L^1 and L^2 Regularized Deep Residual Network Model for Automated Detection of Myocardial Infarction (Heart Attack) Using Electrocardiogram Signals

Arijit Ukil[a], Leandro Marin[b] and Antonio J. Jara[c]

[a] TCS Research and Innovation, Tata Consultancy Services, Kolkata, India
[b] University of Murcia, Murcia, Spain
[c] HOP Ubiquitous (HOPU), Spain

Abstract

Automated interpretation of fundamental physiological signals like Electrocardiogram (ECG) plays an important role as part of smart healthcare ecosystem to build reliable and on-demand cardio-vascular diseases (CVD) screening system. Myocardial infarction is a life-threatening condition. Acute myocardial infarction leads to fatal condition commonly known as heart attack. It is medically established that myocardial infarction is often ignored initially, and treatment starts late. In this paper, we present an automated method of myocardial infarction detection from off-the-shelf single lead ECG signals so that early warning can be generated, and timely diagnosis can take place. We propose regularized deep neural network based model that is capable of classifying myocardial infarction condition from normal heart rhythm in single lead ECG signals. More precisely, we propose intensely regularized deep residual networks (ResNet) where both L^2 (also known as Tikhonov regularization) and L^1 (commonly known as Lasso) regularizations are used to construct a compact residual learning model. We demonstrate through empirical study on publicly available relevant ECG dataset from UCR timeseries archive that the proposed method demonstrates considerably superior performance over baseline methods and current state-of-the-art algorithms. We have also performed ablation study to depict the efficacy of the proposed intense regularization over only L^2 or L^1 regularizations.

Keywords 1

Deep Learning, regularization, residual networks, Electrocardiogram, time series, sensor, classification, automation.

1. Introduction

Electrocardiogram (ECG) is one of the fundamental markers for preliminary investigation of Cardio-Vascular Diseases (CVDs). Owing to the affordable availability of personal ECG Sensors like Alivecor [1], and wide-scale adoption of Internet of Things (IoT) infrastructure, remote diagnosis of critical CVD like myocardial infarction (commonly known as heart attack) is envisaged in smart clinical management system. In this paper, we propose a deep neural network based ECG classification algorithm, which gets executed in a typical cloud platform and the inference on the test or on-field ECG recording is shared to the nearest medical facility for automated emergency services. While, the ECG gets recorded by the individual at the comfort of their home, the analysis requires engagement of trained

Name and year of the event: Proceedings of the CIKM 2021 Workshops, November 01- 05, 2021
Editors of the proceedings (editors):
EMAIL: arijit.ukil@tcs.com (a); leandro@um.es (b); jara@hopu.org (c)

cardiologists, which is likely to hinder the scalability and larger scale basic diagnosis of the heart condition. Computerized analysis and automated interpretation of ECG signal paves the way to substantially reduce the frequency of clinician intervention and provides quick assessment of the heart condition. The future smart infrastructure facility is invariably necessitating remote diagnosis capability. Future smart cities with IoT based applications and eco-system demand such automated clinical management system that ensures on-demand diagnosis and quick response to the critical treatment requirements. In a general setup, patients interact with their smartphone applications to get associated in the smart healthcare systems and derive the benefit of remote assistance of healthcare facility with digital therapeutics.

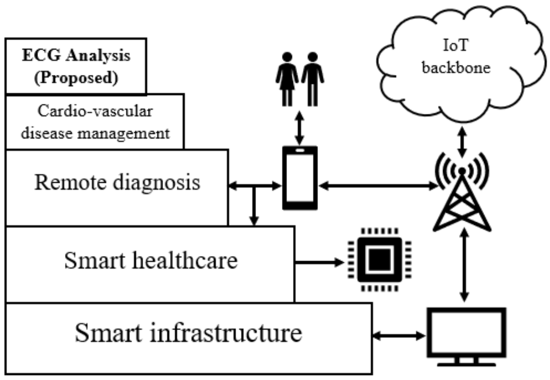

Figure 1: ECG analysis as a component of automated cardio-vascular management system. We depict the hierarchical relationship in smart infrastructure, which is an integral part of smart health care systems.

