Master Thesis:
Combining multi-scale kernels and transformer encoder for ECG classification

Kilian Kramer

Dr. Pietro Bonizzi
Dr. Stef Zeemering
Dr. Joël Karel

Maastricht University

Index

Problem Introduction

- Cardiovascular diseases, such as Atrial Fibrillation, are among the leading causes of death in the population (32% of deaths world-wide 2019 WHO and 42.5% in European region WHO) resulting in an increased demand for efficient, timely, and reliable cardiac assessment

- Today, much research and accurate models are available for simpler arrhythmia classification tasks, e.g. binary AFIB (Atrial Fibrillation) classification or grouped common arrhythmia types (i.e. AAMI standard)

- Problem: Professional treatment requires individual and detailed ECG assessment

Problem Introduction

- In recent years, Transformer models have gained considerable popularity due to the attention mechanism and research shows promising results when these models are applied to less comprehensive ECG arrhythmia classification tasks

- Research about Transformer models on comprehensive arrhythmia classification is still limited

Thesis objective

- Question in mind: Are Transformer models also effectively for comprehensive (26 classes) ECG classification and show improved performance?
- Detailed performance evaluation of Transformer models with some other SOTA deep-learning approaches
- Tasks: single-lead and multi-lead ECG classification, a subset of four specific arrhythmia classes and the generalisation ability across datasets

Thesis objective

- Training and evaluation: Public Physionet 2021 challenge database, which consists of 88.251 12-lead ECGs with multi-label annotations of more than 100 arrhythmia diseases (here 26 classes are used, which are scored within the challenge)

- Evaluation across datasets: Internal MyDiagnostick dataset, provided by the Department of Physiology at Maastricht University, annotations for Sinus Rhythm, Atrial Fibrillation, Atrial Flutter, Premature Atrial and Premature Ventricular Contractions

Research Questions

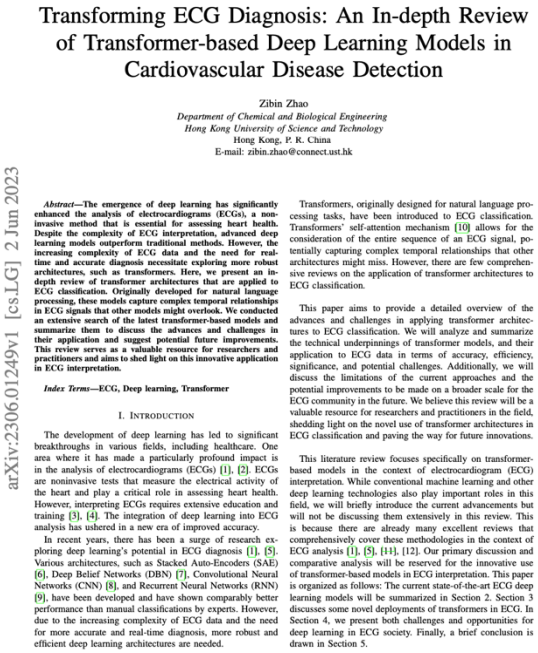
1. How well does a Transformer-based model perform on the Physionet 2021 challenge data compared to a feature-based model or a Convolutional Network?

2. Can an ensemble Transformer model and Convolutional Network effectively capture spatio-temporal information from multi-lead ECGs and improve accuracy?

3. How is the performance of the proposed approach at discriminating SR, AF, AFL, PAC and PVC?

4. What are the challenges in transferring the pre-trained models from the Physionet 2021 challenge data to the MyDiagnostick database? Do the models generalise well, even though different ECG devices were used?

Related Work: Recent (2023) comprehensive review about Transformer models applied to ECG classification



Transforming ECG Diagnosis: An In-depth Review of Transformer-based Deep Learning Models in Cardiovascular Disease Detection

Zibin Zhao
Department of Chemical and Biological Engineering
Hong Kong University of Science and Technology
Hong Kong, P. R. China
E-mail: zibin.zhao@connect.ust.hk

***Abstract*—The emergence of deep learning has significantly enhanced the analysis of electrocardiograms (ECGs), a non-invasive method that is essential for assessing heart health. Despite the complexity of ECG interpretation, advanced deep learning models outperform traditional methods. However, the increasing complexity of ECG data and the need for real-time and accurate diagnosis necessitate exploring more robust architectures, such as transformers. Here, we present an in-depth review of transformer architectures that are applied to ECG classification. Originally developed for natural language processing, these models capture complex temporal relationships in ECG signals that other models might overlook. We conducted an extensive search of the latest transformer-based models and summarize them to discuss the advances and challenges in their application and suggest potential future improvements. This review serves as a valuable resource for researchers and practitioners and aims to shed light on this innovative application in ECG interpretation.**

***Index Terms*—ECG, Deep learning, Transformer**

I. Introduction

The development of deep learning has led to significant breakthroughs in various fields, including healthcare. One area where it has made a particularly profound impact is in the analysis of electrocardiograms (ECGs) [1], [2]. ECGs are noninvasive tests that measure the electrical activity of the heart and play a critical role in assessing heart health. However, interpreting ECGs requires extensive education and training [3], [4]. The integration of deep learning into ECG analysis has ushered in a new era of improved accuracy.

In recent years, there has been a surge of research exploring deep learning's potential in ECG diagnosis [1], [5]. Various architectures, such as Stacked Auto-Encoders (SAE) [6], Deep Belief Networks (DBN) [7], Convolutional Neural Networks (CNN) [8], and Recurrent Neural Networks (RNN) [9], have been developed and have shown comparably better performance than manual classifications by experts. However, due to the increasing complexity of ECG data and the need for more accurate and real-time diagnosis, more robust and efficient deep learning architectures are needed.

Transformers, originally designed for natural language processing tasks, have been introduced to ECG classification. Transformers' self-attention mechanism [10] allows for the consideration of the entire sequence of an ECG signal, potentially capturing complex temporal relationships that other architectures might miss. However, there are few comprehensive reviews on the application of transformer architectures to ECG classification.

This paper aims to provide a detailed overview of the advances and challenges in applying transformer architectures to ECG classification. We will analyze and summarize the technical underpinnings of transformer models, and their application to ECG data in terms of accuracy, efficiency, significance, and potential challenges. Additionally, we will discuss the limitations of the current approaches and the potential improvements to be made on a broader scale for the ECG community in the future. We believe this review will be a valuable resource for researchers and practitioners in the field, shedding light on the novel use of transformer architectures in ECG classification and paving the way for future innovations.

This literature review focuses specifically on transformer-based models in the context of electrocardiogram (ECG) interpretation. While conventional machine learning and other deep learning technologies also play important roles in this field, we will briefly introduce the current advancements but will not be discussing them extensively in this review. This is because there are already many excellent reviews that comprehensively cover these methodologies in the context of ECG analysis [1], [5], [11], [12]. Our primary discussion and comparative analysis will be reserved for the innovative use of transformer-based models in ECG interpretation. This paper is organized as follows: The current state-of-the-art ECG deep learning models will be summarized in Section 2. Section 3 discusses some novel deployments of transformers in ECG. In Section 4, we present both challenges and opportunities for deep learning in ECG society. Finally, a brief conclusion is drawn in Section 5.

- Transforming ECG Diagnosis: An In-depth Review of Transformer-based Deep Learning Models in Cardiovascular Disease Detection (Zhao, 2023)
- Most models focus on binary classification (Atrial Fibrilliation) or 8 grouped arrhythmia classes, i.e. AAMI standard (Association for the Advancement of Medical Instrumentation): Normal Beat (N), Supraventricular Ectopic Beat (S), Ventricular Ectopic Beat (V), Fusion Beat (F), Atrial Fibrillation (AF), Junctional Escape Beat (J), Unknown Beat (U),Others (O)
- Report very high accuracies (99%) on the MIT-BIH dataset
- Transformer models often in combination with Convolutional Networks applied (limited receptive field for capturing fine-granular features + capturing temporal relations)
- Transformer models used for capturing spatio-temporal information from multi-lead ECGs, e.g. Hao & Nugroho

Related Work: Winning paper of Physionet 2021 challenge

Classification of ECG Using Ensemble of Residual CNNs with Attention Mechanism

Petr Nejedly, Adam Ivora, Radovan Smisek, Ivo Viscor, Zuzana Koscova, Pavel Jurak, Filip Plesinger

Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

Abstract

This paper introduces a winning solution (team ISIBrno-AIMT) to the PhysioNet Challenge 2021. The method is based on the ResNet deep neural network architecture with a multi-head attention mechanism for ECG classification into 26 independent groups. The model is optimized using a mixture of loss functions, i.e., binary cross-entropy, custom challenge score loss function, and sparsity loss function. Probability thresholds for each classification class are estimated using the evolutionary optimization method. The final model consists of three submodels forming a majority voting classification ensemble. The proposed method classifies ECGs with a variable number of leads, e.g., 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead. The algorithm was validated and tested on the external hidden datasets (CPSC, G12EC, undisclosed set, UMich), achieving a challenge score 0.58 for all tested lead configurations. The total training time was approximately 27 hours, i.e., 9 hours per model. The presented solution was ranked first across all 39 teams in all categories.

