

# Impact of Neural Architecture Design on Cardiac Abnormality Classification Using 12-lead ECG Signals

Najmeh Fayyazifar<sup>1</sup>, Selam Ahderom<sup>1</sup>, David Suter<sup>1</sup>, Andrew Maiorana<sup>2</sup>, Girish Dwivedi<sup>3</sup>

<sup>1</sup>Edith Cowan University, Perth, Australia

<sup>2</sup>Curtin University, Perth, Australia

<sup>3</sup>University of Western Australia, Perth, Australia

## Abstract

*Cardiac rhythm abnormality, as associated with irregular heart activity, presents as changes in an electrocardiogram (ECG). In this paper, as part of the PhysioNet Challenge 2020, we propose two cardiac abnormality detection and classification neural models, using 12-lead ECG signals. Our ECU team proposes a hand-designed RCNN model, consisting of 49 one-dimensional convolutional layers, 16 skip connections, and one Bi-Directional LSTM layer. This model scored 61.9% on training dataset using 10-fold cross validation and 62.3% on the challenge's hidden test subset using the PhysioNet Challenge 2020 evaluation metric, ranking 13 out of 100 teams who successfully participated in official phase of challenge. Our second neural model, designed through neural architecture search, scored 64.4% using 10-fold cross validation on training dataset. This model did not score on the Challenge's hidden test subset due to an error; however, we expect performance improvement compared to our hand-designed neural model.*

## 1. Introduction

Cardiovascular disease was reported to be the leading cause of death worldwide in 2012 (accounting for 17.3 million deaths per year), and has been projected to grow to 23.6 million deaths by 2030 [1]. According to the American Heart Association, cardiovascular disease causes 25% of annual deaths in the United States (US) [2]. Development of an automatic cardiac abnormal activity detection system could assist medical personnel in timely and more accurate diagnosis of cardiovascular disease related to rhythm abnormalities. The PhysioNet/Computing in Cardiology (CinC) Challenge 2020, provides an opportunity for the machine learning community to propose scientific solutions to the problem of automatic detection of cardiac abnormalities from standard 12-lead ECGs [3].

In the literature, most successful machine learning based cardiac abnormality detection and classification methods have utilized manually engineered feature extraction schemes [4]. Whilst, these techniques when combined with classical classifiers provide promising classification performance, they require expert knowledge

to manually design feature extraction methods. In this paper, we propose two neural architectures that provide competitive results, without requiring feature extraction and/or feature engineering.

Recurrent Neural Networks (RNNs) which model the temporal dependencies of data, have proved to perform well for time series classification [5]. The heart's electrical conduction system follows a periodic process; thus, such temporal dependencies naturally exist in ECG signals [6]. In this paper, we have experimentally shown that our proposed RCNN model, constructed by combination of a proper CNN model, that effectively extract features from ECG waveforms, and a Long-Short Term Memory (LSTM) model, that models the dependencies in ECG waveforms, can provide comparable results to methods utilizing extensive manual feature engineering. Although this model has provided competitive results on a subset of the Challenge's hidden test dataset, the manual design of the network's structure is a time-demanding task that requires expert knowledge in deep learning optimization.

In recent years, Neural Architecture Search (NAS) has become a popular research topic within machine learning community, assisting the community to automatically design suitable architecture with higher performance for a specific problem. However, do date, these studies have mostly targeted image classification and language modelling tasks [7, 8]. There are very few studies that have applied a NAS method to design a CNN model for biomedical signal classification. In a previous study [9], we explored the performance of a well-known NAS method, Efficient Neural Architecture Search (ENAS) [7], on Atrial Fibrillation (AF) detection, confirming that NAS-Based methods can provide competitive results to state of the art methods for biomedical signal classification tasks.

In this study, we proposed an automatically designed CNN architecture by employing Differentiable Architecture Search (DARTS) method [10], modifying its search space to tailor it to 1-Dimensional ECG waveforms. DARTS algorithm proved to provide accurate results while accelerating the search algorithm compared to previous methods that utilized evolutionary algorithms [8] or reinforcement learning [7].

