Classification of Heart Diseases Based On ECG Signals Using Long Short-Term Memory

Ming Liu and Younghoon Kim†

***Abstract*— Heart disease classification based on electrocardiogram(ECG) signal has become a priority topic in the diagnosis of heart diseases because it can be obtained with a simple diagnostic tool of low cost. Since early detection of heart disease can enable us to ease the treatment as well as save people's lives, accurate detection of heart disease using ECG is very important. In this paper, we propose a classification method of heart diseases based on ECG by adopting a machine learning method, called Long Short-Term Memory (LSTM), which is a state-of-the-art technique analyzing time series sequences in deep learning. As suitable data preprocessing, we also utilize symbolic aggregate approximation (SAX) to improve the accuracy. Our experiment results show that our approach not only achieves significantly better accuracy but also classifies heart diseases correctly in smaller response time than baseline techniques.**

I. Introduction

Heart is the most sophisticated organ in our body which pumps blood of the circulatory system. There are various forms of heart disease, for example, arrhythmias, prolapsed mitral valve, coronary artery disease, congestive heart failure, congenital heart disease, and so on [1]. Traditionally, heart diseases have been diagnosed by various methods such as blood test and chest X-ray. Among such tests, Electrocardiogram (ECG) has been widely used and known as the best method for heart disease detection today. ECG assesses the heart rate and rhythm by recording electrical currents generated in muscles during its contraction representing neuromuscular activities over a period. Since the investigation of ECG signal is an easy and non-invasive tool measuring the electrical signal by simply placing electrodes on the skin, accurate detection of abnormal heartbeat based on ECG may help early detection of heart disease.

Heart disease diagnosis using automatic classification techniques has been extensively studied in both medical/health and data mining research communities. H. Khorrami et al. [2] evaluated the performance of well-known supervised machine learning techniques, such as multilayer perceptron and Support Vector Machine (SVM) for heart disease classification. By using features extracted by discrete wavelet and discrete cosine transforms, they found that the performance of heart disease classification largely depends on the features used in the algorithm. M. Korürek et al. [3] presented ECG classification using radial basis function neural network (RBFNN) with adopting particle swarm optimization (PSO). R. Saini et al. [4] improved the classification efficiency using k-nearest neighborhood (KNN) classifier based on the wavelet features. E. D. Übeyli [5] found that the Lyapunov exponents, wavelet coefficients and the power levels of power spectral density values obtained by the eigenvector methods are very suitable for representing ECG signals. J. Park et al. [6] proposed a cascade classification method of arrhythmia detection by using simple features obtained by heuristic peak detection. For further related work, a survey in [7] provides a comprehensive roadmap of heart disease classification with a list of machine learning techniques and features used in those methods.

Most of the traditional work including those mentioned above utilizing feature extraction designed for ECG signal mainly resorting to heuristic heartbeat segmentation. Recently, the latest advances in deep learning enable us to achieve high accuracy of classification in many applications such as speech and image recognition with relying on feature extraction customized for each application.

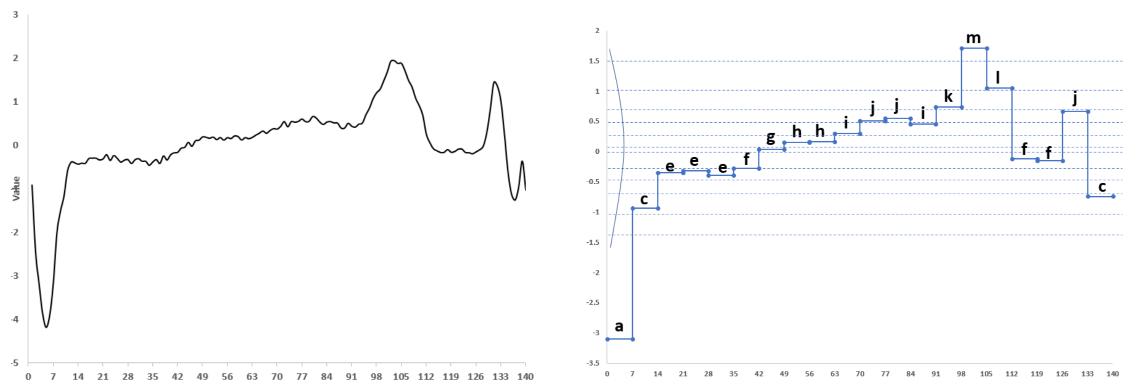
In this paper, we propose a method of classifying heart disease with ECG signals using Long Short-Term Memory (LSTM), which is developed to encode a sequence input of variable lengths for supervised learning. We also employ a fast and efficient quantization technique called Symbolic Aggregate approXimation (SAX) [8] and show that it is very suitable for numeric sequences with periodic pulses like ECG records by evaluating the performance of simple usage of LSTM with applying SAX. Our experiments confirm that using SAX and LSTM significantly improves the performance of heart disease classification as well as the execution time of training, and also suggest the optimal setting of hyper-parameters for using SAX.

