Detecting Cardiac Abnormalities Using 12-Lead ECG and Deep Learning

Roneel V. Sharan
Australian Institute of Health Innovation
Macquarie University
Sydney, Australia
roneel.sharan@mq.edu.au

***Abstract*—The standard 12-lead electrocardiogram (ECG) is widely used by cardiologists in diagnosing cardiac abnormalities. However, manual interpretation of ECG signals can be time consuming and dependent on the skills of the clinicians. In this work, an approach for detection of cardiac abnormalities using automatic analysis of 12-lead ECG is presented and validated on a comprehensive dataset with eight cardiac abnormalities and one normal sinus rhythm. The proposed approach uses the raw ECG signals as a direct input to a model comprised of convolutional neural network and bi-directional long short-term memory. The dataset includes subjects with multiple cardiac conditions to account for which a binary classification strategy is utilized, in particular, the one-against-all classification method. A weighted cross entropy loss function is used to compensate for the imbalance in the class sizes. While the ECG signals in the dataset are up to 60 seconds long, the proposed approach utilizes only the first 15 seconds of the signals as it was seen to produce comparable performance with lower computational costs. An average validation accuracy of 96.19%, *F*-score of 0.8026, and AUC of 0.9624 is achieved using the proposed method.**

Keywords*—bi-directional long short-term memory, cardiac abnormalities, convolutional neural networks, electrocardiogram

I. Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally. It is estimated that 17.9 million people died from CVDs in 2016, accounting for 31% of all deaths [1]. In Australia, CVD was the underlying cause of death in 41,800 deaths in 2018 (26% of all deaths) and an associated cause of death in 70,600 deaths [2].

CVDs are a group of disorders of the heart and circulation. This includes various cardiac abnormalities such as cardiac arrhythmias. Early detection of cardiac abnormalities could help improve clinical outcome [3]. Electrocardiogram (ECG) is a non-invasive measure of the electrical activity of the heart using electrodes placed on the surface of the body. The standard 12-lead ECG is a widely used tool in the diagnosis of various cardiac abnormalities [4].

Manual interpretation of ECG is, however, time consuming and dependent on the training and skills of the personnel [5]. Automatic analysis of the ECG signal can assist the physicians in the diagnosis of cardiac abnormalities. As such, the use of signal processing and machine learning techniques in the detection of cardiac abnormalities has generated a lot of interest.

Conventionally, this has been achieved using various signal processing, feature extraction, and feature selection techniques. For example, studying the heart rate variability based on the R-R interval, analysis of the P, QRS, and T patterns [6], analyzing the power spectrum of the rhythms [7], and entropy-based measures [8]. However, more recently, these methods have been surpassed by deep learning methods. In particular, convolutional neural networks (CNN) which have shown to learn distinguishing ECG signal characteristics [9], sometimes without the need for any signal processing, feature extraction, and feature selection.

In this work, the 12-lead ECG signals are used as a direct input to a one-dimensional convolutional neural network (1-D CNN) for detecting cardiac abnormalities, forgoing all preprocessing steps. Furthermore, long short-term memory (LSTM) is an artificial recurrent neural network [10], a deep learning technique that is particularly suited to sequence and time series data classification. Bidirectional LSTM (BiLSTM) is an extension of the conventional LSTM that can learn from past and future states. In this work, a BiLSTM network is combined with a CNN for the classification of multi-channel ECG signals, as seen to be useful in [9]. This allows the network to learn the feature representations together with the temporal relationship between the features.

The proposed approach is evaluated on a comprehensive dataset containing multiple cardiac abnormalities. In addition, a number of ECG recordings in the dataset contain multiple cardiac abnormalities. To address this multi-label classification problem, a binary classification strategy, in particular, the one-against-all classification method is explored whereby the classifier is trained to recognize one disease against all other diseases combined. The dataset also has an imbalance of class sizes which is accounted for using a weighted classification layer.

The rest of the paper is organized as follows. The dataset used in this work is described in Section II along with the CNN-BiLSTM classification method and the performance evaluation metrics. The experimental results are presented in Section III and discussion and conclusion in Section IV.

II. Method

A. Dataset

The dataset used in this work has 6,877 recordings of 12-lead ECG signals with a sampling frequency of 500 Hz [11]. The recordings come from 11 hospitals and the duration of the recordings varies from 6 seconds to 60 seconds. Moreover, 3,699 recordings are from male subjects and 3,178 from female subjects.

978-1-7281-9052-5/20/$31.00 ©2020 IEEE

TABLE I. DISTRIBUTION OF THE DIFFERENT CARDIAC ABNORMALIATIES AND NORMAL SINUS RHYTHM IN THE DATASET.