1. Introduction

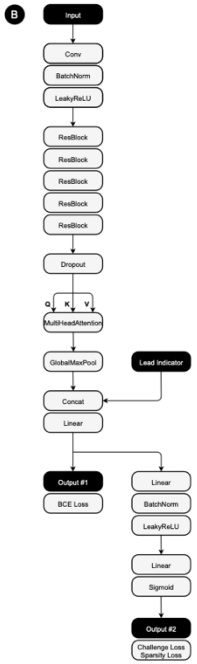
Cardiovascular diseases are the most common cause of death globally, reaching 32 percent by 2019 [1]. Heart disorders are usually analyzed using electrocardiographic signals (ECG) at a length of 10-60 seconds, acquired from the body surface. The ECG signal shows the electrical activity of heart atria and ventricles and, therefore, informs about heart rhythm and a beat morphology.

Current automated algorithms to analyze the ECG signal are based on machine-learning (using expert features) or deep-learning methods. A specialty of deep-learning methods is that they extract features by themselves during a training process from a raw or transformed ECG signal. These deep-learning methods are usually based on convolutional layers and are called convolutional neural networks (CNN). A need to train very complex CNNs led to the invention of the Residual Networks (ResNet) architecture [2], implementing residual blocks to improve gradient propagation during training.

For this Computing in Cardiology 2021 challenge [3] we introduce a solution using an ensemble of a custom variant of ResNet neural networks accompanied by a multihead attention mechanism. This solution arises from our last year Computing in Cardiology 2020 challenge solution [4], where we were using ResNet-GRU with attention mechanism [5]. With this method, we were able to achieve an acceptable validation score, however, the final test scores showed poor performance while testing on the undisclosed testing database. This indicated that our method was able to classify data that originated from the same hospital very well, however, generalizability for other institutions was missing. This year's solution tries to improve drawbacks from previous years by introducing several changes to our method.

The investigation of literature from previous year solutions [6] led us to change the preprocessing step by introducing data filtering to maximize generalizability across the institutions. The filtering should minimize the ECG frequency band as much as possible (for the cost of possibly discarding some ECG information that might be useful). We believe that this might be a good idea since we are not aware of data quality, types of artifacts, and distortions in undisclosed testing sets. Secondly, we utilize z-score normalization, while last year, we were using physical units in mV. In addition, models are trained using a custom loss function which consists of three parts, i.e., cross-entropy, custom challenge loss, and custom sparsity loss.

The custom challenge loss optimization was proposed by [7], where the continuous equivalent of binary OR operator was used to design differentiable approximation of challenge metric. This helps the model to learn the similarity of diagnoses. Next, we introduce the custom sparsity loss, which forces the model to output probability values close to 0 or 1, which helps with the final threshold optimization to binarize the data output. Lastly, the class-specific thresholds are found using differential evolution genetic algorithms. The final model consists of three subunits creating the model ensemble.



(c) Full model architecture.

(applies 15 and 9 large convolutions)

Follow-up study findings:
Multi-head Attention block did not improve performance

IOP Publishing Physiol. Meas. 43 (2022) 044001 https://doi.org/10.1088/1361-6579/ac647c

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

CrossMark

PAPER

Classification of ECG using ensemble of residual CNNs with or without attention mechanism

RECEIVED 7 January 2022
REVISED 16 March 2022
ACCEPTED FOR PUBLICATION 5 April 2022
PUBLISHED 28 April 2022

Petr Nejedly, Adam Ivora, Ivo Viscor, Zuzana Koscova, Radovan Smisek, Pavel Jurak and Filip Plesinger

Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

E-mail: nejedly@isibrno.cz

Keywords: ECG, classification, deep learning, PhysioNet challenge 2021, attention mechanism

Abstract

Objective. This paper introduces a winning solution (team ISIBrno-AIMT) to the official round of PhysioNet Challenge 2021. The main goal of the challenge was a classification of ECG recordings into 26 multi-label pathological classes with a variable number of leads (e.g. 12, 6, 4, 3, 2). The main objective of this study is to verify whether the multi-head-attention mechanism influences the model performance. *Approach.* We introduced an ECG classification method based on the ResNet architecture with a multi-head attention mechanism for the official round of the challenge. However, empirical findings collected during model development suggested that the multi-head attention layer might not significantly impact the final classification performance. For this reason, during the follow-up round, we removed a multi-head attention layer to test the influence on model performance. Like the official round, the model is optimized using a mixture of loss functions, i.e. binary cross-entropy, custom challenge score loss function, and custom sparsity loss function. Probability thresholds for each classification class are estimated using the evolutionary optimization method. The final architecture consists of three submodels forming a majority voting classification ensemble. *Main results.* The modified model without the multi-head attention layer increased the overall challenge score to 0.59 compared to the 0.58 from the official round. *Significance.* Our findings from the follow-up submission support the fact that the multi-head attention layer in the proposed architecture does not significantly affect the classification performance.

1. Introduction

Cardiovascular diseases are the most common reason for death by World Health Organization (WHO), reaching 32 percent in 2019. Several disorders of heart rhythm are directly life-threatening as ventricular tachycardia (VT), ventricular fibrillation (VF), or long pauses, and others. However, several other disorders and arrythmias make life just less comfortable when they occur at first. Nevertheless, with increasing time, these none life-threatening arrhythmias increase the risk of heart failure or stroke, as in the case of atrial fibrillation (AF) or atrial flutter (AFL). Also, several disorders may affect the heart conduction system while the rhythm stays untouched for some time. These disorders affect signal propagation to ventricular myocardial cells, as the Left Bundle Branch Block (LBBB) or the Right Bundle Branch Block (RBBB). These conduction abnormalities change morphology, prolong the QRS complex, and lower ejection fraction. Fibrosis and sclerosis of the conduction system may also lead to so-called atrioventricular blocks (AVB I/IIa/IIb/III), ranging from prolonged P-Q interval to complete dissociation between atrial and ventricular activation.

Most of the known heart disorders may be captured using the electrocardiogram (ECG). The CinC/PhysioNet challenge 2020 (Perez Alday *et al* 2020) demonstrated that machine-learning approaches could reliably classify ECG recorded in standard 12-lead setup into 26 classes. However, while the 12-lead configuration is routinely used in clinical ambulatory practice, another arrangements exists, and they are used, for example, in telemedicine. In this field, many devices record only a limited number of leads (1–3) as wearables

Related Work: 2nd and 4th ranked Physionet 2021 challenge papers

Towards High Generalization Performance on Electrocardiogram Classification

Hyeongrok Han[†1], Seongjae Park[†2], Seonwoo Min[1,3], Hyun-Soo Choi[4], Eunji Kim[1], Hyunki Kim[1], Sangha Park[1], Jinkook Kim[2], Junsang Park[2], Junho An[2], Kwanglo Lee[2], Wonsun Jeong[2], Sangil Chon[2], Kwonwoo Ha[2], Myungkyu Han[2], Sungroh Yoon[*1,5]

[1]Department of Electrical and Computer engineering, Seoul National University, Seoul, South Korea
[2]HUINNO Co., Ltd,. Seoul, South Korea
[3]LG AI Research, Seoul, South Korea
[4]Department of Computer Science and Engineering, Kangwon National University, Chuncheon, South Korea
[5]Department of Biological Sciences, Interdisciplinary Program in Bioinformatics, Interdisciplinary Program in Artificial Intelligence, ASRI, INMC, and Institute of Engineering Research, Seoul National University, Seoul, South Korea

Abstract

Recently, many electrocardiogram (ECG) classification algorithms using deep learning have been proposed. The characteristics of ECG vary from dataset to dataset for various reasons (i.e., hospital, race, etc.). Therefore, it is important for a model to have high generalization performance consistently over all datasets. In this paper, as part of the PhysioNet / Computing in Cardiology Challenge 2021, we present a model developed to classify cardiac abnormalities from 12 lead and reduced-lead ECGs. In particular, to upgrade our previous model for improving generalization performance, we newly adopt constant-weighted cross-entropy loss, additional features, Mixup augmentation, and squeeze/excitation block, OneCycle learning rate scheduler, which are selected via evaluation of generalization performance using leave-one-dataset-out cross-validation setting. With the present model, our DSAIL_SNU team has received challenge scores of 0.55, 0.58, 0.58, 0.57 and 0.57 (ranked 2nd, 1st, 1st, 2nd, 2nd out of 39 teams) for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead versions of the hidden test set, respectively. The present model achieves higher generalization performance over all versions of the hidden test set than the model submitted last year.

