C137 Report: Exploring Time-series Data Classification in Time-frequency Domain by CNN

Shoaib Bin Masud, Boyang Lyu, Tiange Wang

1. Introduction

Time series classification (TSC) problems are a popular research topic in the machine learning field. Predicting the label for time series data can be widely used in different areas, such as biological engineering, trading markets, hydrological, and even astronomy [1]. However, it still remains challenging since the desired model needs to handle both temporal and spatial information from the data. Moreover, due to the intrinsic property of time-series data, the analysis need also take care of the frequency property of the data. Current deep neural network methods for TSC mainly focus either on time domain or frequency domain. Recurrent neural networks [14], LSTM [13] are two of widely used networks for time series classification .The paper we referenced [2] in this project proposed their model “Multiscale Convolutional Neural Network (MCNN)” to solve the TSC problem in the time domain. The reason to switch from conventional time series classification model Convolutional neural network (CNN) is that CNN is capable of capturing complex feature representation using its convolutional layers. To better improve the model’s performance, we considered a combination of the information obtained from both time and frequency domain by applying wavelet transform to the time-series data. MCNN is an end-to-end neural network model, which incorporates feature extraction and classification in a single framework. The model is robust to small inputs variations and doesn’t require any specific feature extractor choice [15]. Since MCNN pre-process the input signal in the transformation stage, it can automatically extract features at different scales without changing the filter size inside the model. MCNN model performs well due to its ability of extracting good feature representations [2], but all the transformation the model applied to the input signal is still kept in time domain. Thus we proposed our assumption: with more diverse features added, the performance of the model should be improved. Specifically, we use wavelet transform as a data augmentation technique, hoping to see an enhancement of the model’s performance. To verify our hypothesis, we use three models on electrocardiogram (ECG) data: a modified MCNN model with an extra wavelet transform branch, an original MCNN model and a simple deep neural network model. We are able to show that the MCNN model combined with wavelet transform achieves higher classification accuracy than other other two models.

2. Related Work

In recent years, people have paid much attention to time series analysis. The daily lives of the human race constantly produce time series data. In former studies, a kind of deterioration warning system based on lectrodogram (ECG) data powered by TSC has achieved unprecedented performance compared with traditional clinical approaches and has been applied in major hospitals [3].

Previously, the approaches of TSC mainly focused on feature extracted analysis or distance based analysis [4]. Distance-based methods are to measure the similarity between any given two time series. Once the similarity metrics is computed, the classification can be done by k-nearest neighbor (KNN) or support vector machine (SVM). For feature-based methods, each time series data is characterized with a feature vector and SVM and logistic regression models are applied [2]. In recent decades, Convolutional Neural Networks (CNN) have generated impressive results in time series data classification [5]. CNN has the ability that can learn complex features by convolutional layers. Among these CNN-based models, Multi-scale Convolutional Neural Networks (MCNN) is a kind of specially designed neural network for time series data. It has a distinctive feature that its first layer contains multiple branches that perform various transformations of time series data, thus extracting features of different time scales.

While considering frequency domain analysis, the first consideration is Fourier transform. However, Fourier Transform cannot clearly separate the frequency information (decomposing the signal into a few frequency components) when the frequency of the signal varies across time. Since ECG signal is not stationary, after Fourier Transform, it's difficult to clearly separate frequency domain components. Moreover, In frequency domain, we will lose the information for the variation of frequencies over time.

Wavelet transform (WT) is a widely used tool that can deal with signals whose frequencies vary over time. A wavelet means a small wave and in brief, a wavelet is an oscillation that decays quickly [6]. WT decomposes a signal at different timescales [7]. Wavelet transform has two broad classes, the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT). We applied CWT in this project because CWT can be used to analyze transient behavior, rapidly changing frequencies, and slowly varying behavior [8].The wavelet transformed data contains information from both time and frequency domain. So CWT data augmentation will be a useful approach to improve MCNN performance on ECG signals.