Computational analysis using machine learning methods for automated interpretation of digital ECG signal has been in the research foray from quite a few years. ECG is a typical time series signal. ECG classification problem is often considered as a time series classification task and consequently, attempts were made to solve it in the light of time series classification-based model generation. Typically, training model is developed from the given training examples. The trained model is evaluated on the test dataset. Classically, distance-based metrics like one nearest neighbour (1-NN) dynamic time warping (DTW) is used to understand the characteristics of the different class of signals [2]. Test signal is inferred based on the rule of minimum DTW distance of the training signals set. In fact, similarity measure is often considered a good approach for classification particularly in the area of data mining. DTW-based elastic distance measurement and 1-NN based classification is the conventional choice of baseline time series classification algorithm [3]. Lately, machine learning algorithms tailored for time series signals are evolve [3]. Collective of transformation ensembles (COTE) is an ensemble classifier in the time, autocorrelation, power spectrum and shapelet domains [4]. Another important time series classification algorithm is Bag of Symbolic Fourier Approximation (SFA) symbols (BOSS) [5]. BOSS considers a truncated Discrete Fourier Transform (DFT) in extracting features through sliding windowed time series. These algorithms namely, COTE and BOSS are dependent on the customized feature space development and specific pattern recognition.

Recently Residual Network (ResNet) have been emerged to solve time series classification problem. In fact, ResNet based approach seems to be a promising one [6]. Similarly, we get inspired by the promise of the capability of deep residual network's representation learning by residual mapping, we propose a regularized deep residual neural network with both L^2 and L^1 penalties such that a compact learned model can be constructed. We investigate on publicly available ECG datasets with normal rhythm and myocardial infarction labels and comparative study indicates superior performance of our proposed method over baseline algorithms and state-of-the-art methods.

2. Proposed deep residual network architecture

Myocardial infarction detection from ECG signal considers number of input ECG signals as part of the overall training space. We can represent each of them as: $\mathcal{X} = [x_1, x_2, x_3, \dots, x_T] \in \mathbb{R}^T$ is an ordered set of real values T number of time steps at each of the training samples. Each of the training samples is associated with a with class label $y \in [NR, MI]$, where NR indicates normal sinus rhythm and MI indicates myocardial infarction. Thus, the training dataset consists of labeled ECG signals as $\Omega = \{\mathcal{X}^n, y^n\}$, for $n =$

$1,2,3, \dots, N$ with total N number of training examples.

Typically, normal sinus rhythm consists of regular ECG morphology while ECG signal of myocardial infarction patients have different morphology [7] as shown in Figure 2.

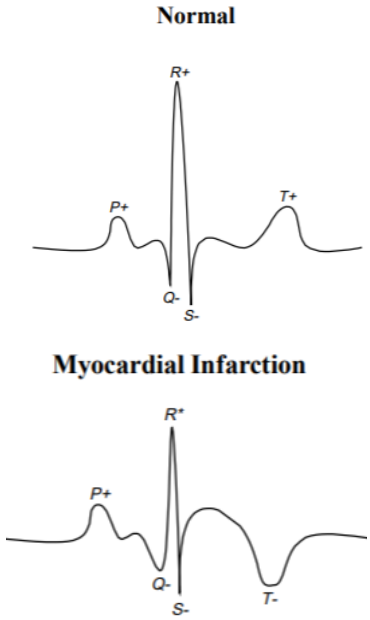

Figure 2: Morphological differences in ECG signals for normal sinus rhythm and myocardial infarction.

In order to extract the distinctive features from ECG signal to distinguish normal rhythm and myocardial infarction, we use deep residual network based representation learning [8].

The immense success of deep residual networks for computer vision applications motivates us. In deep residual networks, the underlying layers fit a residual mapping instead of directly stacking layers to overcome learning degradation problem through layer-wise recursive learning using skip-connection [6]. Let us consider a deep neural network that consists of L number of layers, where each of the layer implements a non-linear transformation $\mathcal{H}_l(\cdot)$, where a layer is indexed by l. In accordance to the ResNet topology as shown in Figure 3, the residual mapping $\mathcal{R}_l$ of l^{th} layer is denoted as:

$$\mathcal{R}_l = \mathcal{R}_{l-1} + \mathcal{H}_l(\mathcal{R}_{l-1})$$

The main intuition of a residual network is that the skip connection or short-circuit connection provides direct path of input signal propagation and thus, it helps to prevent the phenomenon of propagation loss of signal characteristics.