1. Introduction

Electrocardiogram (ECG) is an important tool for diagnosing cardiac abnormalities, and more than 300 million ECGs are obtained worldwide each year [1]. Standard ECGs, which are used to diagnose heart diseases, consist of 12 leads. However, it is not always possible to obtain all 12 leads due to the cost and limitations of measurement devices. Recently, it has been demonstrated that a subset of 12 leads also contains sufficiently meaningful information [2].

In last year, we developed a model to classify clinical cardiac abnormalities from 12-lead ECG classification [3]. Although our model showed a high challenge score on the validation dataset, it showed a much lower score on the hidden test dataset due to the lack of dataset-wise generalization performance. The characteristics of ECG vary from dataset to dataset for various reasons, *i.e., hospital, race, etc.* It is important to design a model to have generalization performance on dataset unseen during training. Therefore, it is necessary to check whether the proposed model shows high generalization performance consistently over various datasets.

In this paper, as part of the PhysioNet / Computing in Cardiology Challenge 2021, we present a model developed to classify cardiac abnormalities from 12 and reduced-lead ECGs [4–6]. To get the present model having high generalization performance, we have attempted various techniques and evaluated them using the leave-one-dataset-out cross-validation for model selection. The present model achieves 0.1 higher challenge score in average over all test set versions than the model we submitted last year [3].

2. Methods

2.1. Data

Table 1 shows the statistics of the data provided by the challenge with 26 scored SNOMED-CT labels [13] from

†: equal contribution (Hyeongrok Han and Seongjae Park)
*: corresponding author (Sungroh Yoon)

Computing in Cardiology 2021; Vol 48 Page 1 ISSN: 2325-887X DOI: 10.22489/CinC.2021.080

2nd

Channel Self-Attention Deep Learning Framework for Multi-Cardiac Abnormality Diagnosis from Varied-Lead ECG Signals

Apoorva Srivastava[1], Ajith Hari[1], Sawon Pratiher[1], Sazedul Alam[2], Nirmalya Ghosh[1], Nilanjan Banerjee[2], Amit Patra[1]

[1]*Department of Electrical Engineering, Indian Institute of Technology Kharagpur, West Bengal, India.*
[2] *Computer Science and Electrical Engineering, University of Maryland-Baltimore County, Maryland, USA.*

Abstract

*Electrocardiogram (ECG) signals are widely used to diagnose heart health. Experts can detect multiple cardiac abnormalities using the ECG signal. In a clinical setting, 12-lead ECG is mainly used. But using fewer leads can make the ECG more pervasive as it can be integrated with wearable devices. At the same time, we need to build systems that can diagnose cardiac abnormalities automatically. This work develops a channel self-attention-based deep neural network to diagnose cardiac abnormality using a different number of ECG lead combinations. Our approach takes care of the temporal and spatial interdependence of multi-lead ECG signals. Our team participates under the name "**cardiochallenger**" in the "PhysioNet-Computing in Cardiology Challenge 2021". Our method achieves the challenge metric score of 0.55, 0.51, 0.53, 0.51, and 0.53 (ranked 2nd, 5th, 4th, 5th and 4th) for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead cases, respectively, on the test data set.*

1. Introduction

With over 17.9 million deaths, cardiovascular diseases are the leading cause of mortality worldwide [1]. The heart's activity from different angles can be studied from a 12-lead ECG. Detection of multiple cardiac abnormalities like coronary occlusion, myocardial infarction, etc., can be done using a 12-lead ECG.

Early-stage prognosis and timely interventions aid clinicians in identifying different cardiac irregularities and provide improved clinical outcomes. The PhysioNet/CinC-2021 challenge is dedicated to cardiac abnormality classification (CAC) from 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECG recordings [2]. Early and accurate detection of diseases with fewer leads makes ECG greater pervasive as it can be incorporated with wearable devices. Conventional CAC methods regularly employ machine learning models on the extracted domain-aware handcrafted features using raw ECG signal processing. Of late deep learning (DL) methods have democratized the CAC task with superior performance [3],[4],[5]. DL models can abstract explanatory ECG feature representations in an automated fashion and predict CACs in an end-to-end manner [6],[7]. This paper proposes an attention-based DL model, which will help medical practitioners judiciously inspect and categorize the inter-beat and intra-beat patterns. The proposed model acknowledges the spatial interrelation among the channels and the important temporal segments of the signal.

The rest of the paper is organized as follows. Section 2 summarizes the data pre-possessing and our channel self-attention-based DL model. Experimental results are discussed in sections 3 and 4. Section 5 concludes the paper.

2. Methodology

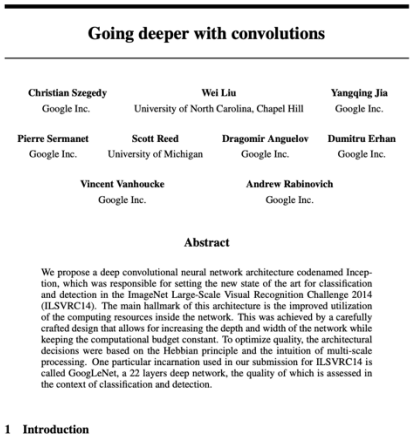
Cardiac abnormality detection using ECG signals can be formulated as a time-series classification problem. We aim to detect 29 multi-labeled cardiac abnormalities along with sinus rhythm using varying lead ECG signals [2]. The model is trained on 12-lead ECG and tested on:

- 12-lead: I,II,III,aVR,aVL,aVF,V1,V2,V3,V4,V5,V6
- 6-lead: I,II,III,aVR,aVL,aVF
- 4-lead: I,II,III,V2
- 3-lead: I,II,V2
- 2-lead: I,II

In this paper, a channel self-attention (CA)-based framework, as depicted in Figure 1,is proposed for the diagnosis of multi-labeled cardiac abnormalities. The model is inspired by squeeze and excitation network [8]. The global spatial information is squeezed, and channel-wise statistic is generated by CA framework. Higher weight is given to the more imperative channel, which leads to enhanced performance. Here, it is applied with the inception and residual neural model. In the following section, we provide a detailed description of the system's components.

4th (applies 3, 4 and 5 convolutions in parallel)

Related Work: Multi-scale convolutions and channel-attention



Going deeper with convolutions

Christian Szegedy
Google Inc.

Wei Liu
University of North Carolina, Chapel Hill

Yangqing Jia
Google Inc.

Pierre Sermanet
Google Inc.

Scott Reed
University of Michigan

Dragomir Anguelov
Google Inc.

Dumitru Erhan
Google Inc.

Vincent Vanhoucke
Google Inc.

Andrew Rabinovich
Google Inc.

Abstract

We propose a deep convolutional neural network architecture codenamed Inception, which was responsible for setting the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main hallmark of this architecture is the improved utilization of the computing resources inside the network. This was achieved by a carefully crafted design that allows for increasing the depth and width of the network while keeping the computational budget constant. To optimize quality, the architectural decisions were based on the Hebbian principle and the intuition of multi-scale processing. One particular incarnation used in our submission for ILSVRC14 is called GoogLeNet, a 22 layers deep network, the quality of which is assessed in the context of classification and detection.

1 Introduction

In the last three years, mainly due to the advances of deep learning, more concretely convolutional networks [10], the quality of image recognition and object detection has been progressing at a dramatic pace. One encouraging news is that most of this progress is not just the result of more powerful hardware, larger datasets and bigger models, but mainly a consequence of new ideas, algorithms and improved network architectures. No new data sources were used, for example, by the top entries in the ILSVRC 2014 competition besides the classification dataset of the same competition for detection purposes. Our GoogLeNet submission to ILSVRC 2014 actually uses 12× fewer parameters than the winning architecture of Krizhevsky et al [9] from two years ago, while being significantly more accurate. The biggest gains in object-detection have not come from the utilization of deep networks alone or bigger models, but from the synergy of deep architectures and classical computer vision, like the R-CNN algorithm by Girshick et al [6].

Another notable factor is that with the ongoing traction of mobile and embedded computing, the efficiency of our algorithms – especially their power and memory use – gains importance. It is noteworthy that the considerations leading to the design of the deep architecture presented in this paper included this factor rather than having a sheer fixation on accuracy numbers. For most of the experiments, the models were designed to keep a computational budget of 1.5 billion multiply-adds at inference time, so that the they do not end up to be a purely academic curiosity, but could be put to real world use, even on large datasets, at a reasonable cost.