3. Background

3.1 MCNN Model for TSC

A time series is a sequence of real-valued data points with timestamps [3]. In this project, we mainly focus on building a model such that given a new piece of time series data, it can predict the label for the data.

For the given dataset, we denote a time series as $T = \{t_1, t_2, ..., t_n\}$, where t_i is the value at time stamp i and there are n timestamps for each time series. Consider $\{(T_i, y_i)\}_{i=1}^{N}$ as a collection of data points, where N is the number of segments, the i_{th} segment is denoted as $T_i \in R^{d \times w}$, where d is the number of channels and w is the number of time points inside the segment. $y_i \in [0, m]$ is the corresponding label.

In the multi-scale branches of MCNN, firstly downsampling is used for generating sketches of the time series T with downsampling rate k, which means only every k^{th} data points is kept in the new time series, as shown in Eq.1:

$$T^k = \{t_{1+k*i}\},\ i = 0, 1, ..., \left[\frac{n-1}{k}\right] \quad (1)$$

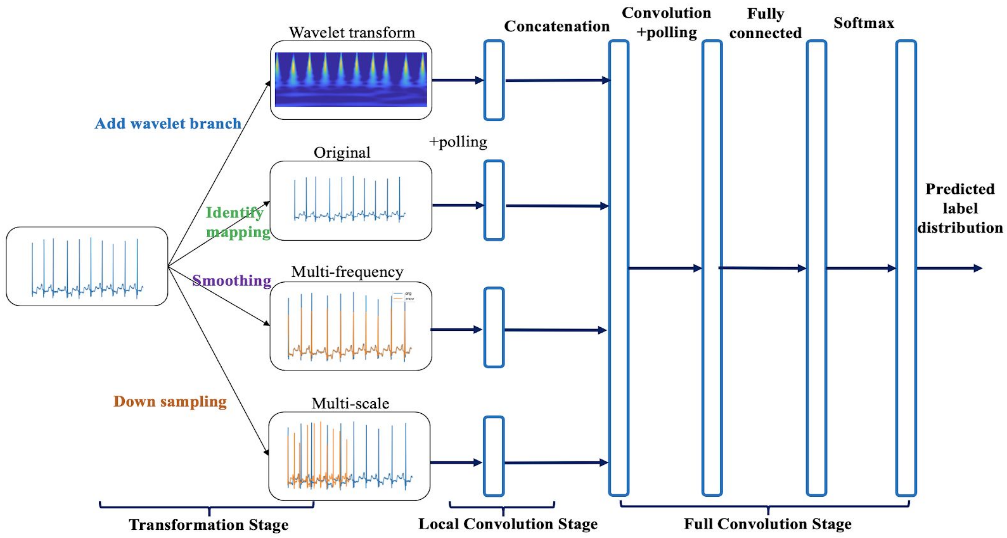
Besides multi-scale branches, the multi-frequency branch is another transformation stage of MCNN. MCNN adopts low frequency filters with multiple degrees of smoothness to decrease the variance of the signal. For the time series T, the low frequency filter is implemented by a moving average window which convert the original time series into a new time series as shown in Eq.2:

