>
- 10 M. Lin, Q. Chen, and S. Yan: arXiv preprint (2013) arXiv:1312.4400. <https://arxiv.org/pdf/1312.4400.pdf>
- 11 B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba: Proc. 2016 IEEE Computer Vision and Pattern Recognition Conf. (IEEE, 2016) 2921–2929. <https://doi.org/10.1109/cvpr.2016.319>
- 12 C. Zhou, C. Sun, Z. Liu, and F. Lau: arXiv preprint (2015) arXiv:1511.08630. <https://arxiv.org/pdf/1511.08630>
- 13 T. Mikolov, K. Chen, G. Corrado, and J. Dean: arXiv preprint (2013) arXiv:1301.3781. <https://arxiv.org/pdf/1301.3781.pdf>
- 14 T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean: Advances in Neural Information Processing Systems Conf. (NeurIPS, 2013) 3111–3119. <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>
- 15 P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casadei, M. Castella, H. C. Diener, H. Heidbuchel, J. Hendriks, G. Hindricks, A. S. Manolis, J. Oldgren, B. A. Popescu, U. Schotten, B. V. Putte, and G. Hindricks: *Eur. J. Cardiothorac. Surg.* **50** (2016) e1. <https://doi.org/10.5603/kp.2016.0172>
- 16 W. S. Aronow: *Cardiol. Rev.* **16** (2008) 181. <https://doi.org/10.1097/crd.0b013e31816de1e3>
- 17 Y. Xia, N. Wulan, K. Wang, and H. Zhang: *Comput. Biol. Med.* **93** (2018) 84. <https://doi.org/10.1016/j.combiomed.2017.12.007>
- 18 A. John Camm: *Clin. Cardiol.* **35** (2012) S1. <https://doi.org/10.1002/clc.21961>
- 19 H. J. Lin, P. A. Wolf, M. Kelly-Hayes, A. S. Beiser, C. S. Kase, E. J. Benjamin, and R. B. D’Agostino: *Stroke* **27** (1996) 1760. <https://doi.org/10.1161/01.str.27.10.1760>
- 20 H. P. Adams, B. H Bendixen, L. J. Kappelle, J. Biller, B. B. Love, D. L. Gordon, and E. E. Marsh: *Stroke* **24** (1993) 35. <https://doi.org/10.1161/01.str.24.1.35>
- 21 Hallym University Chuncheon Sacred Heart Hospital: <https://chuncheon.hallym.or.kr/> (accessed May 2020).

- 22 BioWordVec: <https://github.com/ncbi-nlp/BioSentVec/> (accessed May 2020).
- 23 S. H. Lau, A. N. Damato, W. D. Berkowitz, and R. D. Patton: *Circulation* **40** (1969) 71. <https://doi.org/10.1161/01.cir.40.1.71>
- 24 S. Patanè, F. Marte, and S. A. Villari: *Int. J. Cardiol.* **138** (2010) e42. <https://doi.org/10.1016/j.ijcard.2008.06.061>
- 25 V. Carbone, S. Carerj, and M. P. Calabrò: *J. Electrocardiol.* **37** (2004) 67. <https://doi.org/10.1016/j.jelectrocard.2003.11.002>

## About the Authors



**Tak-Sung Heo** received his B.S. degree in biology from Hallym University. He is pursuing a master's degree in convergence software at Hallym University. His research interests are in deep learning, data mining, and natural language processing. (gjxkrtjd221@gmail.com)



**Chulho Kim** received his Ph.D. degree from the College of Medicine, Gangwon National University. He is currently an associate professor in the Department of Neurology, Chuncheon Sacred Hospital. His research interests are in stroke epidemiology, acute stroke care, and biomedical informatics. (gumdol52@naver.com)



**Jong-Dae Kim** received his M.S. and Ph.D. degrees in electrical and electronics engineering from Korea Advanced Institute of Science and Technology, Seoul, Korea, in 1984 and 1990, respectively. He worked for Samsung Electronics from 1988 to 2000 as an electronic engineer. He is currently a professor in the School of Software at Hallym University. His recent interests are in biomedical systems and bioinformatics. (kimjd@hallym.ac.kr)



**Chan-Young Park** received his B.S. degree from Seoul National University and his M.S. and Ph.D. degrees from Korea Advanced Institute of Science and Technology in 1989 and 1995, respectively. From 1991 to 1999, he worked at Samsung Electronics. He is currently a professor in the School of Software at Hallym University. His research interests are in bio-IT convergence, intelligent transportation systems, and sensors. (cypark@hallym.ac.kr)



**Yu-Seop Kim** received his Ph.D. degree in computer engineering from Seoul National University. He is currently a professor in the School of Software at Hallym University. His research interests are in bioinformatics, computational intelligence, and natural language processing. (yskim01@hallym.ac.kr)