ECG signal classification using capsule neural networks

Tejashwini Neela | Swetha Namburu

Department of Electrical and Communication Engineering, GRIET, Hyderabad, India

Correspondence
Tejashwini Neela and Swetha Namburu,
Department of Electrical and Communication Engineering, GRIET, Hyderabad, India.
Email: swethakarima@gmail.com

Abstract
Cardiovascular diseases (CVD) are the dominant cause of deaths in the world, of which 90% are curable. The electrocardiogram (ECG) measures the electrical stimulus of the heart noninvasively. Convolutional neural networks (CNN) act as one of the powerful machine learning techniques to classify ECG arrhythmia classification and other CVDs. Nonetheless, they have some functional flaws like ignorance of spatial hierarchies between the features and are unable to acquire a rotational invariance. To overcome these problems of CNN, a novel neural network named capsule network (CapsNet) is proposed as an efficient algorithm to provide error-free implementation of deep learning over the databases. The main focus of this work is to apply and implement CapsNet for ECG signal classification from the MIT-BIH database and compare its efficiency with the pretrained CNN networks.

1 | INTRODUCTION

According to the World Health Organization (WHO), an average of 17.7 million people died from cardiovascular diseases in 2015 [1, 2]. The electrocardiogram (ECG) signals, simply, the human heart beat signals are recorded based on the muscular and electrical activity of the heart. ECG signals provide the information about the health status of a heart [3]. Classification of ECG signals is thus important to know the specifications of signals recorded from patients for predicting their heart health. When a person’s ECG is recorded, the signal characteristics differ from person to person, time to time and also at a particular instant of time [4]. The most considered areas for ECG signal classifications are arrhythmia detection and other cardio vascular diseases (CVDs). This classification of ECG signals with different classes is used to keep track on the patient’s heart health, which when ignored may cause fatal circumstances. For example, on ignoring premature ventricular contractions (PVCs), the condition may lead to ventricular tachycardia or ventricular fibrillation which may cause heart failure [1]. This shows the necessity for classification of signals to categorize between the normal and the abnormal signals [5].

General classification of ECG signals is carried out in four levels—preprocessing of signals followed by signal segmentation, future extraction and signal classification [6]. In [7], the authors proposed detection algorithms using digital analysis of signal amplitude, slope and width for feature extraction and signal classification. Also, wavelet transformation techniques are also incorporated for PVC detection from ECG signals [8]. However, these techniques have low accuracy due to loss of critical information and nonrecurrent implementations. Furthermore, advanced machine learning (ML) algorithms like support vector machine, principal component analysis and artificial neural networks are applied for improving the classification results [9, 10]. However, the ML algorithms (a) result in poor performance with dataset containing overlapping classes, and (b) find difficulty in the selection of hyperparameters when applied over complex data. These drawbacks then paved a way for the evolution of deep learning techniques for data and image classification.

Deep learning algorithms involve automatic feature extraction while training and resemble the ML algorithms but with larger functionalities and complex architectures. This is a hierarchical process, having multiple layers for subsequent processing of data [11]. In the recent study, deep learning methods are most commonly used in data classifications and are more particularly applied in case of signals with large number of features. Convolution neural network (CNN) being the most widely used deep learning method, has many advantages over ML algorithms. Many classification based state-of-the-art studies, earlier this decade, used CNN as a basic tool for classification of ECG signals [1, 2, 7, 11–13]. In addition, hybrid techniques are adopted by combining CNN with other neural networks like long short memory networks (LSTM) in order to achieve

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Networks* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.
sensitive results. Diagnosis of ECG signals was automated with the help of 2D-CNN and LSTM to classify the arrhythmia signals [14]. Despite having many references for ECG classification using well-renowned CNN, there are still some limitations: (a) CNN suffers Picasso problem, (b) CNN algorithms violate spatial inheritance between the features [3], (c) CNN algorithms need larger amounts of data for efficient progress, (d) pooling in CNN causes great loss of information used for data segmentation and makes it invariant and (e) CNN algorithms are not capable of recognizing pose, texture and deformation of the images [15].