Label	Number of Recordings with:			Total Label Count
	One Label	Two Labels	Three Labels	
AF	976	242	3	1,221
I-AVB	686	36	0	722
LBBB	179	54	3	236
Normal	918	0	0	918
PAC	533	80	3	616
PVC	607	93	0	700
RBBB	1,533	321	3	1,857
STD	784	82	3	869
STE	185	32	3	220
Total	**6,401**	**940**	**18**	**7,359**

Each recording is marked either normal or one or more labels from eight cardiac abnormalities. The nine labels are:

1. atrial fibrillation (AF),
2. first-degree AV block (I-AVB),
3. left bundle branch block (LBBB),
4. normal sinus rhythm (Normal),
5. premature atrial contractions (PAC),
6. premature ventricular contractions (PVC),
7. right bundle branch block (RBBB),
8. ST depression (STD), and
9. ST elevation (STE).

As summarized in Table I, of the 6,877 recordings, 6,401 recordings have only one label, 470 recordings have two labels, and 6 recordings have 3 labels. This gives a total of 7,359 labels. The distribution of the recordings is uneven with RBBB having the most number of occurrences at 1,857 and STE having the least number of occurrences at 220.

An illustration of the first 2 seconds of a 12-lead ECG recording labeled as AF is shown in Fig. 1. More details on the dataset can be found in [12].

B. *Deep Learning Model*

In this work, the 12 lead ECG signals are used as a direct input to a 1-D CNN which is followed by a BiLSTM layer. The durations of the ECG signals in the dataset vary considerably but the CNN requires input of fixed length. Some preliminary experiments showed that fixing the length of the signals to 15 seconds (7,500 sample points) achieved similar performance to fixing it to 60 seconds (30,000), the length of the longest duration signal. A significant advantage of using 15 second long signals over 60 long signals is the reduced computational costs. Therefore, the length of all the signals was fixed at 15 seconds using zero-padding and cropping.

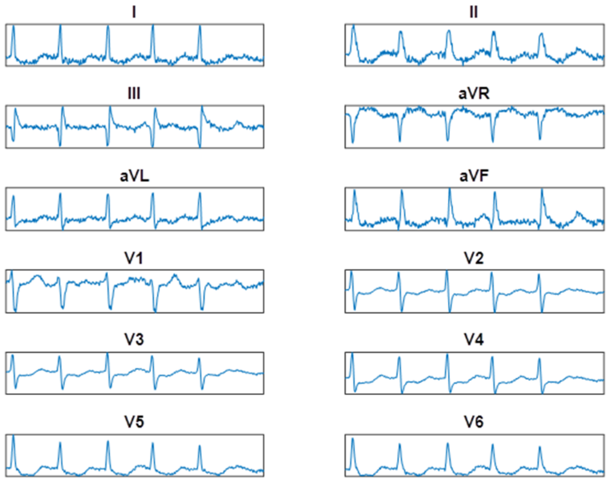

Fig. 1. Plot of the first 2 seconds of a 12-lead ECG signal labeled as atrial fibrillation.

A total of 476 recordings have 2 or more labels making this a multi-label classification problem. This problem is solved in this work by transforming it into a binary classification task. In particular, a technique similar to the one-against-all classification method is used. In this approach, the classifier is trained to separate one class against all other classes combined, one at a time. This results in 9 different models with each model making a prediction for each class.

An overview of the CNN architecture used in this work is provided in Fig. 2. The network has five convolutional layers and each convolutional layer is followed by a batch normalization layer [13] and rectified linear unit (ReLU) [14]. The filter size in the four convolutional layers is 1×15, 1×15, 1×10, and 1×4, respectively, each with a stride of 1×3. The number of filters in each convolutional layer is 128. Each ReLU layer is followed by a max pooling layer [15] of size is 1×10, 1×10, 1×5 and 1×3, respectively, each with stride 1×2. The final pooling layer is followed by a flatten layer, a BiLSTM layer [16], ReLU layer, dropout layer [17], fully connected layer, and a softmax layer [18]. The dropout layer has a probability of 0.2 and the number of hidden units in the BiLSTM layer is set to 200.

The distribution of labels in the dataset is imbalanced. In this work, this is accounted for using a weighted cross entropy loss in the final layer as

$$L = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K} c_i T_{ni} \log\left(Y_{ni}\right) \tag{1}$$

where Y represents the prediction scores, T the training targets, N the number of observations, K the number of classes, and c the class weights.