SOTA performances in 2014 on ImageNet

Squeeze-and-Excitation Networks

Jie Hu[0000−0002−5150−1003] Li Shen[0000−0002−2283−4976] Samuel Albanie[0000−0001−9736−5134]
Gang Sun[0000−0001−6913−6799] Enhua Wu[0000−0002−2174−1428]

Abstract—The central building block of convolutional neural networks (CNNs) is the convolution operator, which enables networks to construct informative features by fusing both spatial and channel-wise information within local receptive fields at each layer. A broad range of prior research has investigated the spatial component of this relationship, seeking to strengthen the representational power of a CNN by enhancing the quality of spatial encodings throughout its feature hierarchy. In this work, we focus instead on the channel relationship and propose a novel architectural unit, which we term the "Squeeze-and-Excitation" (SE) block, that adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels. We show that these blocks can be stacked together to form SENet architectures that generalise extremely effectively across different datasets. We further demonstrate that SE blocks bring significant improvements in performance for existing state-of-the-art CNNs at slight additional computational cost. Squeeze-and-Excitation Networks formed the foundation of our ILSVRC 2017 classification submission which won first place and reduced the top-5 error to 2.251%, surpassing the winning entry of 2016 by a relative improvement of ~25%. Models and code are available at https://github.com/hujie-frank/SENet.

Index Terms—Squeeze-and-Excitation, Image representations, Attention, Convolutional Neural Networks.

1 Introduction

Convolutional neural networks (CNNs) have proven to be useful models for tackling a wide range of visual tasks [1], [2], [3], [4]. At each convolutional layer in the network, a collection of filters expresses neighbourhood spatial connectivity patterns along input channels—fusing spatial and channel-wise information together within local receptive fields. By interleaving a series of convolutional layers with non-linear activation functions and downsampling operators, CNNs are able to produce image representations that capture hierarchical patterns and attain global theoretical receptive fields. A central theme of computer vision research is the search for more powerful representations that capture only those properties of an image that are most salient for a given task, enabling improved performance. As a widely-used family of models for vision tasks, the development of new neural network architecture designs now represents a key frontier in this search. Recent research has shown that the representations produced by CNNs can be strengthened by integrating learning mechanisms into the network that help capture spatial correlations between features. One such approach, popularised by the Inception family of architectures [5], [6], incorporates multi-scale processes into network modules to achieve improved performance. Further work has sought to better model spatial dependencies [7], [8] and incorporate spatial attention into the structure of the network [9].

In this paper, we investigate a different aspect of network design - the relationship between channels. We introduce a new architectural unit, which we term the *Squeeze-and-Excitation* (SE) block, with the goal of improving the quality of representations produced by a network by explicitly modelling the interdependencies between the channels of its convolutional features. To this end, we propose a mechanism that allows the network to perform feature recalibration, through which it can learn to use global information to selectively emphasise informative features and suppress less useful ones.

The structure of the SE building block is depicted in Fig. 1. For any given transformation $\mathbf{F}_{tr}$ mapping the input $\mathbf{X}$ to the feature maps $\mathbf{U}$ where $\mathbf{U} \in \mathbb{R}^{H \times W \times C}$, e.g. a convolution, we can construct a corresponding SE block to perform feature recalibration. The features $\mathbf{U}$ are first passed through a *squeeze* operation, which produces a channel descriptor by aggregating feature maps across their spatial dimensions ($H \times W$). The function of this descriptor is to produce an embedding of the global distribution of channel-wise feature responses, allowing information from the global receptive field of the network to be used by all its layers. The aggregation is followed by an *excitation* operation, which takes the form of a simple self-gating mechanism that takes the embedding as input and produces a collection of per-channel modulation weights. These weights are applied to the feature maps $\mathbf{U}$ to generate the output of the SE block which can be fed directly into subsequent layers of the network.

It is possible to construct an SE network (SENet) by simply stacking a collection of SE blocks. Moreover, these SE blocks can also be used as a drop-in replacement for the original block at a range of depths in the network architec-

- *Jie Hu and Enhua Wu are with the State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, 100190, China.*
  They are also with the University of Chinese Academy of Sciences, Beijing, 100049, China.
  Jie Hu is also with Momenta and Enhua Wu is also with the Faculty of Science and Technology & AI Center at University of Macau.
  E-mail: hujie@ios.ac.cn ehwu@umac.mo
- *Gang Sun is with LIAMA-NLPR at the Institute of Automation, Chinese Academy of Sciences. He is also with Momenta.*
  E-mail: sungang@nicemeta.ai
- *Li Shen and Samuel Albanie are with the Visual Geometry Group at the University of Oxford.*
  E-mail: {lishen,albanie}@robots.ox.ac.uk

SOTA performances in 2017 on ImageNet

ImageNet: Image classification of more than one million images into 1000 classes

Related Work: Atrial Fibrilliation vs Atrial Flutter

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.xxxx/xxxx.2024.DOI

Can Deep Learning Models Differentiate Atrial Fibrillation from Atrial Flutter?

ESTELA RIBEIRO[1,2], QUENAZ BEZERRA SOARES[1,3], FELIPE MENEGUITTI DIAS[1,3], JOSE EDUARDO KRIEGER[1], AND MARCO ANTONIO GUTIERREZ.[1,2,3]

[1]Heart Institute, University of Sao Paulo (INCOR), Sao Paulo, SP, Brazil
[2]University of Sao Paulo Medical School (FMUSP), Sao Paulo, SP, Brazil
[3]Polytechnique School, University of Sao Paulo (POLI - USP), Sao Paulo, SP, Brazil

Corresponding author: Estela Ribeiro (e-mail: estela.ribeiro@hc.fm.usp.br).

This work was supported in part by São Paulo Research Foundation (FAPESP) – grant n° 2021/12935-0, the Foscons Brazil, and the Zerbini Foundation as part of the research project "Machine Learning in Cardiovascular Medicine"

ABSTRACT Atrial Fibrillation (AFib) and Atrial Flutter (AFlut) are prevalent irregular heart rhythms that poses significant risks, particularly for the elderly. While automated detection systems show promise, misdiagnoses are common due to symptom similarities. This study investigates the differentiation of AFib from AFlut using standard 12-lead ECGs from the PhysioNet CinC Challenge 2021 (CinC2021) databases, along with data from a private database. We employed both one dimensional-based (1D) and image-based (2D) Deep Learning models, comparing different 1D and 2D Convolutional Neural Network (CNN) architectures for classification. For 1D models, LiteVGG-11 demonstrated the highest performed, achieving an accuracy (Acc) of 77.91 (±1.73%), area under the receiver operating characteristic curve (AUROC) of 87.17 (±1.29%), F1 score of 76.59 (±1.90%), specificity (Spe) of 71.69 (±4.73%), and sensitivity (Se) of 86.53 (±5.33%). On the other hand, for 2D models the EfficientNet-B2 outperformed other architectures, with an Acc of 75.20 (±3.38%), AUROC of 85.50 (±1.14%), F1 of 71.59 (±3.66%), Spe of 74.76 (±13.85%) and Se of 75.74 (±13.85%). Our findings indicate that distinguishing between AFib and AFlut is non-trivial, with 1D signals exhibiting superior performance compared to their 2D counterparts. Furthermore, it's noteworthy that the performance of our models on the CinC2021 databases was considerably lower than on our private dataset.

INDEX TERMS ECG, Atrial Fibrillation, Atrial Flutter, Deep Learning, Convolutional Neural Network

I. INTRODUCTION

Atrial Fibrillation (AFib) and Atrial Flutter (AFlut) are distinct irregular heart rhythms originating from abnormal activity in the heart's upper chambers, the Atria [1], [2]. These conditions pose significant risks, especially for the elderly [3]. AFib involves chaotic electrical activity, causing rapid, irregular atrial contractions at 350-500 beats per minute, compromising heart function and raising stroke risk [4], [5]. AFlut, often misdiagnosed as AFib, features a single electrical circuit driving atrial contractions at 250-350 beats per minute, disrupting heart function [4]. Early diagnosis and treatment are crucial for managing AFib and AFlut and reducing severe complications such as stroke [3].

Subtle or absent symptoms often accompany irregular heart rhythms, including chest pain, dizziness, shortness of breath, fainting, and palpitations [3], [5], associated with rapid ventricular rate and inadequate diastolic ventricular filling [2]. Automated detection systems can significantly aid in promptly and accurately identifying these conditions, improving healthcare efficiency and reducing patient wait times. This is especially beneficial for underprivileged hospitals with limited access to experienced cardiologists, alleviating strain on their healthcare infrastructure.

The electrocardiogram (ECG), an essential tool for diagnosing cardiac issues, is utilized extensively worldwide, with millions of exams conducted annually. ECG involves the measurement of the heart's electrical activity using electrodes affixed to patient's skin and is considered the gold standard for noninvasive diagnosis of various heart disorders [4]. Clinical assessment of AFib and AFlut predominantly relies on non-invasive 12-lead ECGs where distinct patterns of electrical activity on the ECG signal enable differentiation between these two conditions [1], [5].