Hence, there is a need to develop a robust deep learning algorithm that overcomes all the problems and provides high efficiency and enhanced sensitivity. This leads to the path for the evolution of capsule networks (CapsNet), that is a consolidated version of CNN. CapsNet is a supervised deep learning algorithm with small neural blocks called capsules in each layer [5, 16]. Recent studies show how CapsNet has evolved as an efficient algorithm for image and signal classification with minimal amount of data. The main functionality of CapsNet revolves around dynamic routing between capsules and expectation-minimization routing [5, 16]. It is used in various classification and detection algorithms such as electroencephalogram emotion signal classification [17] and remote sensing image scene classification [18] to overcome the challenges faced in generative modelling of data [19]. It has provided greater classification accuracy when compared to CNN [3, 20, 21]. In addition, CapsNet has provided minimal reconstruction loss compared to CNN for processing of complex data [22]. CapsNet is also capable of classifying time series data in various stages providing accurate results [21].

The main focus is to implement the ECG signal classification using CapsNet to improve accuracy compared to other existing techniques. The proposed method used a dataset comprising of 18 classes in which 17 classes are from MIT-BIH database and one outlier class from noise signals. The reminder of the study is organized as follows: Section 2 describes the materials and methods adopted. Section 3 discusses the experimental results and Section 4 concludes the study.

## 2 | MATERIALS AND METHODS

### 2.1 | ECG dataset

In this study, the classification approach for different classes of ECG signals of the same size was proposed. The normal ECG signals of various classes have been obtained from dataset namely Physionet database. Outlined ECG signals having noisy data of same length have been added to the above-mentioned dataset, there by including an additional outlier class to the training and validation data. The training dataset now consists of 1096 ECG signals with 18 different classes. Each data signal consists of 3600 data samples. Table 1 describes the different classes of signals and their numerical distributions. Figure 1 illustrates the proposed methodology. The experiment is performed to separate the test set signals to their respective classes including the outliers class. The signals are converted to their respective spectrogram images and then fed to the network for training and testing of data.

### 2.2 | Capsule network

CapsNet is a new deep learning technique that consists of a number of capsules that executes inverse graphics [5]. Each capsule in the network consists of group of neurons. The group of neurons in each layer carry out internal computations to predict the existence of instantiation parameters of a feature at a given position. The main idea of CapsNet implementation is to add capsules to CNN in order to reuse the outputs from...
primary capsules for representing more stable higher capsules. The output is a probability of observation vector. The major implementation of CapsNet requires the knowledge over the following functions,

A. Capsule: A capsule is a building block of CapsNet. It is a collection of neurons whose vector shows the instantiation parameters of an entity. The active capsules at primary level predict the parameters for higher levels. Each capsule gets activated individually for each type of object with various parameters such as position, hue, size and so on.

B. Pooling: CapsNet does not encourage the idea of pooling strategy followed by CNN. It explains the inefficiency of usage of pooling layers as they have no coordinate frame, violating shape perception of input data. They discard the position of information, do not view significance of linear manifold, route statistically and delete the information that relies upon the feature detectors.

C. Layers of CapsNet: The architectural implementation of CapsNet is carried out mainly by the capsule layers identified as primary capsules, higher layer capsules and loss calculation.

The CapsNet is implemented for the classification of signals efficiently. The basic CapsNet model is developed on the basis of dynamic routing between capsules for further processing of ECG data. The CapsNet has three different layers—primary capsule layer, higher capsule layer and decoder network layer shown in Figure 2. The activation function used for the layer is rectified linear unit (ReLU). The kernel size of the layer is defined as 9 with 1 stride and valid padding. The input data set is first fed to the convolutional layer of primary capsule layer, where the convolution between data samples takes place. The layer then proceeds the data to reshape layer of the network, which can be considered as 2D convolutional layer. The dimensions of primary capsule layer are defined as 32×8 with kernel size 9, 2 strides and valid padding. The main function of the primary capsule layer is to reshape the input using the squash activation function. The next process is carried out by the higher capsule layers also called as secondary capsule layer which performs dynamic routing between capsules to inherit the data features and perform the routing function explicitly. The data thus routed dynamically is then processed through three fully connected layers of the decoder network which decodes the explicit features of each signal, leading to efficient classification. After the decoder output is generated, the loss calculation is then carried out to secure the accuracy of the network performance. Back propagation is carried out to enhance the functioning of the network while it is under the training process.