$$T^l = (t_i + t_{i+1} + ... + t_{i+l-1})/l \quad (2)$$

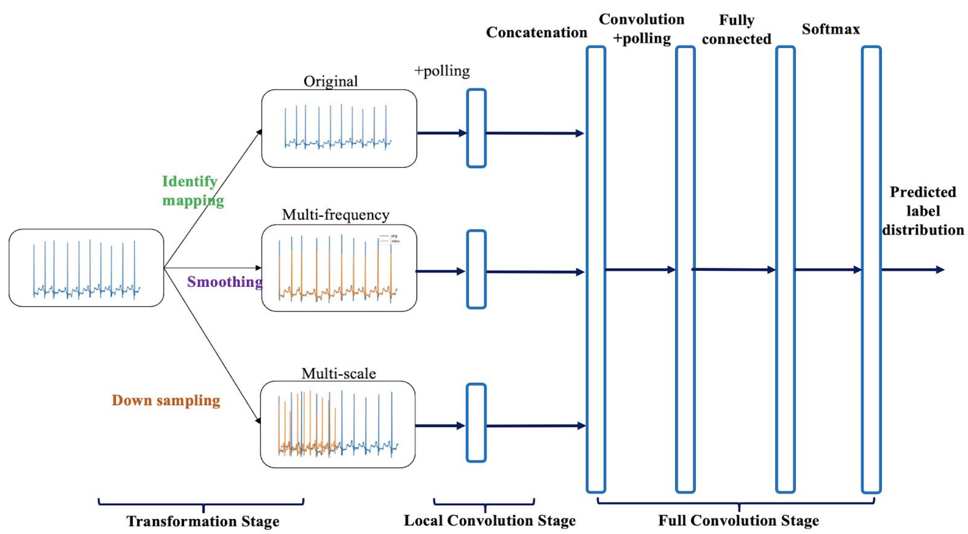
To increase the amount of data, window slicing is proposed for the data augmentation. For the time series T of length n and length of the slice window is s, a slice is a snipped of the original time series. The slicing operation will generate a set of sliced time series as shown in Eq.3:

$$slicing(T, s) = \{S_{1:s}, ..., S_{n-s+1:n}\} \quad (3)$$

All time series will have the same label as they originally do.



(a) MCNN with wavelet transform data augmentation



(b) MCNN without wavelet transform data augmentation

Figure 1: Architecture of Multi-scale CNN

The overall architecture of MCNN framework is depicted in Figure 1(a), similar to the original MCNN structure in Figure 1(b), the model has a transformation stage, a local convolution stage and a full convolution stage. The key different part is in the transformation stage, a wavelet branch is added as an additional feature.

3.2 Wavelet Transform Data Augmentation

In continuous wavelet transform(CWT), the analyzing function is ψ, a wavelet that is a small wave. By comparing signals to the wavelet at various scales and positions, two variables can be obtained. For a scale parameter $a > 0$, and a position parameter b, the formula for CWT is:

$$C(a,b;f(t),\psi(t)) = \int_{-\infty}^{\infty} f(t)\frac{1}{a}\psi^*(\frac{t-b}{a})dt \tag{4}$$

Where * denotes the complex conjugate. Eq.4 shows that the values of scale and position affect coefficients of CWT. Also, the wavelet affects values of coefficients as well [10].

4. Method

4.1 ECG Dataset

In this project, the dataset is ECG data. The dataset was obtained from three groups of people: persons with cardiac arrhythmia (ARR), persons with congestive heart failure (CHF), and persons with normal sinus rhythms (NSR). There are 162 ECG recordings in total and the data matrix is of the shape 162-by-65536, where each row is an ECG recording sampled at 128Hz from 1 channel. There are 96 recordings from persons with arrhythmia, 30 recordings from persons with congestive heart failure, and 36 recordings from persons with normal sinus rhythms. Our goal is to train a classifier to distinguish between ARR, CHF, and NSR [9,10,16,17].