On the ECG, AFib is characterized by the absence of P

Automated detection of atrial fibrillation and atrial flutter in ECG signals based on convolutional and improved Elman neural network☆

Jibin Wang

School of Mathematics, Tianjin University, Tianjin 300354, China
Visual Pattern Analysis Research Lab, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 21 October 2019
Received in revised form 13 December 2019
Accepted 26 December 2019
Available online 27 December 2019

MSC:
68Uxx
62P10

Keywords:
Atrial fibrillation
Atrial flutter
Convolutional neural network
Improved Elman neural network

ABSTRACT

Atrial fibrillation (AF) and atrial flutter (AFL) are two common life-threatening arrhythmias. Both are not only easily transformed into each other, but also often cause misdiagnosis due to the similar clinical symptoms. The early efficient and accurate detection of AF and AFL is helpful to reduce the pain and injury of patients suffering from these diseases, and the traditional detection is often inefficient and laborious. Therefore, we propose a new model mechanism using an 11-layers network architecture to automatically classify AF and AFL signals. It is mainly constructed with the convolutional neural network (CNN) and the improved Elman neural network (IENN). Besides, we specifically design two relative models as control subjects to validate the classification performance of the proposed model. 10-fold cross-validation is also implemented on the MIT-BIH AF database (AFDB) and the MIT-BIH arrhythmia database (MITDB), respectively. The obtained results show that the model achieved the accuracy, specificity, and sensitivity of 98.8%, 98.6%, and 98.9% on the AFDB database and 99.4%, 99.1%, and 99.6% on the MITDB database, respectively. The model mechanism has been demonstrated to have more superior performance than two relative models and some advanced algorithms, which can be considered as a reliable and efficient identification system to aid physicians and save lives.

1. Introduction

According to the World Health Organization (WHO), arrhythmia is a common life-threatening heart disease, affecting more than 12 million people in the United States and over 350 million people worldwide [1,2]. Among the diverse arrhythmia conditions, atrial fibrillation (AF) and atrial flutter (AFL) have higher morbidity and mortality. AF is an irregular abnormal rhythm along with the absence of P-wave and AFL is a relatively regular atrial rhythm due to the macroreentrant circuit phenomenon [3]. Notably, both are not only easily transformed into each other, but also often easily confused by cardiologists due to their similar clinical symptoms [2]. This undoubtedly increases the difficulty of diagnosis and even causes irreparable medical accidents if misdiagnosis occurs. Therefore, it is indispensable to combat this situation for the sake of public health.

However, the traditional AF and AFL detection are mainly dependent on visual observation of electrocardiogram (ECG) data by professional doctors, their personal experience often determines the final diagnosis. Thus, this visual analysis is relatively laborious and subjective. Also, a great quantity of ECG data is putting a heavy burden on physicians [2,4]. These limitations have motivated efforts to develop a computer-automated diagnosis system to discriminate AF and AFL signals from mass ECG data more efficiently and accurately.

The ECG is a primary non-invasive diagnostic tool that has been widely used in clinical practice due to its low-cost and simplicity. It is installed on the human body by electrodes for a period, which provides a large amount of key information about the electrical activity of the heart [1,5]. At the same time, the time-domain, frequency-domain and other features of ECG signals can be well revealed [6,7]. In recent years, many algorithms based on the ECG features have seen great success in boosting the performance of ECG signals analysis.

Currently, the majority of them mainly depend on machine learning algorithms [8–10]. The algorithm structure mainly includes feature extraction, feature selection, and classification. For instance, Henzel et al. [11] combined four statistical characteristics extracted from RR intervals (adjacent heartbeat intervals) with the generalized linear classifier for the discrimination of AF signals. Kumar et al. [12] performed the decision mechanism using the ECG features from log-energy entropy (LEE) and the random forest (RF) classifier to distinguish AF segments. Safarbali

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.knosys.2019.105446.

E-mail address: wangjibin0429@163.com.

https://doi.org/10.1016/j.knosys.2019.105446
0950-7051/© 2019 Elsevier B.V. All rights reserved.

- Trained several CNN models on Physionet 2021 challenge data

- State that performance on Physionet data (50-60% accuracy) is considerable lower than same models trained on own dataset (70-95% accuracy) (see comparison)

- Trained RNN on MIT-BIH database (99% accuracy)

- List more models with this high performances for AF and AFL classification on the same database

Related Work: ECGBERT fine-tuned on multiple tasks (AFIB classification, beat classification, ECG verification, sleep apnea detection)

ECGBERT: Understanding Hidden Language of ECGs with Self-Supervised Representation Learning

Seokmin Choi[1,2,†] Sajad Mousavi[1,†*] Phillip Si[1,3,†]
Haben G. Yhdego[1] Fatemeh Khadem[1] Fatemeh Afghah[4]
[1]*CardioPhi LLC*, CA, USA
[2]University at Buffalo, SUNY, NY, USA
[3]Carnegie Mellon University, PA, USA
[4]Clemson University, SC, USA
{seokmin.choi,sajad.mousavi,phillip.si}@cardiophi.com
{haben.yhdego,fatemeh.khadem}@cardiophi.com
fatemeh.afghah@clemson.edu

Abstract

In the medical field, current ECG signal analysis approaches rely on supervised deep neural networks trained for specific tasks that require substantial amounts of labeled data. However, our paper introduces ECGBERT, a self-supervised representation learning approach that unlocks the underlying language of ECGs. By unsupervised pre-training of the model, we mitigate challenges posed by the lack of well-labeled and curated medical data. ECGBERT, inspired by advances in the area of natural language processing and large language models, can be fine-tuned with minimal additional layers for various ECG-based problems. Through four tasks, including Atrial Fibrillation arrhythmia detection, heartbeat classification, sleep apnea detection, and user authentication, we demonstrate ECGBERT's potential to achieve state-of-the-art results on a wide variety of tasks.

1 Introduction

The Centers for Disease Control (CDC) reported that heart disease is the leading cause of death in the United States [for disease control, 2022]. Specifically, one person dies every 34 seconds from cardiovascular disease and about 697,000 people died from heart disease in 2020, which is one in every five deaths. The electrocardiogram (ECG) is the most essential bio-signal used by cardiologists and physicians to keep track of heart activity and detect different heart-related diseases and is used by cardiologists and physicians. One of the most critical limitations of ECG signals is that it requires manual analysis and annotation. Furthermore, the interpretation of the ECG signals varies from physician to physician as different heart diseases are associated with complex patterns within the ECG which can be hard to detect. The resulting inconsistencies may affect diagnostic accuracy or the trust between the patient and the physician. Therefore, to mitigate the aforementioned limitations in regard to manual ECG interpretation, several studies have proposed alternative ECG analysis techniques to achieve higher accuracy in real-time. Among these, deep learning-based approaches have recently gained traction in this domain [Pyakillya et al., 2017, Mousavi et al., 2020]. Compared with machine learning-based approaches where features need to be extracted manually, deep learning-based approaches automatically extract relevant features [Rajpurkar et al., 2017, Isin and Ozdalili, 2017], allowing for improved performance given enough data and a sufficiently expressive model. As a result, deep learning techniques have been widely applied to the medical domain in recent years to

*Corresponding author. †Equal contribution.