The CNN model was trained using adaptive moment estimation (Adam) [19] algorithm. Adam uses the moving average of the first and second moments of the gradients to adapt the learning rate which, for iteration t, is given as

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \text{ and } \hat{v}_t = \frac{v_t}{1-\beta_2^t}, \tag{2}$$

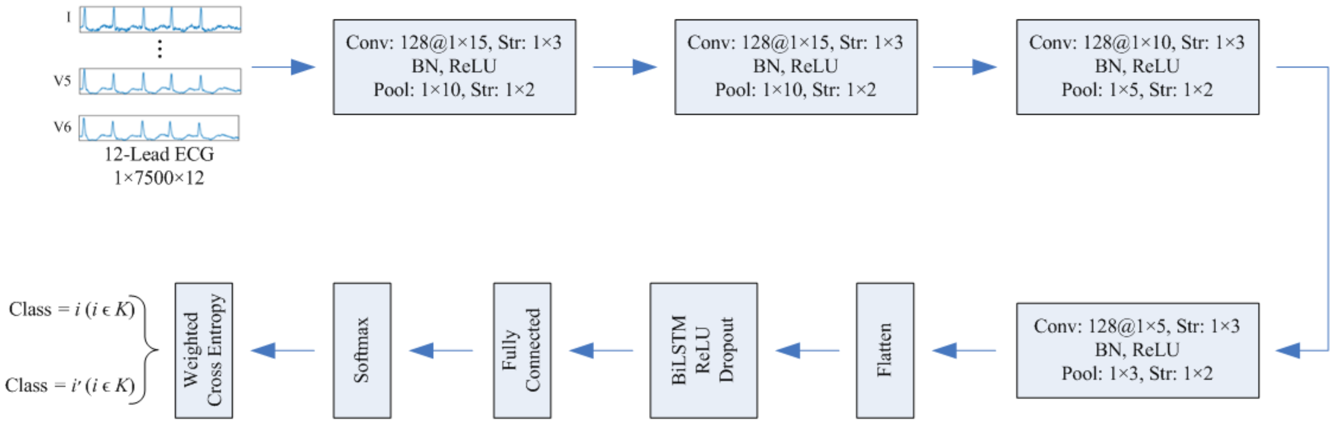

Fig. 2. An overview of the CNN-BiLSTM one-against-all classification architecture used in this work.

respectively, where β_1 and β_2 are the algorithm hyperparameters. The weight w of the model is then updated as

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \tag{3}$$

where η is the step size and ε a small scalar.

The initial learn rate of the algorithm was set to 0.003 with a drop factor of 0.1 at a drop period of 10. The mini batch size was set to 128 and the training was stopped at 20 epochs.

C. Performance Metrics

The classification performance of the proposed method is evaluated in 10-fold cross validation using the accuracy and F-score (or F_1 score) as

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}, \tag{4}$$

$$F\text{-}Score = \frac{2TP}{2TP + FP + FN} \tag{5}$$

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, and FN is the number of false negatives. The area under the curve (AUC) of the receiver operating characteristics (ROC) curve is also used as an evaluation measure given as

$$AUC = \int_0^1 f(x)\,dx \tag{6}$$

where $f(x)$ is the ROC function curve. The AUC was computed using trapezoidal approximation.

III. Experimental Evaluation

The 10-fold cross-validation results in detecting cardiac abnormalities using 12-lead ECG signals are given in Table II. The results are presented using two models, a CNN model and a CNN-BiLSTM model. The CNN model is similar to the proposed CNN-BiLSTM model illustrated in Fig. 2 but without the BiLSTM layer and associated layers. The input to both the models is exactly the same, raw 12-lead ECG signals of duration 15 seconds.

With the CNN model, an average accuracy, F-score, and AUC of 95.75%, 0.7836, and 0.9477, respectively, is achieved in detecting normal sinus rhythm and eight cardiac abnormalities. Looking at the individual class performances, the highest accuracy of 99.21% is achieved in detecting LBBB and the highest F-score of 0.9329 and AUC of 0.9902 in detecting RBBB. With an accuracy of 91.32%, F-score of 0.5004, and AUC of 0.8388, the lowest results are in detecting PAC.

Similarly, with the CNN-BiLSTM model, an average accuracy, F-score, and AUC of 96.19%, 0.8026, and 0.9624, respectively, is achieved. For the individual classes, the highest accuracy of 99.16% is once again achieved in detecting LBBB and the highest F-score of 0.9390 and AUC of 0.9900 in detecting RBBB. PAC once again has the lowest detection accuracy of 92.92% and AUC of 0.8849 while the lowest F-score is for class STE at 0.5876.

While the results using both the models show similar trend, the overall results using the CNN-BiLSTM model is slightly better than the CNN model. With the CNN-BiLSTM model, there is a relative improvement of 0.46%, 2.42%, and 1.55% in the accuracy, F-score, and AUC, respectively, over the results using the CNN model.