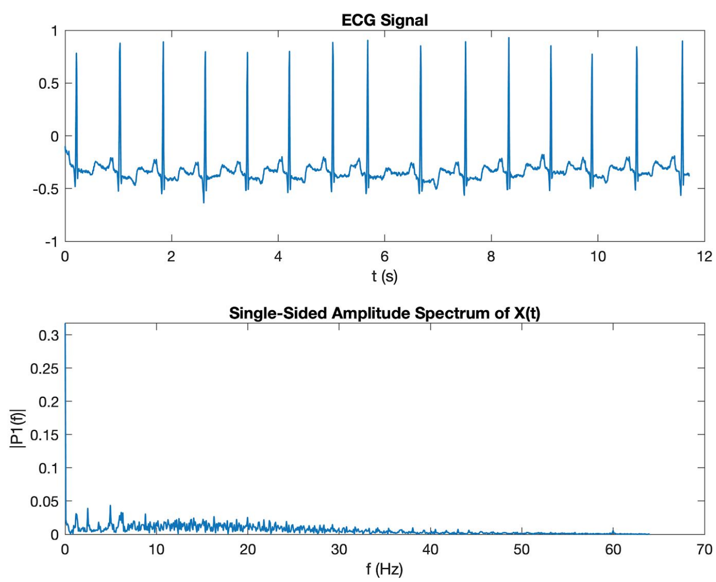


Figure 2: ECG signal and its amplitude spectrum obtained by Fourier Transform

4.2 Data Preprocessing

Before feeding the data set to model, firstly window slicing is applied as shown in Figure 3. The input data is sliced to non-overlapping segments using a window of length 1000 samples with 65 segments per channel. Low pass filter and down-sampling technique is applied to the original signal separately, as shown in Figure 4. The dawn-sampling rate is chosen as k = {2, 3, 4} and parameters for smoothness use the same value.

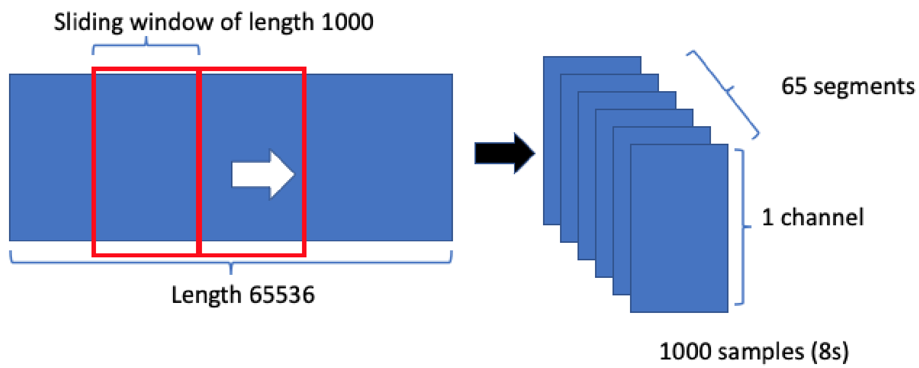


Figure.3: Window slicing for data augmentation

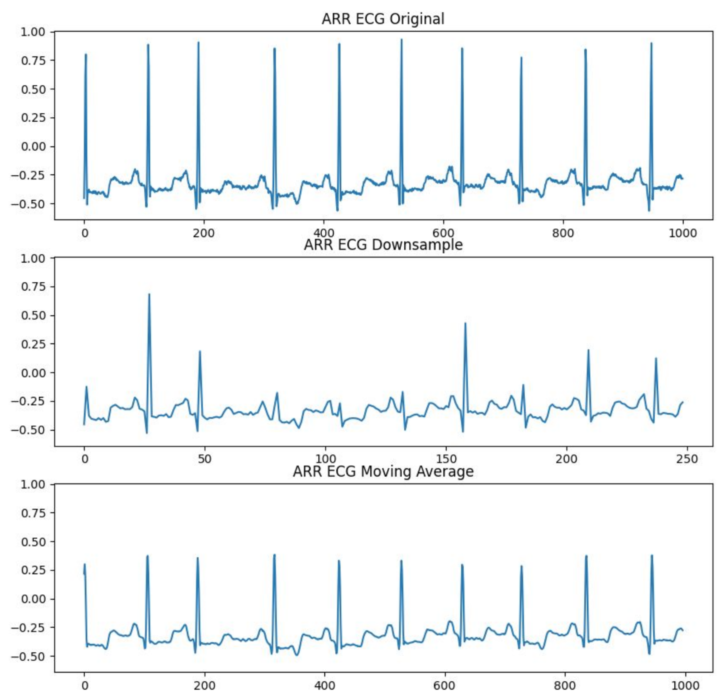


Figure 4: Comparison of original data, down samping data and moving average data. The first row is the original ECG signal, the middle row is the down-sampling signal with a down-sampling rate of 4, the last row is the smoothed signal with the smoothness parameter chosen as 3.