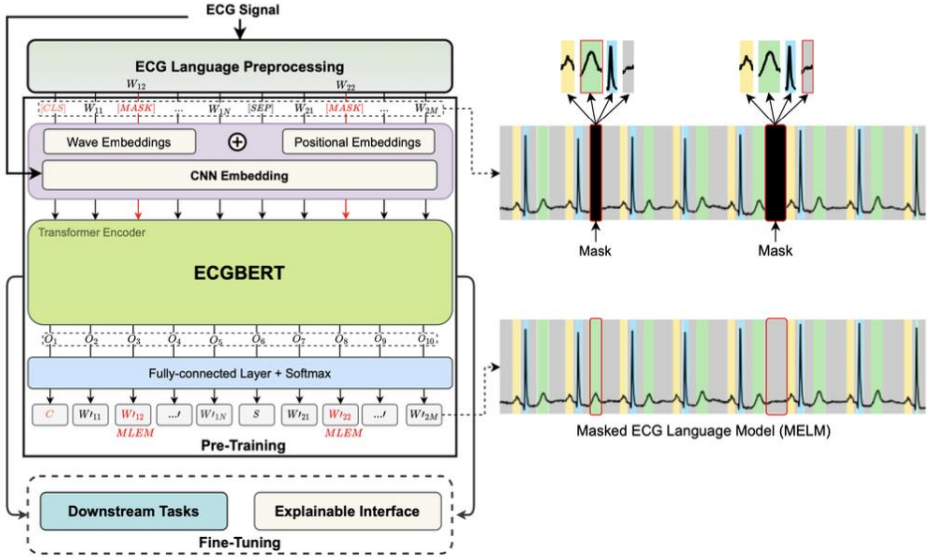
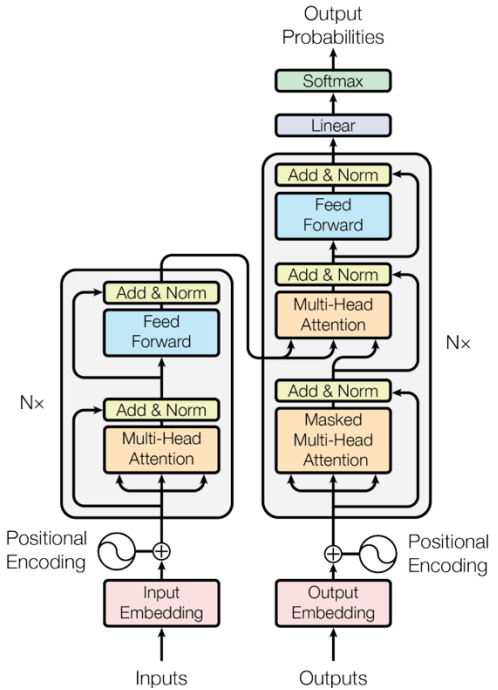


Figure 2: Illustration of the proposed ECGBERT architecture

Methologies: Background about Transformers

BERT

Encoder

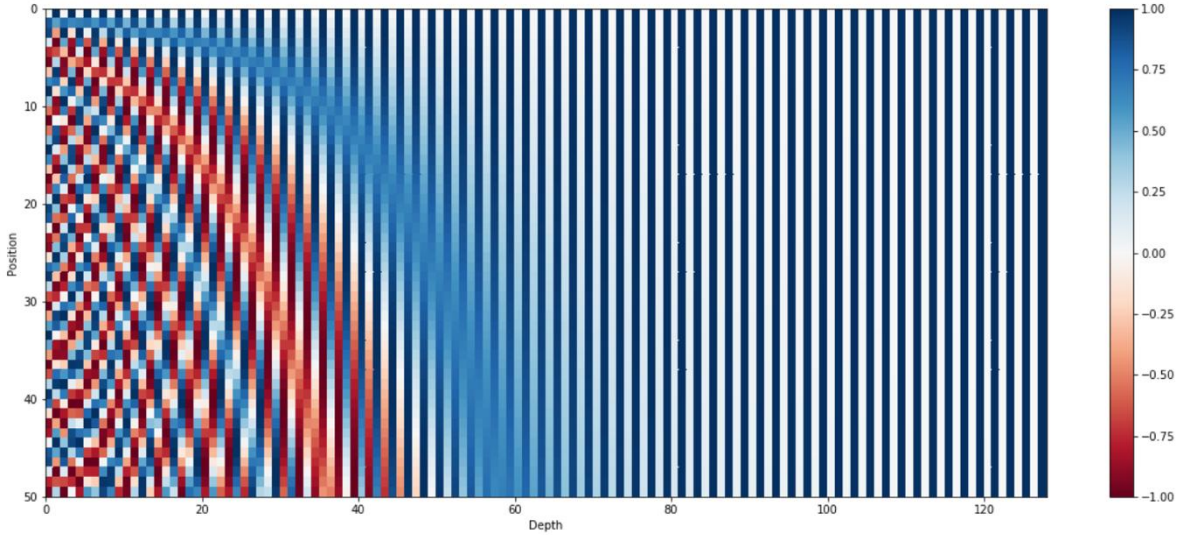


GPT

Decoder

Source: Attention Is All You Need, 2017

Methologies: Background about Transformers

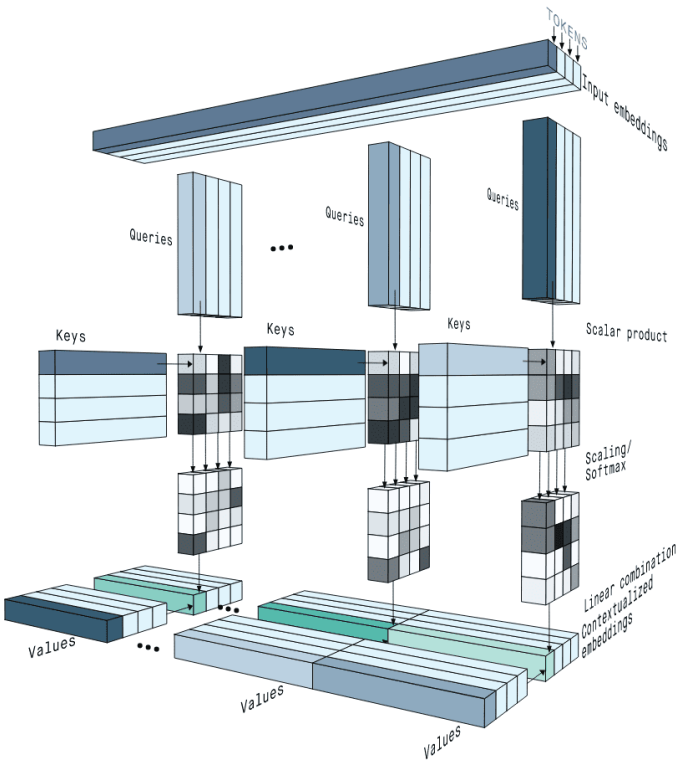


$$\mathrm{PE}(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\mathrm{model}}}}}\right)$$

$$\mathrm{PE}(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\mathrm{model}}}}}\right)$$

Source: https://kazemnejad.com/blog/transformer_architecture_positional_encoding

Multi-head Attention



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Source: https://theaisummer.com/self-attention

Methologies: Random Forest classifier (base-line model)

Biobss features names	
Features (averaged)	**Description**
a_R	Amplitude of R peak
RR0	Previous RR interval
RR1	Current RR interval
RR2	Subsequent RR interval
RRm	Mean of RR0, RR1, and RR2
RR_0_1	Ratio of RR0 to RR1
RR_2_1	Ratio of RR2 to RR1
RR_m_1	Ratio of RRm to RR1
t_PR	Time between P and R peak locations
t_QR	Time between Q and R peak locations
t_SR	Time between S and R peak locations
t_TR	Time between T and R peak locations
t_PQ	Time between P and Q peak locations
t_PS	Time between P and S peak locations
t_PT	Time between P and T peak locations
t_QS	Time between Q and S peak locations
t_QT	Time between Q and T peak locations
t_ST	Time between S and T peak locations
t_PT_QS	Ratio of t_PT to t_QS
t_QT_QS	Ratio of t_QT to t_QS
a_PQ	Difference of P wave and Q wave amplitudes
a_QR	Difference of Q wave and R wave amplitudes
a_RS	Difference of R wave and S wave amplitudes
a_ST	Difference of S wave and T wave amplitudes
a_PS	Difference of P wave and S wave amplitudes
a_PT	Difference of P wave and T wave amplitudes
a_QS	Difference of Q wave and S wave amplitudes
a_QT	Difference of Q wave and T wave amplitudes
a_ST_QS	Ratio of a_ST to a_QS
a_RS_QR	Ratio of a_RS to a_QR
a_PQ_QS	Ratio of a_PQ to a_QS
a_PQ_QT	Ratio of a_PQ to a_QT
a_PQ_PS	Ratio of a_PQ to a_PS
a_PQ_QR	Ratio of a_PQ to a_QR
a_PQ_RS	Ratio of a_PQ to a_RS
a_RS_QS	Ratio of a_RS to a_QS
a_RS_QT	Ratio of a_RS to a_QT
a_ST_PQ	Ratio of a_ST to a_PQ
a_ST_QT	Ratio of a_ST to a_QT