## 2.1. Dataset and data preparation

The dataset used in this study, provided by PhysioNet Challenge 2020, has 43101 multi-label ECG samples, containing 111 classes of heartbeat rhythms [3]. However, the performance metric defined by challenge organizers uses a subset of 27 of these classes for scoring the performance of competing algorithms. Therefore, we discarded samples belonging to any of the unused classes from our training dataset, reducing the total number of training samples to 37619. Amongst the 27 classes, 3 pairs of classes are considered equivalent for scoring performance, where we have treated each pair as the same class, leading the final classification problem to be a 24-class, multi-label classification task.

The length of ECG samples varies between 6 seconds to 30 minutes, with a median of 10 seconds. As CNN models accept signals with a constant length, waveforms shorter than 10 seconds were zero-padded, and signals longer than 10 seconds were truncated to the last 10 seconds (as a medical expert would generally look at the last few seconds of a signal to assess patient's condition).

In order to evaluate the performance of our proposed models, we have used 10-fold cross validation on the training dataset, that is, we have randomly partitioned the training dataset into 10 equally sized subsets, then trained our models 10 times using one subset for testing, 8 for training as well as one for validation (for determination of early stopping).

In order to reduce the size of input to our NAS-based model without losing meaningful information, and consequently reducing the amount of required computing resources such as memory, the input time domain signal was transformed into wavelet presentation. This was completed by using Db1 mother wavelet with one level of decomposing, providing a set of details and approximation coefficients. In the corresponding experiments, approximation coefficients were used as the input to our model while details were discarded. Our main intention for using wavelet transformation was to reduce the required computing resources; however, Christov et al. [11] showed that wavelet transforms are the most suitable method for ECG analysis. Accordingly, we expect performance improvement.

## 2.2. Our proposed RCNN architecture

Our proposed architecture is illustrated in Figure 1. Inspired by architecture proposed in [12], the CNN section of our proposed architecture is constructed of 16 blocks with three 1-Dimensional convolution layers within each block. All convolutions are followed by a Batch Normalization (BN) layer, a Rectifier Linear Unit (ReLU) activation function, and a dropout layer. All blocks have been connected to their previous block by a skip connection, similar to that of Residual Networks (ResNet) [13]. These skip connections concatenate features at the

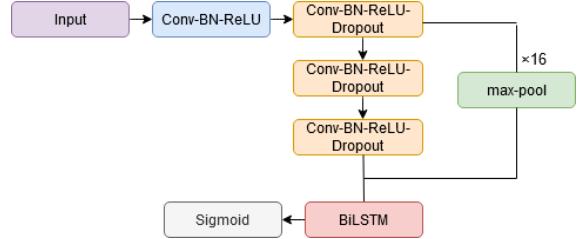


Figure 1. Our proposed RCNN model

end of the last two blocks, allowing information to propagate through the deep CNN structure, thus reducing the effect of vanishing gradient descent. Prior to our first block, the 12-lead ECG inputs were passed through a stem convolution layer with 32 filters, expanding the total number of convolution layers in our model to 49. The number of output channels has been doubled at the end of the 4<sup>th</sup>, 8<sup>th</sup>, and 12<sup>th</sup> blocks, leading to a total of 256 output channels at the last convolution layer. The length of all convolution filters is set to 24. Within each block, the first 2 convolution layers have a stride of 1, and the last layer's stride is 2, leading the size of input signal to be divided by 2 by the end of each block. At each skip connection, a max-pooling layer with a stride of 2 was applied to make the output size of the previous block consistent with output size of current block (to allow concatenation). The output of the last convolution layer is then passed through a BiLSTM layer. The number of hidden units of the BiLSTM layer is empirically adjusted to 300 where the evaluation metric was most promising. The last layer of the network is a sigmoid layer with 24 nodes (corresponds to 24 output classes) which predicts a value between 0 and 1 for each class in our multi-label classification problem, indicating if a sample belongs to a class. The threshold in a binary classification problem is usually 0.5; however, as studied in [14], if the classification problem is imbalanced this threshold can vary. In this paper, by applying moving decision threshold [15] we selected 0.2 as the classification threshold for testing our proposed model.