4.3 Wavelet Transform

Figure 5 shows the visualization of an ECG data with timestamps, and the related CWT scalogram. To convolve with the original signal, we have used the Morlet wavelet as mother wavelet. The instantaneous frequencies can be accurately captured by tiling time-frequency planes with variable-sized windows [18]. CWT is shift-invariant which implies that one can shift the signal before applying wavelet transform but the resulting energy will still be the same.

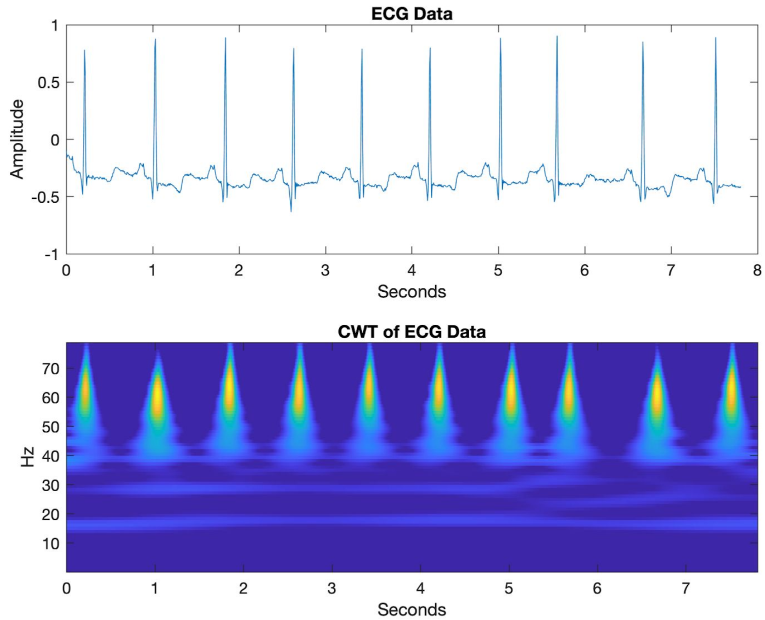


Figure 5: ECG data (upper row) and its corresponding CWT coefficient amplitude (lower row), the brighter color in the lower plot represents higher magnitudes.

4.4 Model Design

In this project, the problem is conducted in a supervised way and input data is pre-processed and concadacated time series data. To make the comparison, the Continuous Wavelet Transform (CWT) will be applied to the original time-series data, making each piece of data a 2-D image with x-axis as time and y-axis as frequency. The transformed time-frequency data will be added as an additional feature to the original time-series data and feed to the model we defined above. Based on the specific structure of the model we referenced, we design our own wavelet layer to extract the feature from the transformed data and concatenate it with the original features. And finally, our output should be labels of each time series data. As stated above, we conduct three neural networks and compare them. Their structures are represented in the table as follows.

Table 1. Design of multi-scale convolutional neural network

Architecture	Operation	Output
Transformation Stage	Add Wavelet branch	Wavelet Transformed Data
	Identify Mapping	Original Data
	Smoothing	Multi-frequency Data

	Down Sampling	Multi-scale Data
Local Convolution Stage	Polling	
	Concatenation	Base Model
Full Convolution Stage	Convolution and Polling	
	Fully Connected	
	Softmax Activation	Prediction Results

Table 1 is our main model, the MCNN with wavelet transform as data augmentation. Our second model is to remove the operation of adding continuous wavelet transforms as augmentation. And the third model is by removing multi-scale and multi-frequency operations. Classification accuracy of the three models will be presented for comparison.