Figure 3 shows the flow of data through the neural network. The first layer is a simple 2D convolution layer that performs spatial convolution over data, where the data is processed over 256 filters and ReLu activation function. The array of signal features obtained from the convolution layer forms the input vector for the primary capsule layer.

The primary capsule layer is made of 32 capsules with output vector size defines as 8. The layer parameters are listed in Table 2. The reshape function defines the output of certain shape. We reshape the input shape to fit the output with proper arguments. The data reshaped is then passed to the layer that performs squash activation. The squash
activation function of primary CapsNet is a nonlinear activation function used in capsules. The function drives the length of large vector to 1 and small vector to 0, in simple words, this function makes sure the length of all sample vectors to be either 1 or 0. For any sample vector ‘x’, the squash function can be defined as

\[
squash(x) = \frac{\|x\|^2}{1 + \|x\|^2} \frac{x}{\|x\|}
\]  

(1)

Furthermore, the actual functionality of CapsNet is carried out with the capsule layer. This layer performs the dynamic routing between the capsules, whose inputs are the output vectors obtained from squash function, using Equation (1), that is either 1 or 0. The number of capsules in this layer is equal to the number of classes of the input data. The number of capsules in higher capsule layer is 18 as there are 18 classes of input data. The output vector dimension of the layer is 16 and minimum number of routings between the capsules is 3.

The output vectors of secondary capsule layer are fed as input to the decoder network. The decoder network is designed by four layers—the first layer is a mask layer that is used to mask a vector by the input shape of the capsule. This is immediately followed by three fully connected layers which can be eventually dense with first two layers having 512 and 1024 units and third layer having units same as that of input size. The activation functions used for the first two layers and third layer are ReLu and sigmoid, respectively. When test input is given to the network, the output of these layers yields classification results which can be used to measure accuracy.

The algorithmic flow of CapsNet is described in Algorithm 1. It consists of three procedures namely softmax, squash and routing. Softmax pooling and squash functioning are the integral parts of routing algorithm of the network, described in Algorithms 2 and 3.

Algorithm 2 Pseudocode for softmax pooling

\[
\text{softmax}(b, i) \\
1: \text{foreach index } i, j \\
2: \quad d \leftarrow c_i \cdot \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}
\]

Algorithm 3 Pseudocode for squash function

\[
squash(s_i) \\
1: \text{foreach index } i, j \\
2: \quad s \leftarrow \frac{\|s_i\|^2}{1 + \|s_i\|^2} \frac{s_i}{\|s_i\|}
\]

3 | EXPERIMENTAL RESULTS

The ECG signal dataset from MIT-BIH database is used in simulations. In this database, PhysioNet service has signals with a frequency of 360 Hz (3600 samples are recorded for 10 seconds). All the signals are nonoverlapping signals. The dataset statistics are also tabulated in Table 3. In the training process, data augmentation is performed on the signals to enhance the efficiency of the network. The predefined dataset has 17 classes plus an additional outlier class results (total 18 classes). The outlier class consists of noisy ECG signals which have the same specifications as mentioned with the signals in other classes. This class has been included for testing the robustness of the network for classifying even low amplitude noise. In addition, AlexNet, GoogleNet, SqueezeNet and Vgg-19 pretrained CNN networks are implemented with depths 8, 18, 22 and 19, respectively.

3.1 | Implementation

The implementation of the network is carried out with the back end TensorFlow libraries and Keras deep learning packages as the basic work stations. The methodology adopted is described in Figure 1. In each class, 80% of data is used for training network and the rest is used for evaluating test accuracy. The number of epochs considered for training are 50. This depicted a substantial improvement in network performance and accuracy.