II. Data Preprocessing using SAX

To classify the ECG signals, we first preprocess the original ECG signals and transform it to the sequences that have strongly featured. Then we use our proposed neural network model to identify heart diseases with the preprocessed data as input. For the preprocessing, we suggest using Symbolic Aggregate approXimation (SAX) which was introduced in [8] to represent a numeric sequence with a sequence of discrete symbols aiming the dimension reduction. In many applications handling numeric sequences, using SAX achieves the performance as good as other well-known discretization methods such as discrete wavelet transform and discrete Fourier transform while requiring less storage space [9]. SAX follows two steps; The first step is to reduce the dimension using Piecewise Aggregate Approximation

†Corresponding author. E-mail: *nongaussian@hanyang.ac.kr*

M. Liu and Y. Kim are with the Department of Computer Science, Hanyang University at Ansan, Korea.



(a) Original ECG Signals **(b)** Processing with SAX

Fig. 1: SAX Algorithm

(PAA) [10] and the second step is to convert a PAA sequence into a series of letters.

In the first step of PAA dimension reduction, we divide the long sequences into smaller equally spaced segments or bins and compute the average of each segment. Note that we normalize the input data to have a zero mean and a standard deviation of one before it converts it to the PAA representation. In the second step, we perform discretization by replacing the average value of each bin with a symbol. The numerical average is mapped into a symbol according to the range on which it falls, where the ranges are determined to slice the continuous space to have equiprobability by assuming normal distribution [11] of $N(0, 1)$, since we have normalized the input data to follow the distribution. Table 1 shows the splitting values for determining the interval. α and δ represent the number of ranges to split with and the splitting values between two adjacent intervals. For example, to discretize the data using 4 symbols, we use 3 splitting values as shown in the column with α=4.

δ \ α	4	5	6	7	8	9
δ_1	-0.67	-0.84	-0.97	-1.07	-1.15	-1.22
δ_2	0	-0.25	-0.43	-0.57	-0.67	-0.76
δ_3	0.67	0.25	0	-0.18	-0.32	-0.43
δ_4		0.84	0.43	0.18	0	-0.14
δ_5			0.97	0.57	0.32	0.14
δ_6				1.07	0.67	0.43

TABLE I: Breakpoints lookup table.

Figure 1 presents an example of a raw ECG signal and the discretized sequence of symbols using SAX. We reduce the dimension (i.e., the length of input sequence) from 140 to 20. Each equally divided bins then encoded into one of 13 symbols. The splitting values we have mentioned above are illustrated as horizontal dashed-line in Figure 1(b).

III. LSTM Network Architectures

Recurrent neural network(RNN) is developed to handle the sequence data with variable lengths in neural networks. Furthermore, it is designed to keep some information that is the key for the final result even after long steps, and it is significant to drop some information during the network running, as not everything affects the performance positively. The traditional RNN has a problem that it can be hard to solve problems when requiring learning long-term temporal dependencies. That is because the problem called the vanishing gradient makes the gradient of the loss function decay exponentially with time. LSTM network [12] is a kind of RNN that uses special units in addition to standard units, and it adds a memory cell that can preserve information in memory for long periods. LSTM keeps the error at a more constant level, allowing the circulatory network to learn in many time steps and establish long-range connections. LSTM does not lose the information in the gate which outside the normal flow of the recurrent networks. These gates can store, write or read information, it learns when to allow data to enter, leave, or be deleted by guessing, propagating errors backward, and iterative processes that adjust the weights with gradient descent.