5. Experiments

5.1 Modeling Procedure and Structure

Table 2 includes the hyperparameters we have selected to train the augmented model. Table 3 shows the input and output shape of the base convolutional model. After building the base convolutional model, we concatenated all the flattened features and applied dense layers shown in Table 4, and got the prediction for 3 different classes of our datasets using 'softmax'.

Table 2. Number of parameters for modeling procedure

Procedure	Parameter
Train/Test split	80% training, 20% testing
Validation percentage	10%
Optimizer	Adam
Loss function	Categorical_Crossentropy
Learning rate	0.001
Batch size	32
Training epochs	100

Table 3. Base model architecture

CNN Architecture	Operation
input	-

conv1D	filter(64), kernel size(5), activation(relu), stride(2)
conv1D	filter(32), kernel size(5), activation(relu), stride(2)
Dropout	0.2
MaxPolling1D	2
-	Flatten
Dense	128

Table 4. Wavelet model structure

CNN Architecture	**Operation**
input	-
conv2D	filter(64), kernel size(5), activation(relu), stride(2)
conv2D	filter(64), kernel size(5), activation(relu), stride(2)
Dropout	0.2
MaxPolling2D	2
-	Flatten
Dense	128

5.3 Results Comparison

To analyze model performances, we have chosen the ‘accuracy’ metric. We have compared three different models, one mentioned in MCNN [2], classification using only wavelet features and lastly on combination of the features obtained from both multiscale convolutional neural network and wavelet features. Table 5 includes the accuracy of these three models. We can see that only using wavelet features gave us the lowest accuracy. Using the MCNN model, the accuracy was more than 95%. However, if after combining both the features, we got the highest accuracy.

Table 5. Classification accuracy for each method

Method	**Classification accuracy**
Wavelet transform	86.42%
Multiscale CNN	95.77%
Multiscale CNN + Wavelet transform	97.86%

In Figure 6 we have observed the loss and accuracy vs. epochs. It is obvious from the figure, is that if we use both wavelet and multiscale features, then accuracy and loss reach the plateau much faster than using only a wavelet or multiscale feature.

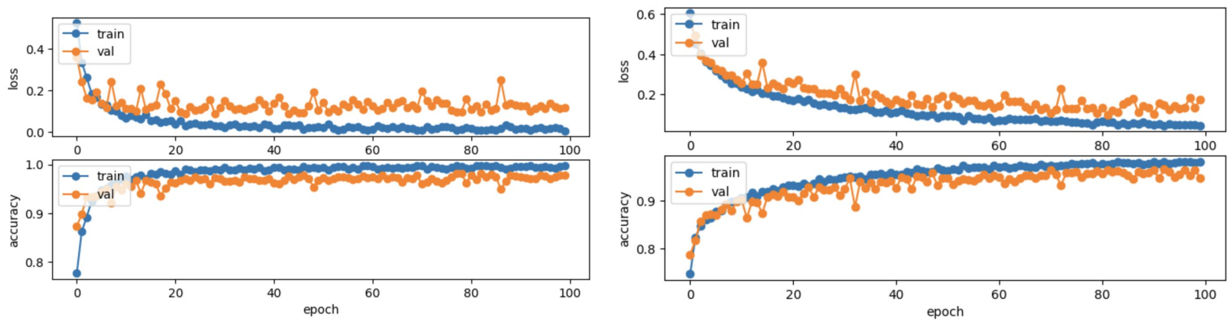


Figure 6 left: MCNN+ wavelet, right: MCNN

6. Conclusion

This report has presented a multi-scale convolutional neural network augmented with wavelet transformed data, which compared with two other models. The MCNN leverages both time and frequency domain features of ECG datasets. Moreover, the WT augmented dataset performs better than single MCNN, with higher classification accuracy, and less loss and larger testing accuracy with the same model run epochs.