## 2.3. Our NAS-based architecture

NAS approaches consist of two phases [7, 8]. In first phase, the search algorithm looks for the best possible set of operations within the search space and builds up the best possible architecture. In the second phase, the discovered architecture is trained from scratch to provide the final neural model. Our search space consists of: a) 1-dimensional convolutions (with possible filter lengths: 3,5,7,9,11), b) dilated convolution (filter length 3,5), c) max-pooling (size of 3), d) average-pooling (size of 3), and e) skip connections. Convolution operations have a ReLU-Conv-BN order. The DARTS search method considers each of these blocks as building units of a CNN architecture and optimizes possible operations within the

block, whilst forming the final architecture by stacking these blocks.

In the search algorithm, the total number of blocks were set to 8 where block numbers 3 and 6 are reduction blocks. In reduction blocks the number of convolution channels are doubled while input signal length is reduced by a factor of two. Each of our proposed blocks has two input nodes (outputs from two previous blocks or input data in the case of first two blocks), two middle nodes, and one output node that concatenates the outputs of the middle nodes. Each middle node can be connected to any two previous nodes. Similar to our RCNN model, the last layer is a sigmoid layer with 24 nodes.

The search algorithm in DARTS uses a bi-level optimization strategy, as formulated in Equation 1. In this equation,  $W$ , the weights of the network, are optimized by minimizing training loss, while  $\alpha$ , the selected architectures, are optimized by minimizing validation loss.

Each epoch of the search algorithm samples a neural architecture and evaluates its performance by computing its validation loss. This algorithm moves towards an optimized architecture by minimizing validation loss. We have chosen the architecture with minimum validation loss as the final architecture and trained it from scratch. The discovered normal and reduction blocks are illustrated in Figures 2 and 3, with the training results summarized in Section 3.2.

$$\begin{aligned} \min_{\alpha} \quad & \mathcal{L}_{val}(w^*(\alpha), \alpha) \\ \text{s.t.} \quad & w^*(\alpha) = \operatorname{argmin}_w \mathcal{L}_{train}(w, \alpha) \end{aligned}$$

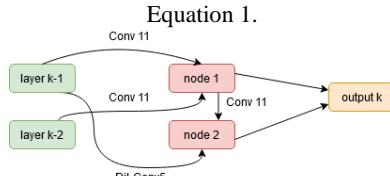


Figure 2. Our discovered normal block

## 2.4. Experimental setup

The parameters of the Our RCNN model are initialized with the ‘He’ initializer [16]. An Adam optimizer was used to train the model parameters. The learning rate was initialized at  $10^{-3}$  and a “ReduceLROnPlateau” learning scheduler with a decay rate of 0.1 and a patience of 3 was used. The minimum learning rate was set to  $10^{-6}$ . The algorithm was trained for 50 epochs, and early stopping with patience of 15 was set. On average, each training epoch of the algorithm took about 300s, leading to a maximum training time of 250 minutes on a single Nvidia Geforce GTX 1080 Ti GPU.

The parameters of our NAS-based network were trained using an SGD optimizer with momentum of 0.9. The learning rate was initiated at 0.025, and a “cosine annealing” learning rate scheduler with a minimum learning rate of  $10^{-3}$  was used. The search algorithm was

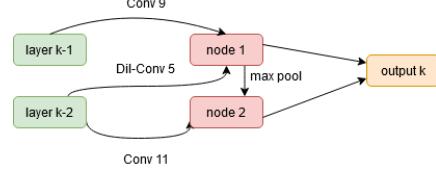


Figure 3. Our discovered reduction block

run for 50 epochs and the architecture with minimum validation loss was chosen for training from scratch. Each epoch of search algorithm took about 50 minutes on a single Nvidia Quadro RTX 8000 GPU. The training from scratch was run for 200 epochs using early stopping with a patience of 20. The train algorithm run time was reported 20 hours on the same Quadro GPU machine.