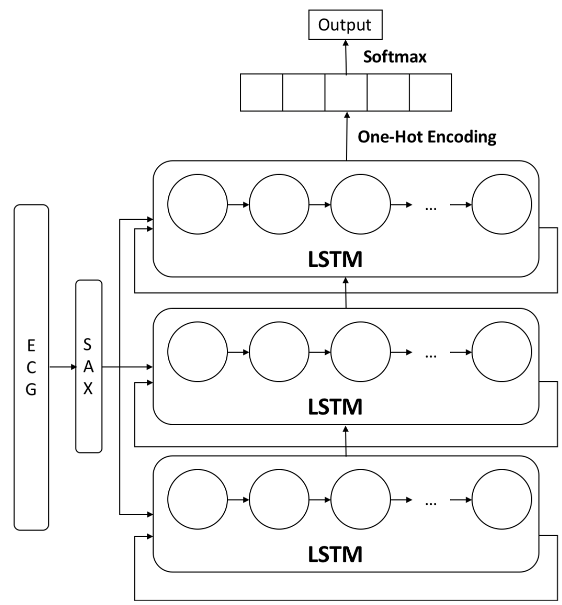


Fig. 2: Our LSTM Model

In our proposed network, we use LSTM to encode the input sequence data as shown in Figure 2. The input data is a series of letters which is the output of preprocessing by SAX. Then we obtain a k-dimensional vector as a final output from

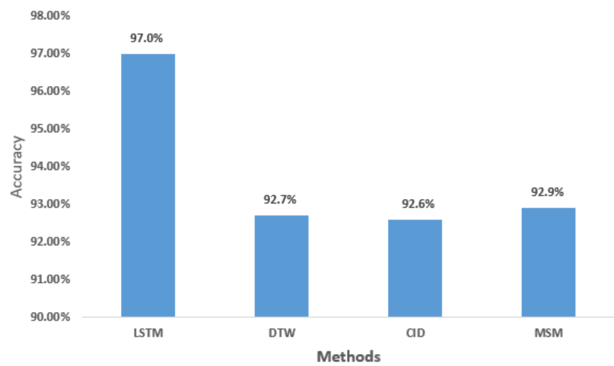


Fig. 3: Accuracy with different methods

the LSTM network, where k may implicitly represent the number of classification in ECG data. Finally, using one-hot encoding for normalizing the output, we infer the class of heart disease by applying softmax function.

IV. Experiments

For an extensive evaluation, we compared the performance of our proposed heart disease classification based on SAX to various signal pre-processing techniques such as SAX and FFT. We also discuss the result of experiments by using different methods of ECG signal classification, varying preprocessing methods of ECG signal classification, and changing the SAX parameters to see the influence on the performance.

Data set: The dataset used in our experiments was originally collected by [13] and we obtained it from UCR Time Series Archive [14]. It is reported in the data description that the raw ECG signal data was pre-processed in two steps: (1) extracting heartbeats, (2) normalizing the length of heartbeats to have the same length using interpolation. Then, $5,000$ heartbeats were randomly selected. It is said that the patients have severe congestive heart failures and the class values were obtained by automated annotation.

Performance of classifications: We evaluate 4 different classification methods, the data used in the comparison without any preprocessing. We compare the performance of LSTM, Dynamic time warping (DTW), Move-split-merge (MSM) [15], Complexity invariant distance (CID) [16].

- *DTW* is a method that identifies the class by finding the closest sample based on the distance measure called DTW, which is widely used to compute the similarity between two time series sequences.
- *MSM* represents a distance metric, which is conceptually similar to general edit distance-based approaches, that calculates similarity by using a set of operations to transform a given series into a target series.
- *CID* denotes a weighted distance measure to compensate for differences in the complexity of the two series being compared.

We find that even if there is not a preprocessing step, LSTM always achieves the high-accuracy about 97%.