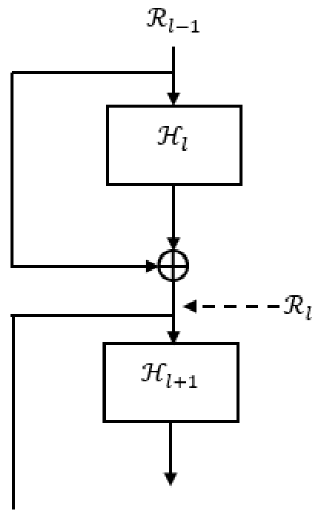

Figure 3: Topology and basic transformation and layer mapping in typical deep residual network model.

We can conceptualize residual networks as micro-structures of residual blocks connected sequentially along with skip connection as shown in Figure 3.

We construct a pretty deep residual neural network with total fifteen convolution layers. There are three residual blocks. First residual block consists of five convolution layers. There are eight convolution layers in second residual block, while third residual block consists of six convolution layers. Including the three skip connection layers, total twenty layers are present in the proposed deep learning model, with nineteen effective convolution layers.

We have used Batch Normalization (BN) and Rectified Linear Unit (ReLU) as the activation layer. After the last convolution layer, we use Global Average Pooling, which acts a structural regularizer. In fact, it natively prevents overfitting for the overall structure with added advantage of no need for hyper-parameter optimization [9]. Softmax activation function is used at the output layer for the classification purpose along with cross-entropy as the loss function. We depict the proposed model in Figure 4.

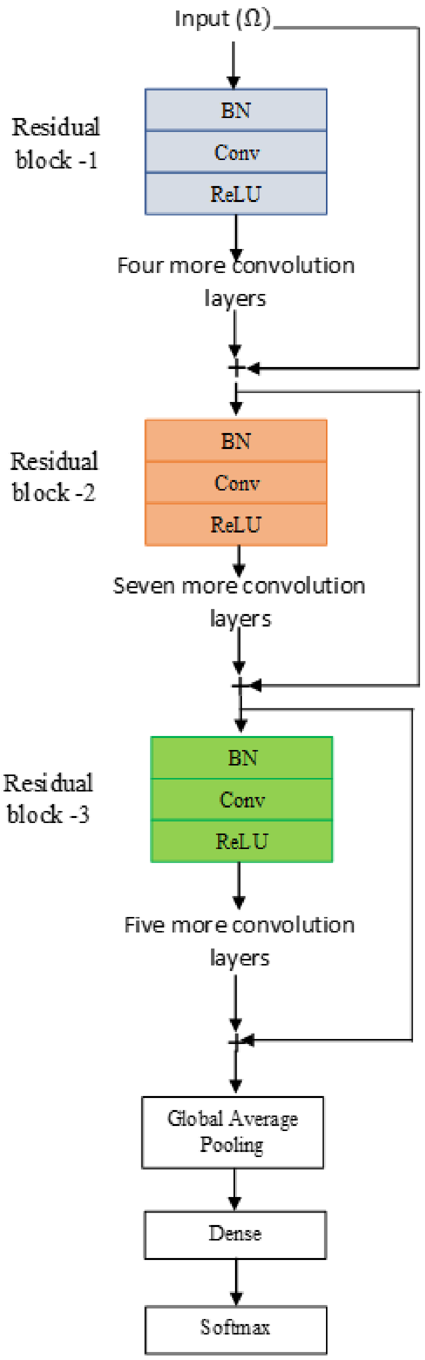

Figure 4: Deep residual network model for ECG classification.

It is understood that ECG signal annotation is an expensive process and it is likely to have scarce number of training examples. While the proposed deep neural network has a sophisticated capability of representation learning over the training space, there exists higher chance of overfitted model generation. In order to eliminate the risk of overfitting to the training examples, we regularize the model using L^2 and L^1 regularizations [10]. The regularized cost function J^* with network parameters θ is:

$$J^*(\theta) = J(\theta) + \lambda J^{pen}(\theta)$$

Where, $J^{pen}(\theta)$ is the penalty function and λ is the regularization co-efficient, with $\lambda \in [0, \infty]$. The L^2 or Tikhonov regularization is expressed as:

$$J^*_{L^2}(\theta) = J(\theta) + \lambda_2 \frac{\theta^{\mathrm{T}}\theta}{2}$$

Where, $J^{pen}(\theta) = \frac{\theta^{\mathrm{T}}\theta}{2}$.

Where, the network parameter gradient is:

Similarly, the Lasso or L^1 regularization is defined as:

$$J^*_{L^1}(\theta) = J(\theta) + \lambda_1 \|\theta\|_1$$

Where, λ_2 and λ_1 are L^2 regularization and L^1 regularization factors respectively.