IV. Discussion and Conclusion

A method for detecting normal sinus rhythm and eight cardiac abnormalities using 12-lead ECG signals is presented in this paper. The proposed method uses deep learning to learn the distinguishing signal characteristics directly from the raw signal, without the need for any feature engineering, and uses the one-against-all strategy for multi-label classification. The classification results using a combined CNN-BiLSTM model are slightly better than using CNN on its own.

The results using the proposed method are comparable to the validation results in [9], one of the best results achieved on this dataset. However, the method presented in this work has some advantages. The model input ECG signal has a duration of 15 seconds in this work compared to 144 seconds in [9] which is 9.6 times longer and, therefore, has higher computational costs. Also, a total of

TABLE II. 10-FOLD CROSS-VALIDATION RESULTS IN DETECTING CARDIAC ABNORMALITIES USING CNN AND CNN-BILSTM MODELS.

	CNN			**CNN-BiLSTM**		
Type	**Accuracy (%)**	***F*-Score**	**AUC**	**Accuracy (%)**	***F*-Score**	**AUC**
AF	96.68	0.9087	0.9796	97.00	0.9169	0.9852
I-AVB	97.24	0.8704	0.9786	97.50	0.8812	0.9833
LBBB	99.21	0.8875	0.9680	99.16	0.8807	0.9815
Normal	93.11	0.7644	0.9584	94.29	0.7986	0.9699
PAC	91.32	0.5004	0.8388	92.92	0.5931	0.8849
PVC	95.77	0.7849	0.9339	96.22	0.8119	0.9572
RBBB	96.31	0.9329	0.9902	96.64	0.9390	0.9900
STD	94.75	0.8004	0.9599	95.25	0.8141	0.9648
STE	97.34	0.6030	0.9215	96.71	0.5876	0.9445
Average	**95.75**	**0.7836**	**0.9477**	**96.19**	**0.8026**	**0.9624**

10 convolutional layers and 5 pooling layers are used in [9] compared to only 4 convolutional and pooling layers in this work. This again reduces the computational costs of the model in this work. This work, however, utilizes a binary classification strategy for the multi-label classification problem whereas they utilize a sigmoid activation function instead of softmax.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (CVDs)," 2017. [Online]. Available: www.who.int [Accessed: 24th August, 2020]

[2] Australian Institute of Health and Welfare, "Cardiovascular disease," 2020. [Online]. Available: www.aihw.gov.au [Accessed: 25th August, 2020]

[3] W. Amara, C. Montagnier, S. Cheggour, M. Boursier, C. Gully, C. Barnay*, et al.*, "Early detection and treatment of atrial arrhythmias alleviates the arrhythmic burden in paced patients: The SETAM study," *Pacing and Clinical Electrophysiology,* vol. 40, no. 5, pp. 527-536, May 2017.

[4] P. Kligfield, L. S. Gettes, J. J. Bailey, R. Childers, B. J. Deal, E. W. Hancock*, et al.*, "Recommendations for the standardization and interpretation of the electrocardiogram," *Journal of the American College of Cardiology,* vol. 49, no. 10, pp. 1109-1127, 2007.

[5] M. Bickerton and A. Pooler, "Misplaced ECG electrodes and the need for continuing training," *British Journal of Cardiac Nursing,* vol. 14, no. 3, pp. 123-132, 2019.

[6] M. Yochum, C. Renaud, and S. Jacquir, "Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT," *Biomedical Signal Processing and Control,* vol. 25, pp. 46-52, 2016.

[7] S. Kara and M. Okandan, "Atrial fibrillation classification with artificial neural networks," *Pattern Recognition,* vol. 40, no. 11, pp. 2967-2973, 2007.

[8] R. Alcaraz and J. J. Rieta, "A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms," *Biomedical Signal Processing and Control,* vol. 5, no. 1, pp. 1-14, 2010.

[9] T.-M. Chen, C.-H. Huang, E. S. C. Shih, Y.-F. Hu, and M.-J. Hwang, "Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model," *iScience,* vol. 23, no. 3, p. 100886, 2020.

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

[11] E. A. Perez Alday, A. Gu, A. Shah, C. Liu, A. Sharma, S. Seyedi*, et al.*, "Classification of 12-lead ECGs: The PhysioNet - Computing in Cardiology Challenge 2020," *PhysioNet,* 2020.

[12] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu*, et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics,* vol. 8, no. 7, pp. 1368-1373, 2018.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167,* 2015.

[14] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *27th International Conference on Machine Learning*, Haifa, Israel, 2010, pp. 807-814.

[15] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?," in *IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 2146-2153.

[16] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing,* vol. 45, no. 11, pp. 2673-2681, 1997.

[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research,* vol. 15, pp. 1929-1958, 2014.

[18] C. M. Bishop, *Pattern Recognition and Machine Learning.* New York: Springer, 2006.

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