## 3.1. Results - RCNN model

We trained our proposed Conv-BiLSTM model using time domain ECG signal. The derived model was evaluated using 10-fold cross validation on training dataset, where it received a score of 61.9%. This model was submitted to the Challenge and scored 62.3% on a subset of hidden test dataset. To study the effect of the LSTM model and compare it with the base CNN model, we trained and evaluated the CNN model with the same setting, except removing the BiLSTM section and decreasing our initial learning rate to  $5 \times 10^{-4}$  for the CNN model as with our default learning rate the network did not converge. The evaluation score on training dataset using 10-fold cross validation was 59.2%.

In order to study the effect of wavelet transformation, we have computed the wavelet coefficients, fed them as input to our Conv-BiLSTM model, and evaluated the results. Our proposed network performance is 62.9% using 10-fold cross validation on training dataset. Whilst this model has not received score on hidden test dataset, we estimate some improvement compared to our current scored model.

## 3.2. Results - NAS-based model

In these experiments, wavelet transformations of ECG waveforms were fed as input to the search algorithm. We performed the search algorithm on our designed search space. We chose the normal and reduction blocks (shown in Figures 2. and 3.) sampled at the last epoch of the algorithm where the validation loss is minimum and trained the architecture constructed by stack of those blocks (20 blocks where block numbers 7 and 14 are reduction). The performance of our model was reported as 64.4% using 10-fold cross validation on training dataset. Whist this model failed to receive a score on the hidden test dataset, we expect a roughly 2-3% improvement compared to our hand-designed model, which might have led us to a higher ranking in the Challenge’s leaderboard. Table 1 summarizes all our proposed models together with their score on training dataset using 10-fold cross validation and Challenge’s hidden subset.

Model	Wavelet decomp.	10-fold cross validation	hidden test subset score
<b>Hand-designed CNN</b>	No	59.2%	-----
<b>Hand-designed Conv-BiLSTM</b>	No	61.9%	62.3%
<b>Hand-designed Conv-BiLSTM</b>	Yes	62.9%	-----
<b>NAS-based CNN</b>	Yes	64.4%	-----

Table 1. Summary of our proposed models' score

## 4. Discussion and conclusion

Early detection of cardiac rhythm abnormalities can improve the quality of treatments that patients receive. In this study, we have proposed two neural models which can be used for early diagnosis of 24 classes of cardiac abnormalities without relying on manual feature engineering methods. Our experiments show that proper neural architecture design has a significant impact on model classification performance. By incorporating long-term and short-term memory, BiLSTM networks can learn temporal features and dependencies more accurately than a hand-designed CNN, leading to higher performance. The forget gate in the LSTM models allow the model to discard parts of ECG waveforms that don't contribute towards classification. Long-term memory facilitates the learning process by remembering discriminative features from all parts of the waveform. Our Conv-BiLSTM model reported a score of 61.9% on training dataset using 10-fold cross validation, which is 2.7% higher than our hand-designed CNN model on the same test set. The Conv-BiLSTM model received 62.3% score on the hidden test dataset.

Pre-processing techniques such as wavelet decomposition can boost classification performance. We would expect it to marginally improve our score on the hidden test dataset. With this technique, our experiments suggest a 1% higher classification score on training dataset using 10-fold cross validation, reaching 62.9%. Our experiments indicate that the most significant performance boost can be derived by using a NAS method to search over a subset of standard operations in CNNs, and automatically design the best possible CNN model. Our NAS-based neural model reported a score of 64.4% using 10-fold cross validation on training subset. This model failed to receive a score on hidden test set; however, we expected around a 3% improvement compared to our current score in leaderboard which might have led to a higher ranking.