3.2 | Classification

In the investigation, the number of signal classes used for the implementation varies from method to method when referred to the state-of-the-art implementations. The reason behind this is the class imbalance present in few classes [21]. In contrast,
we have added an additional outlier class and used the complete 18 class dataset for the classification. The trainable parameters have been increased using a greater number of classes. Our model consists of 61,06,944 trainable parameters belonging to 18 classes of signals.

Table 4 compares the results of the proposed CapsNet implementation with the state-of-the-art results. The CNN-LSTM method has acquired a maximum accuracy of 99% with network being trained with only eight classes, whereas the proposed method has acquired 98.5% of training accuracy over 18 classes of data. Furthermore, the test accuracy obtained is 99.14% and the decoder loss obtained is 4%.

### 3.3 | Parameters

#### 3.3.1 | Accuracy

Accuracy is the measure of data that correctly determines the correctness [14]. It is the ratio of sum of true positives and true negatives to total number of data inputs given. The prerequisite for accuracy calculation is the confusion matrix for the signal classification. Mathematically, the accuracy is computed as:

\[
Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%
\]  

#### 3.3.2 | Sensitivity

Sensitivity is the measure of all positives that are positively classified by the network [14]. It is the ability of a network to detect signals of a class pertaining to the same class. It is the ratio of true positives to the sum of true positives and false negatives. It is also called as recall, detection probability, true positive rate. The prerequisite for specificity calculation is the confusion matrix for the signal classification. Mathematically, sensitivity is computed as:

\[
Sensitivity = \frac{TP}{TP + FN} \times 100\%
\]  

#### 3.3.3 | Specificity

Specificity is the measure of all negatives that are correctly predicted negative by the network [14]. It is the ratio of sum of true negatives to the sum of true negatives and false positives. It can also be called as the true negative rate. The prerequisite for specificity calculation is the confusion matrix for the signal classification. Mathematically, the specificity is computed as:

\[
Specificity = \frac{TN}{TN + FP} \times 100\%
\]
are falsely predicted as negative when the actual value is positive and FN refers to false negative values that are falsely predicted as positive when actual value is negative, respectively. TP, TN, FP and FN are the values that are initially calculated for individual class and then the overall values for parameter calculation are taken as the weighted sum of all the individual classes. Here, for a particular class, true positives are the ECG signals that are correctly classified into its actual class, true negatives are the signals that are incorrectly classified into its actual class, false positives are the signals that are correctly classified into other class and false negatives are the signals that are incorrectly classified into other class.

Table 5 shows the confusion matrix plotted for the classification of signals over the test set. The accuracy, sensitivity and specificity in the matrix are calculated using Equations (2), (3) and (4), respectively.

The training progress of network is recorded graphically as the training accuracy and loss, shown in Figure 4. It can be observed that there is a gradual increase in accuracy and gradual decrease in loss over the number of epochs.
In addition, the efficiency of proposed method is evaluated using various pretrained models of CNN like AlexNet, GoogleNet, SqueezeNet and Vgg-19. Table 6 compares the quantitative parameters like training accuracy, sensitivity and specificity of these pretrained CNN networks with the proposed CapsNet method. The proposed method shows a significant improvement in all three parameters over CNN. This is due to capsules added to the layers of network instead of neurons. These capsules perform dynamic routing among themselves to predict and produce accurate output thereby leading to notable improvement in accuracy compared with CNN.

4 | CONCLUSION

For early stage detection, diagnosis and treatment of any CVDs, it is important to know the nature of the patient's heart functioning from time to time. As ECG signals represent heart functionality, it can be predicted by successful classification. We have implemented ECG signal classification using CapsNet, which resulted in a substantial improvement in the accuracy. The simulation considered 1000 ECG signals of a single dimension from the MIT-BIH arrhythmia database and 9% outlier class signals. These 1D signals are converted to their respective spectrogram images before they are fed to the network. The implementation results show that the proposed method has achieved about 98.57% of the training accuracy and 99.14% of the testing accuracy. The experimental results show a significant improvement in signal classification over other state-of-the-art techniques.