Our proposed cost function is:

$$J^*_{proposed}(\theta) = J(\theta) + \lambda_2 \frac{\theta^{\mathrm{T}}\theta}{2} + \lambda_1 \|\theta\|_1$$

We intend to emphasize that regularization impacts on network parameters for L^2 and L^1 regularizations are different. With or L^1 regularization, we have sparser parameter matrix while L^2 regularization clips the values of the parameters. Thus, a sparser yet controlled network parameter space model is constructed.

3. Experimental Analysis and Results

We consider relevant dataset ECG200 from publicly available UCR time series archive [11]. The intent is to construct a classification model that demonstrates better performance accuracy than state-of-the-art algorithms like DTW_R1_1NN [2], BOSS [5], ResNet [6], HIVE-COTE [4]. From UCR time series archive, the representative dataset ECG 200 is considered. ECG200 dataset is binary-labelled with normal heartbeat and myocardial infarction classes. There are distinct 100 instances for training and 100 instances for testing purposes.

For this experimentation purpose, we consider the hyperparameters as described in Table 1.

Table 1
Hyperparameter description.

Parameter	Brief explanation	Value/ Type
Epoch	Number of training iterations	60
Optimizer	Learning rate optimization	Adam
Batch size	Number of training samples in each of the pass	$min\left(\left\lfloor\frac{T}{10}\right\rfloor, 16\right)$ where T is the number of time steps at each instant
Number of residual blocks	Total number of residual blocks	3
Number of convolution layers at each of the residual blocks	Residual block #1	5
	Residual block #2	8
	Residual block #3	6
Kernel size	Residual block #1	{15,12, 8, 5, 3}
	Residual block #2	{15,10,8,7, 6, 5, 4, 3}
	Residual block #3	{15,10,8,7,5,3 }
Number of filters	Residual block #1	{64, 64, 64,64,64}
	Residual block #2	{128, 128, 128, 128, 128,128,128, 128}
	Residual block #3	{128, 128, 128, 128,128,128}
λ_1	L^1 regularization factor	0.01
λ_2	L^2 regularization factor	0.10

In Table 2, we illustrate the performance efficacy of the proposed regularized deep residual network over ECG200 dataset in comparison with the baseline methods and state-of-the-art algorithms. depict the experimental results of our proposed method. First the model is generated with the provided training dataset and subsequently, the trained model is used to predict the classification of the test dataset. The proposed model performs better than the baseline and outperforms the state-of-the-art algorithms. Currently, BOSS [5] is the benchmark algorithm, and our proposed method shows 2% test accuracy gain ver BOSS [5] and 4% test accuracy gain over ResNet [6]. In fact, our method demonstrates substantial performance gain with similar type of deep learning model- ResNet [6].

Table 2
Performance comparison of our proposed method and related state-of-the-art classification algorithms in terms of test accuracy metric over ECG200 dataset from UCR time series archive.

Algorithm	Test accuracy
DTW_R1_1NN [2]	0.77
HIVE-COTE [4]	0.85
ResNET [6]	0.87
BOSS [5]	0.89
Our method	0.91

In Table 3, we demonstrate ablation study results when L^2 or L^1 regularizations are performed, i.e. when $J^*(\theta) = J^*_{L^2}(\theta)$ or $J^*(\theta) = J^*_{L^1}(\theta)$, whereas our proposed ResNet has cost function as: $J^*(\theta) = J^*_{proposed}(\theta)$.

Table 3
Ablation study to depict the efficacy of the proposed method, showing that L^1 and L^2 regularization is more effective than L^1 or L^2 regularizations.

Regularization	Test accuracy
L^1	0.80
L^2	0.86
Our method (L^1 and L^2)	0.91

4. Plausible Deployment Architecture

We illustrate a plausible deployment architecture of remote diagnosis of myocardial infarction (CVDs in larger scope) in Figure 5. The proposed ECG analysis algorithm is supposed to be hosted over cloud and a smartphone-based application interacts with the user to send the captured ECG signal to the cloud. For ECG signal capturing purpose, off-the-shelf sensors like Alivecor can be used. The cloud server responds the inference result to the user through the same smartphone application as well as alerts nearby medical facilities if abnormal cardiac activity is detected and myocardial infarction is suspected. Thus, a complete diagnosis eco-system can be built using the proposed method, which will have immense benefit in providing early-warning alarms to initiate necessary medical attention. However, it is imperative to mention that physiological signals like ECG consists of potential sensitive information and appropriate trust [13, 15], data security [14, 17] and data privacy [12, 16, 17] mechanisms are to be implemented in order to ensure wide-scale acceptability in the public domain.