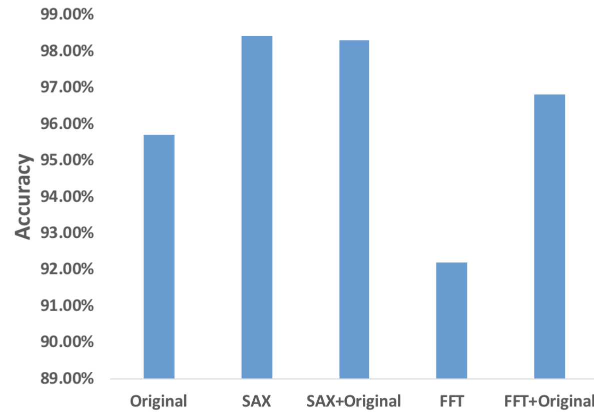


Fig. 4: Accuracy with different preprocessing methods

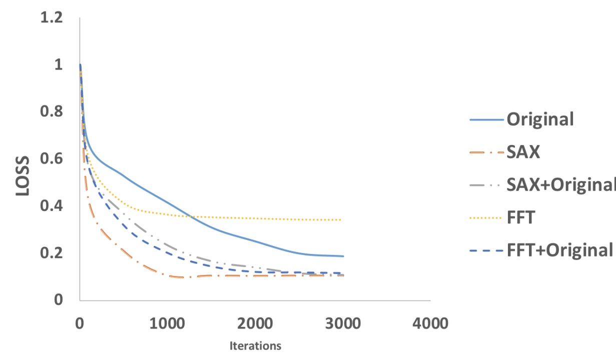


Fig. 5: Loss with different preprocessing methods

Varying preprocessing: Next we conducted five sets of experiments to analyze the data preprocessing results. We use fast Fourier transform(FFT) as a comparison method. A fast Fourier transform (FFT) is an algorithm that samples a signal over a period and divides it into its frequency components. To test the effect of preprocessing methods of the experimental result, we used original data, data which preprocessed with SAX, data which extend by SAX, data which preprocessed with FFT, and data which extend by FFT as experimental data. The extension means that the processed data merged into original data. And we utilize LSTM as the classification model. The Figure 4, 5 show that the accuracy of classification and the change of loss both change with

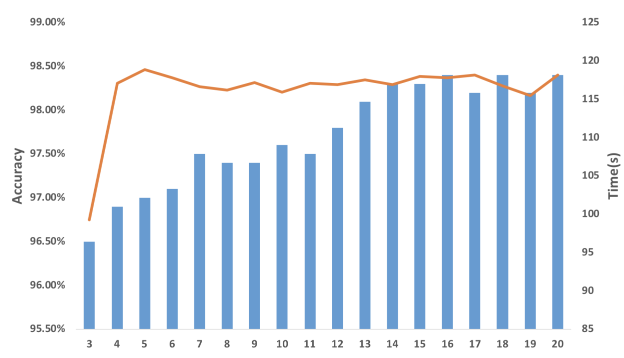


Fig. 6: Parameter: Alphabet Size

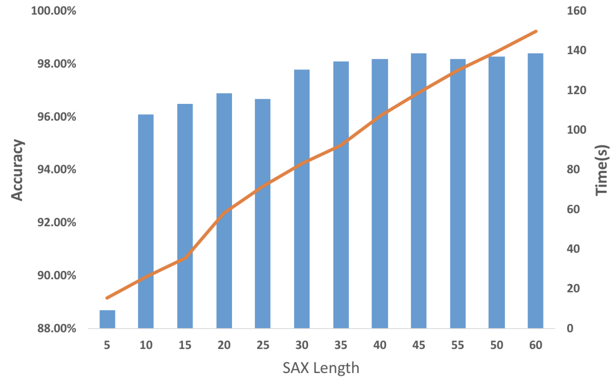


Fig. 7: Parameter: SAX Length

iteration. It is obvious that obtaining the highest accuracy 98.4% using SAX as preprocessing method. However, the accuracy of using FFT alone is the lowest, but the accuracy is obviously improved when the original dataset is extended by FFT. The accuracy of using FFT as preprocessing method is lower than the other method since only use FFT transform the original signals to be the frequency domain to make the features sparse. Furthermore, the method which using SAX as preprocessing method reached the optimal value at about 1300 iterations, and the method which using SAX, FFT to extend the data reached the optimal value at about 2300 iterations. The experimental results show that SAX reduces the dimension of the data and keeping the features of the data very well, thus making the accuracy of the classification up to the highest in the minimum time.