ORCID
Swetha Namburu  https://orcid.org/0000-0001-6364-4207

REFERENCES
1. Jun, T.J., et al.: ECG arrhythmia classification using a 2-d convolutional neural network. arXiv preprint arXiv:180406812 (2018)
2. Acharya, U.R., et al.: A deep convolutional neural network model to classify heartbeats. Comput. Biol. Med. 89, 389–396 (2017)
3. Mukhometzianov, R., Carrillo, J.: Capsnet comparative performance evaluation for image classification. arXiv preprint arXiv:180511195 (2018)
4. Das, M.K., Ari, S.: ECG beats classification using mixture of features. International scholarly research notices (2014)
5. Sabour, S., et al.: Dynamic routing between capsules. Adv. Neural Inform. Process Syst. 3856–3866 (2017)
6. Li, T., Zhou, M.: ECG classification using wavelet packet entropy and random forests. Entropy, 18(8), 285 (2016)
7. Wu, Y., et al.: A comparison of 1-d and 2-d deep convolutional neural networks in ECG classification. arXiv preprint arXiv:181007088 (2018)
8. Shyu, L.Y., et al.: Using wavelet transform and fuzzy neural network for VPC detection from the holtterECG. IEEE Trans. Biomed. Eng. 51(7), 1269–1273 (2004)
9. Dewangan, N.K., Shukla, S.: ECG arrhythmia classification using discrete wavelet transform and artificial neural network. In: 2016 IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), pp. 1892–1896. IEEE, Bangalore (2016)
10. Ranaware, P.N., Deshpande, R.A.: Detection of arrhythmia based on discrete wavelet transform using artificial neural network and support vector machine. In: 2016 International conference on communication and signal processing (ICICSP). IEEE, pp. 1767–1770 (2016)
11. Yıldırım, Ö., et al.: Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Comput. Biol. Med. 102, 411–420 (2018)
12. Kiranyaz, S., et al.: Real-time patient-specific ECG classification by 1-d convolutional neural networks. IEEE Trans. Biomed. Eng. 63(3), 664–675 (2015)
13. Zubair, M., et al.: An automated ECG beat classification system using convolutional neural networks. In: 2016 6th International conference on IT convergence and security (ICTCS). IEEE, pp. 1–5 (2016)
14. Zheng, Z., et al.: An automatic diagnosis of arrhythmias using a combination of CNN and LSTM technology. Electronics, 9(1), 121 (2020)
15. Patrick, M.K., et al.: Capsule networks—a survey. J. King Saud Univ. Comput. Inf. Sci. (2019)
16. Hinton, G.E., et al.: Matrix capsules with EM routing. In: International conference on learning representations (2018)
17. Ali, U., et al.: EEG emotion signal of artificial neural network using capsule network. Int. J. Adv. Comput. Sci. Appl. 11(1), 434–443 (2020)
18. Zhang, W., et al.: Remote sensing image scene classification using CNN-capsnet. Remote Sens. 11(5), 494 (2019)
19. Jaiswal, A., et al.: CapsuleGAN: Generative adversarial capsule network. In: Proceedings of the European conference on computer vision (ECCV), 2018
20. Butun, E., et al.: 1d-cadcapsnet: one dimensional deep capsule networks for coronary artery disease detection using ECG signals. Phys. Med. 70, 39–48 (2020)
21. Jayasekara, H., et al.: Timecaps: Learning from time series data with capsule networks. arXiv (2019). arXiv–1911
22. Xi, E., et al.: Capsule network performance on complex data. arXiv preprint arXiv:171203480 (2017)

How to cite this article: Neela T, Namburu S. ECG signal classification using capsule neural networks. IET Netw. 2021;10:103–109. https://doi.org/10.1049/ntw2.12018

| TABLE 6 | Comparison of CapsNet results with CNN performed over the same dataset |
|---------|------------------------|
| Neural network | Training accuracy (%) | Sensitivity (%) | Specificity (%) |
| AlexNet | 80.79 | 67.76 | 98.74 |
| SqueezeNet | 83.23 | 66.82 | 99.06 |
| GoogleNet | 83.84 | 69.49 | 99.03 |
| Vgg-19 | 85.67 | 65.05 | 99.07 |
| CapsNet (proposed) | 98.57 | 99.70 | 99.03 |