As a result, we can demonstrate that wavelet transform augmented MCNN performs better than neural networks without data augmentation. Both wavelet transform and MCNN are powerful techniques for time series data analysis.

7. Share of Work

	Share of work
Writing the proposal	Shoaib Bin Masud (33%), Boyang Lyu (33%), Tiange Wang (33%)
Coding	Shoaib Bin Masud (40%), Boyang Lyu (30%), Tiange Wang (30%)
Running experiments, collecting data	Shoaib Bin Masud (30%), Boyang Lyu (30%), Tiange Wang (40%)
Discussion	Shoaib Bin Masud (33%), Boyang Lyu (33%), Tiange Wang (33%)
Making Slides	Shoaib Bin Masud (20%), Boyang Lyu (40%), Tiange Wang (40%)
Writing the final report	Shoaib Bin Masud (25%), Boyang Lyu (25%), Tiange Wang (50%)

References

[1] Machiwal, Deepesh, and Madan Kumar Jha. Hydrologic time series analysis: theory and practice. Springer Science & Business Media, 2012.

[2] Cui, Zhicheng, Wenlin Chen, and Yixin Chen. "Multi-scale convolutional neural networks for time series classification." arXiv preprint arXiv:1603.06995 (2016).

[3] Mao, Y., Chen, W., Chen, Y., Lu, C., Kollef, M. and Bailey, T., 2012, August. An integrated

data mining approach to real-time clinical monitoring and deterioration warning. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1140-1148).
[4] Xing, Z., Pei, J. and Keogh, E., 2010. A brief survey on sequence classification. ACM Sigkdd Explorations Newsletter, 12(1), pp.40-48.
[5] Lee, H., Pham, P., Largman, Y. and Ng, A., 2009. Unsupervised feature learning for audio classification using convolutional deep belief networks. Advances in neural information processing systems, 22, pp.1096-1104.
[6] Sifuzzaman, M., Islam, M.R. and Ali, M.Z., 2009. Application of wavelet transform and its advantages compared to Fourier transform.
[7] Rhif, M., Ben Abbes, A., Farah, I.R., Martínez, B. and Sang, Y., 2019. Wavelet transform application for/in non-stationary time-series analysis: a review. *Applied Sciences*, *9*(7), p.1345.
[8] https://www.mathworks.com/discovery/wavelet-transforms.html
[9] Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K. and Stanley, H.E., 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation, 101(23), pp.e215-e220.
[10]https://www.mathworks.com/help/wavelet/gs/continuous-wavelet-transform-and-scale-based-analysis.html
[11]https://blog.octo.com/en/time-series-features-extraction-using-fourier-and-wavelet-transforms-on-ecg-data/
[12] POLIKAR, ROBI. "The Wavelet Tutorial Second Edition Part I."
[13] Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). LSTM fully convolutional networks for time series classification. *IEEE access*, *6*, 1662-1669.
[14] Hüsken, M., & Stagge, P. (2003). Recurrent neural networks for time series classification. *Neurocomputing*, *50*, 223-235.
[15] Buyssens P., Elmoataz A., Lézoray O. (2013) Multiscale Convolutional Neural Networks for Vision–Based Classification of Cells. In: Lee K.M., Matsushita Y., Rehg J.M., Hu Z. (eds) Computer Vision – ACCV 2012. ACCV 2012. Lecture Notes in Computer Science, vol 7725. Springer, Berlin, Heidelberg.
[16] Harten, A., Hyman, J.M., Lax, P.D. and Keyfitz, B., 1976. On finite‐difference approximations and entropy conditions for shocks. Communications on pure and applied mathematics, 29(3), pp.297-322.
[17] Moody, G.B. and Mark, R.G., 2001. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine, 20(3), pp.45-50.
[18] https://www.mathworks.com/discovery/wavelet-transforms.html