Varying hyper-parameters for SAX: From the experimental results of the previous section, we conclude that using SAX to preprocess data can achieve the better performance and lower cost. We attempt to identify the major SAX parameter that influences the result of accuracy and response time. We use the length parameter and the alphabet size parameter to explore what parameters to achieve the maximum performance and minimum response time. Figure 6 shows that the effect of alphabet size parameters on classification accuracy and response time. From the figure we found that the alphabet size does not affect the response time, even if what size is used, the response time always same. But the accuracy increases gradually as the alphabet size increases. The accuracy tends to be stable after the alphabet size is higher than 14.

Figure 7 shows that the effect of length after dimension reduction on classification accuracy and response time. From the figure, we find that the length of SAX is proportional to the response time. The longer length, the model need processes more sequences and take more time. We also found that the accuracy gradually increased with the increase of length and no obvious increase in the length of 35 or more. The accuracy of the classification results tended to be stable at about 98.4%, and the response is still increasing. Therefore, the experimental results show that using the SAX data preprocessing and the RNN with LSTM units model, we get the highest performance when the alphabet size is greater than or equal to 14 and SAX length is greater than or equal to 35. The accuracy of classification is 98.4% and the response time is much less than the method without preprocessing.

V. CONCLUSION

In this paper, we propose a method of classifying heart disease using ECG signals that achieves high accuracy in a short period. We also compared the performance of various classification methods and preprocessed methods and analyzed the preprocess parameters in detail. Experimental results show that the classification accuracy and performance can be improved efficiently.

ACKNOWLEDGMENT

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (2017M3C4A7063570).

REFERENCES

[1] W. Phanphaisarn, A. Roeksabutr, P. Wardkein, J. Koseeyaporn, and P. Yupapin, "Heart detection and diagnosis based on ECG and EPCG relationships," *Medical Devices (Auckland, NZ)*, vol. 4, pp. 133–144, 2011.

[2] H. Khorrami and M. Moavenian, "A comparative study of dwt, cwt and dct transformations in ecg arrhythmias classification," *Expert systems with Applications*, vol. 37, no. 8, pp. 5751–5757, 2010.

[3] M. Korürek and B. Dogan, "ECG beat classification using particle swarm optimization and radial basis function neural network," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7563–7569, 2010.

[4] R. Saini, N. Bindal, and P. Bansal, "Classification of heart diseases from ecg signals using wavelet transform and knn classifier," in *Computing, Communication & Automation (ICCCA), 2015 International Conference on*, 2015, pp. 1208–1215.

[5] E. D. Übeyli, "Statistics over features of ecg signals," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8758–8767, 2009.

[6] J. Park, M. Kang, J. Gao, Y. Kim, and K. Kang, "Cascade classification with adaptive feature extraction for arrhythmia detection," *J. Medical Systems*, vol. 41, no. 1, pp. 11:1–11:12, 2017.

[7] E. J. da S. Luz, W. R. Schwartz, G. C. Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 144–164, 2016.

[8] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2003, pp. 2–11.

[9] J. Lin, E. J. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time series," *Data Min. Knowl. Discov.*, vol. 15, no. 2, pp. 107–144, 2007.

[10] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Locally adaptive dimensionality reduction for indexing large time series databases," *ACM Sigmod Record*, vol. 30, no. 2, pp. 151–162, 2001.

[11] A. Apostolico, M. E. Bock, and S. Lonardi, "Monotony of surprise and large-scale quest for unusual words," *Journal of Computational Biology*, vol. 10, no. 3-4, pp. 283–311, 2003.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] 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," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[14] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The ucr time series classification archive," *URL www. cs. ucr. edu/~ eamonn/time_series_data*, 2015.

[15] A. Stefan, V. Athitsos, and G. Das, "The move-split-merge metric for time series," *IEEE transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1425–1438, 2013.

[16] G. E. Batista, X. Wang, and E. J. Keogh, "A complexity-invariant distance measure for time series," in *Proceedings of the 2011 SIAM international conference on data mining*. SIAM, 2011, pp. 699–710.