In future work, it will be desirable to combine LSTM networks and automatically designed CNN architectures, which might provide more promising classification results. Further, the performance of our NAS search algorithm (DARTS) on cardiac abnormality classification could be validated by randomly sampling a subset of discarded architectures and training them from scratch and testing their performance. Moreover, the search space can be further modified, where the effects of different search space designs on model performance can be further explored.

## References

- [1] Laslett, L.J., Alagona, P., Clark, B.A., Drozda, J.P., Saldivar, F., Wilson, S.R., Poe, C., and Hart, M.: ‘The worldwide environment of cardiovascular disease: prevalence, diagnosis, therapy, and policy issues: a report from the American College of Cardiology’, Journal of the American College of Cardiology, 2012, 60, (25 Supplement), pp. S1-S49
- [2] Benjamin, E.J., Muntner, P., Alonso, A., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Chang, A.R., Cheng, S., and Das, S.R.: ‘Heart disease and stroke Statistics-2019 update a report from the American Heart Association’, Circulation, 2019
- [3] Alday, E.A.P., Gu, A., Shah, A., Robichaux, C., Wong, A.-K.I., Liu, C., Liu, F., Rad, A.B., Elola, A., and Seyedi, S.: ‘Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020’, Physiological Measurement, (Under Review), 2020
- [4] Teijeiro, T., Félix, P., Presedo, J., and Castro, D.: ‘Heartbeat classification using abstract features from the abductive interpretation of the ECG’, IEEE journal of biomedical and health informatics, 2016, 22, (2), pp. 409-420
- [5] Hüskens, M., and Stagge, P.: ‘Recurrent neural networks for time series classification’, Neurocomputing, 2003, 50, pp. 223-235
- [6] Seymour, R.S., Hargens, A.R., and Pedley, T.J.: ‘The heart works against gravity’, American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 1993, 265, (4), pp. R715-R720
- [7] Pham, H., Guan, M.Y., Zoph, B., Le, Q.V., and Dean, J.: ‘Efficient neural architecture search via parameter sharing’, arXiv preprint arXiv:1802.03268, 2018
- [8] Real, E., Aggarwal, A., Huang, Y., and Le, Q.V.: ‘Regularized evolution for image classifier architecture search’, (2019.), pp. 4780-4789
- [9] Fayyazifar, N.: ‘An accurate CNN Architecture for Atrial Fibrillation Detection Using Neural Architecture Search’, (IEEE, 2020.), (to be appered in procceding of EUSIPCO)
- [10] Liu, H., Simonyan, K., and Yang, Y.: ‘Darts: Differentiable architecture search’, arXiv preprint arXiv:1806.09055, 2018
- [11] Christov, I., Gómez-Herrero, G., Krasteva, V., Jekova, I., Gotchev, A., and Egiazarian, K.: ‘Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification’, Medical engineering & physics, 2006, 28, (9), pp. 876-887
- [12] Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., and Ng, A.Y.: ‘Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network’, Nature medicine, 2019, 25, (1), pp. 65
- [13] He, K., Zhang, X., Ren, S., and Sun, J.: ‘Deep residual learning for image recognition’, (2016.), pp. 770-778
- [14] He, H., and Ma, Y.: ‘Imbalanced learning: foundations, algorithms, and applications’ (John Wiley & Sons, 2013. 2013)
- [15] Provost, F.: ‘Machine learning from imbalanced data sets 101’, (AAAI Press, 2000), pp. 1-3
- [16] He, K., Zhang, X., Ren, S., and Sun, J.: ‘Delving deep into rectifiers: Surpassing human-level performance on imagenet classification’, (2015),pp. 1026-1034

### Address for correspondence:

Najmeh Fayyazifar, School of Science, Edith Cowan University, Perth, 6027 Australia - Fayyazifar1@ecu.edu.au