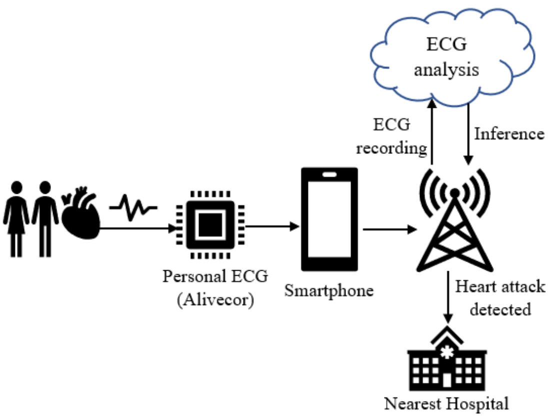

Figure 5: Smart healthcare ecosystem, with plausible deployment architecture that integrates the ECG analytics with user and medical facilities for automated CVD detection.

5. Conclusion

In this paper, our focus is to develop a classification algorithm to reliably detect myocardial infarction or heart attack condition using ECG signal. The performance of the proposed model is superior than the currently available relevant methods. We are confident that ECG analytics as part of a smart healthcare ecosystem will ensure the necessary impetus for human-centered purpose in contributing better quality of life through the development of technology-driven early detection and treatment for life-threatening CVDs like myocardial infarction.

6. Acknowledgements

Leandro Marin is partially funded by Project PID2020-112675RB-C44 funded by MCIN/AEI/10.13039/501100011033.

Antonio J. Jara is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 732679, ACTIVAGE project https://www.activageproject.eu/.

7. References

[1] https://www.alivecor.com/

[2] R.J. Kate, "Using Dynamic Time Warping Distances as Features for Improved Time Series Classification," *Data Mining and Knowledge Discovery*, pp. 283–312, 2016.

[3] A. Bagnall, J. Lines, A. Bostrom, J. Large and E. Keogh, "The Great Time Series Classification Bake Off: A Review and Experimental Evaluation of Recent Algorithmic Advances," *Data Mining and Knowledge Discovery*, volume 31, pp. 606–660, 2017.

[4] A. Bagnall, J. Lines, J. Hills, A. Bostrom, "Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles," *IEEE Transaction on Knowledge and Data Engineering*, pp. 2522–2535, 2015.

[5] P. Schäfer, "The BOSS Is Concerned with Time Series Classification in The Presence of Noise," *Data Mining and Knowledge Discovery*, pp. 1505–1530, 2015.

[6] Z. Wang, W. Yan. And T. Oates, "Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline," *International Joint Conference on Neural Networks (IJCNN)*, pp. 1578-1585, 2017.

[7] R. T. Olszewski, "Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data," February 2001, https://apps.dtic.mil/sti/pdfs/ADA457624.pdf

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *CVPR*, 2016.

[9] M. Lin, Q. Chen, and S. Yan, "Network in Network," *arXiv preprint, arXiv:1312.4400*, 2013.

[10] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning,” *The MIT Press*, 2016.
[11] A. Bagnall, J. Lines, W. Vickers and E. Keogh, “The UEA & UCR Time Series Classification Repository,” www.timeseriesclassification.com
[12] A. Ukil, J. Sen, “Secure Multiparty Privacy Preserving Data Aggregation by Modular Arithmetic,” International Conference On Parallel, Distributed and Grid Computing (PDGC), pp. 344-349, 2010.
[13] A. Ukil, “Secure Trust Management in Distributed Computing Systems,” IEEE International Symposium on Electronic Design, Test and Application, pp. 116-121, 2011.
[14] J. Sen, A. Ukil, “A Secure Routing Protocol for Wireless Sensor Networks,” International Conference on Computational Science and Its Applications, pp. 277-290, 2010.
[15] A. Ukil, “Trust and Reputation Based Collaborating Computing in Wireless Sensor Networks,” IEEE International Conference on Computational Intelligence, Modelling and Simulation, pp. 464-469, 2010.
[16] A. Ukil, A. J Jara, L. Marin, “Data-Driven Automated Cardiac Health Management with Robust Edge Analytics and De-Risking,” Sensors, Volume 19, Issue 12, 2019.
[17] Arijit Ukil, “Security and Privacy In Wireless Sensor Networks,” Smart Wireless Sensor Networks, December, 2010.