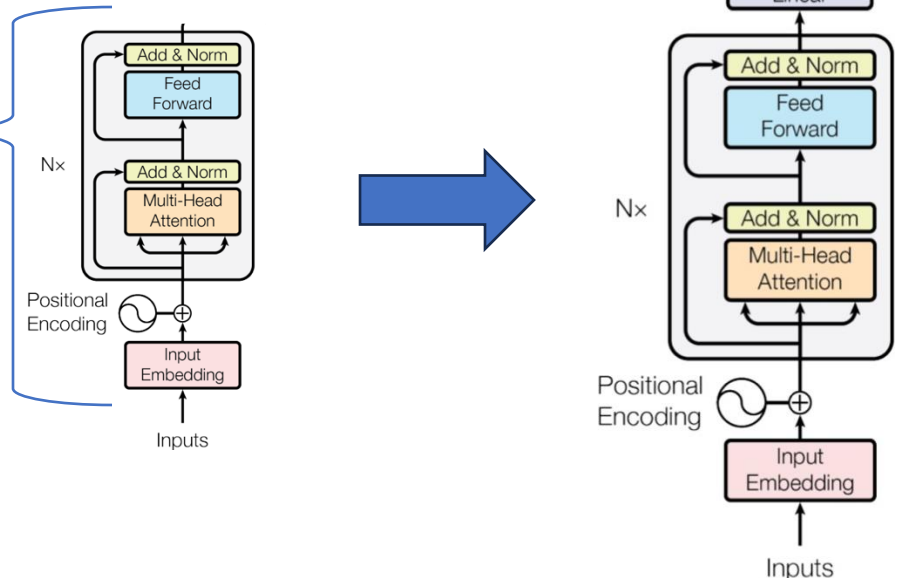
- 39 Biobss features:
  e.g. RR intervals (takes 4 consecutive points),
  P Q R S on/offset locations, ratios

- 100 decision trees
- Bootstrap samples
- Objective function: Gini Impurity

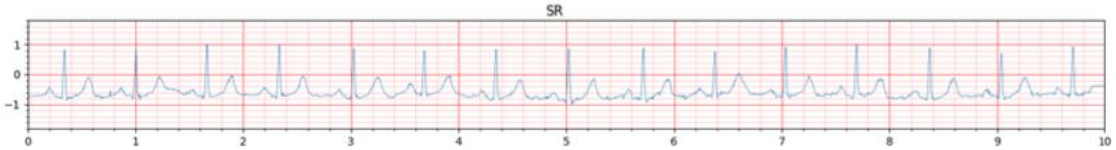
Methologies: Transformer encoder

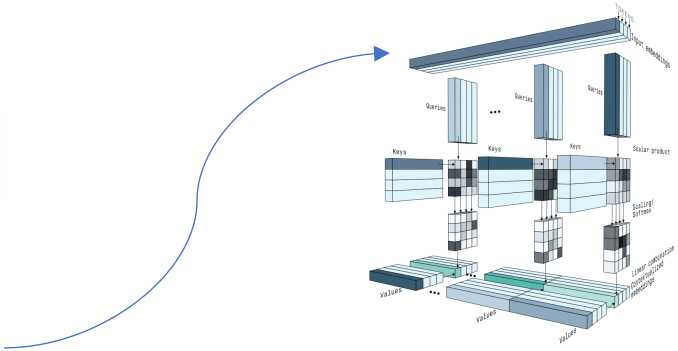
Focus:

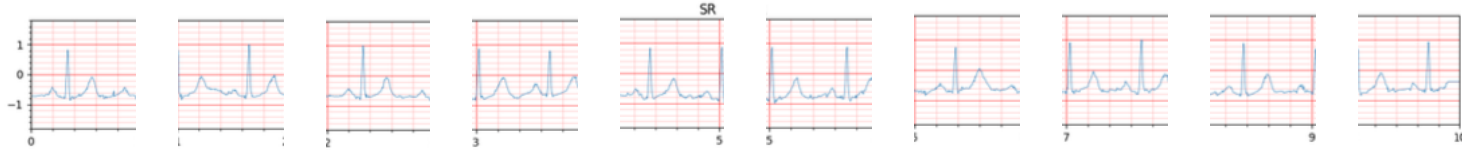
Encoder

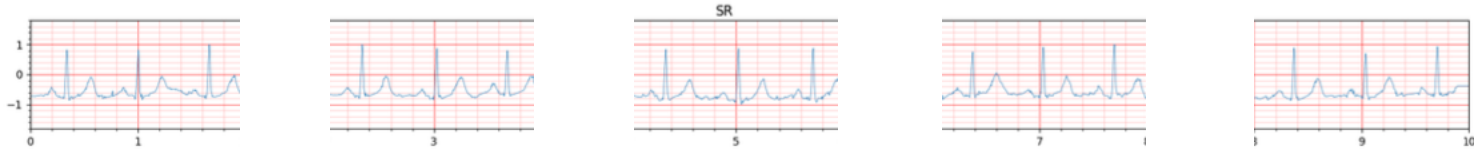


Methologies: Segmenting ECGs as input embeddings for Transformer encoder









Methologies: Residual Convolutional Network

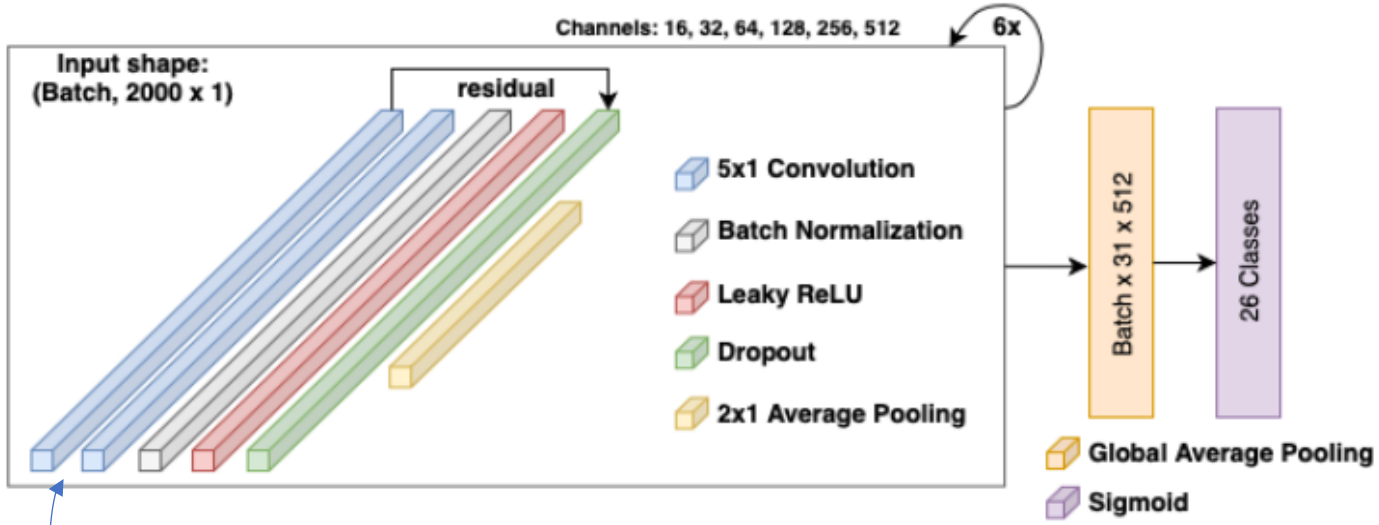


Figure 3.5: Residual Convolutional Network

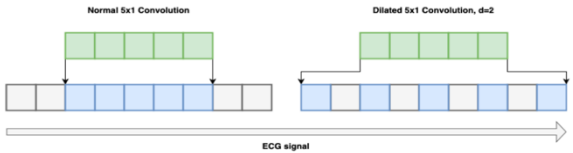


Figure 3.4: Dilated convolution

Methologies: Residual multi-scale Convolutional Network

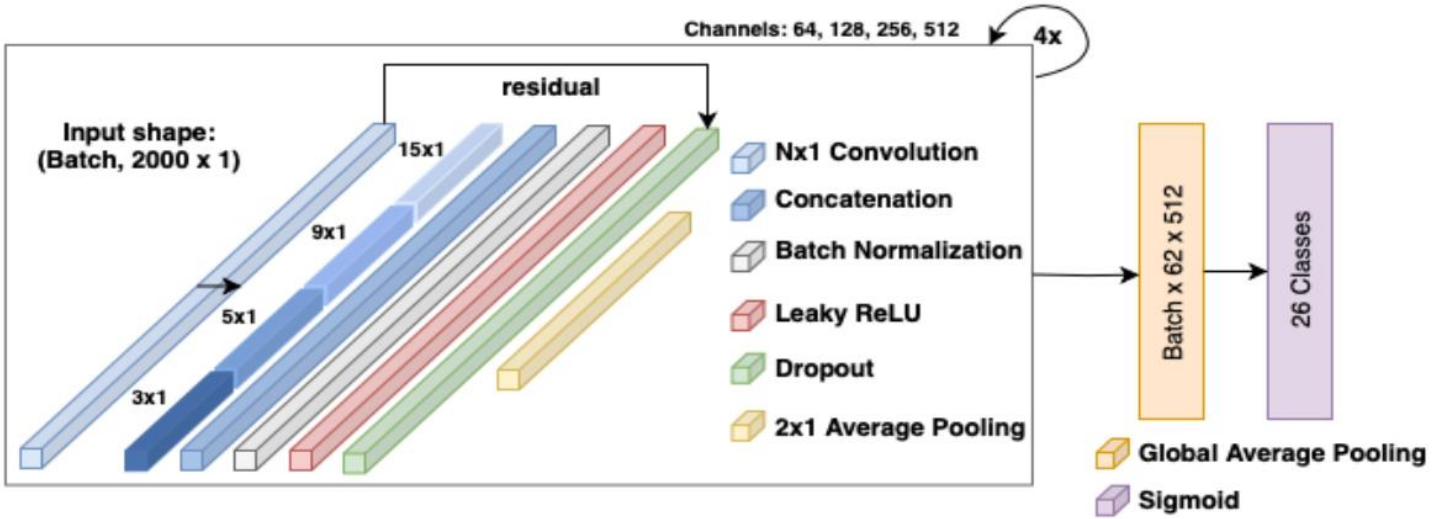
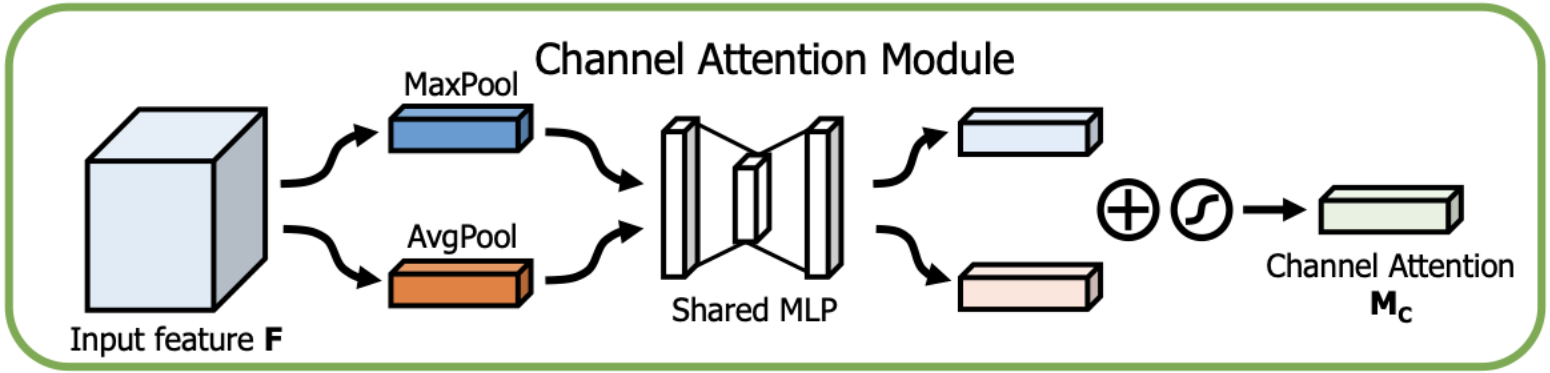


Figure 3.6: Multi-scale residual Convolutional Network

Methologies: Squeeze and Excitation Network



Source: CBAM - Convolutional Block Attention Module

- Modification of Squeeze and Excitation Network
- Learns inter-channel relations
- 3 channel-attention blocks (64, 128, 256 channels)
- In this version, I do not employ spatial-attention

Proposed approach: Residual multi-scale Convolutional Network and Transformer encoder

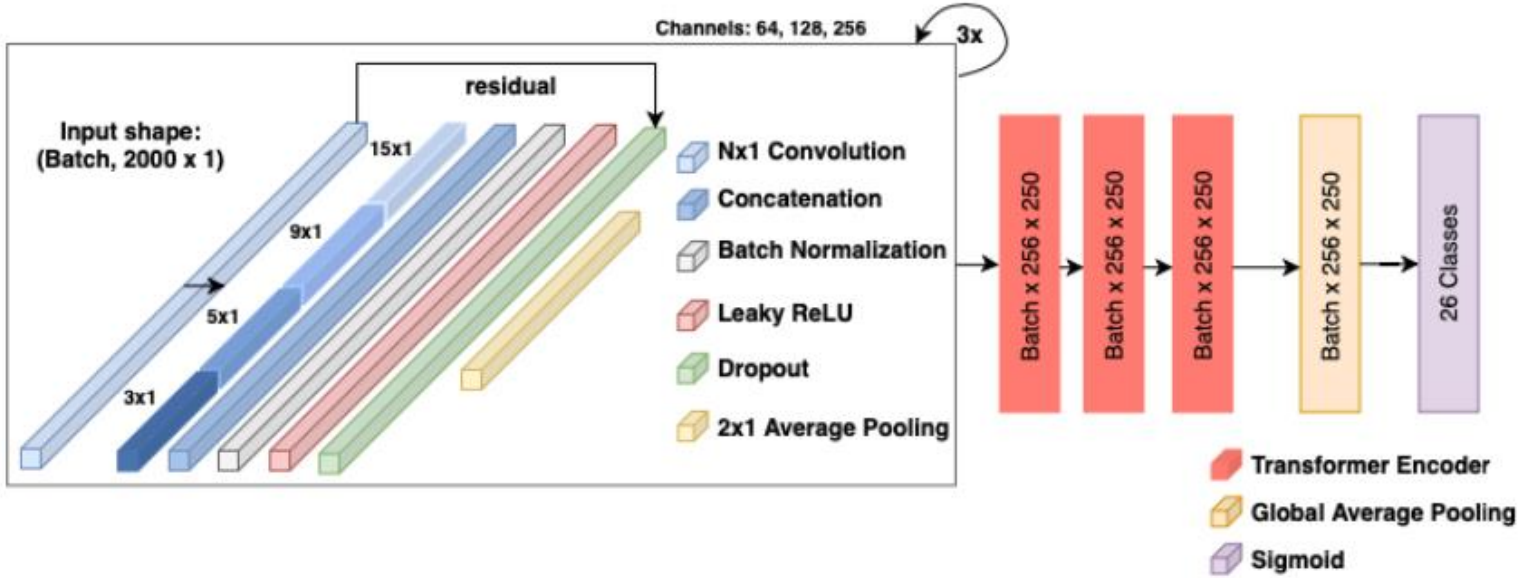
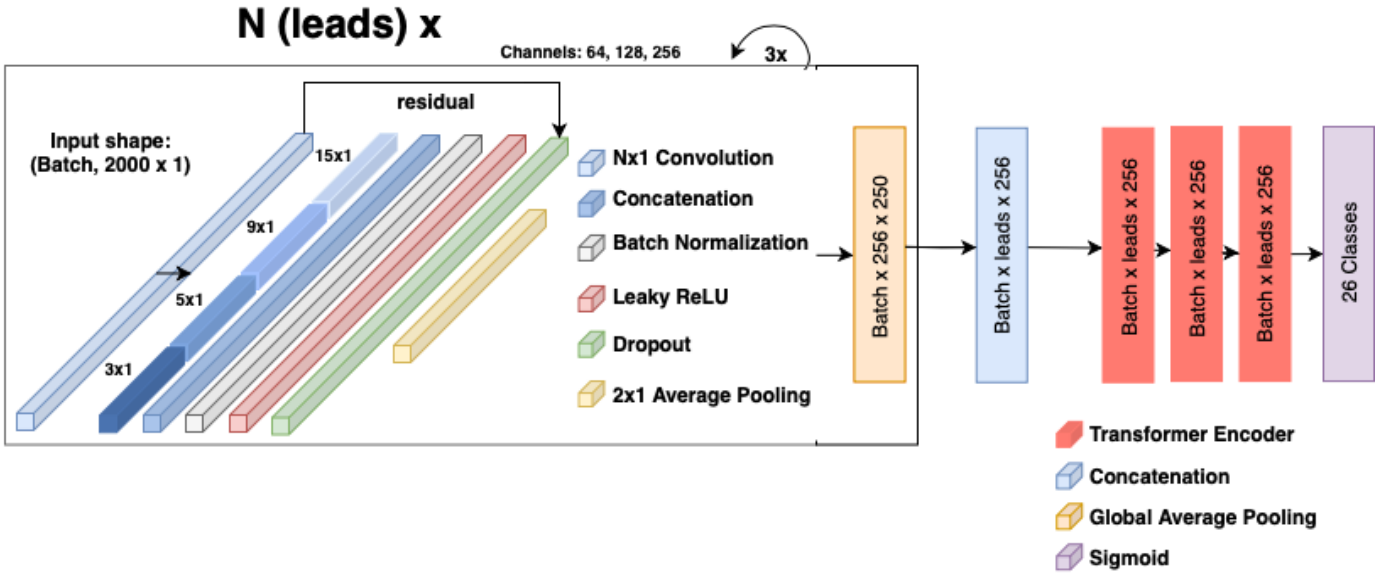
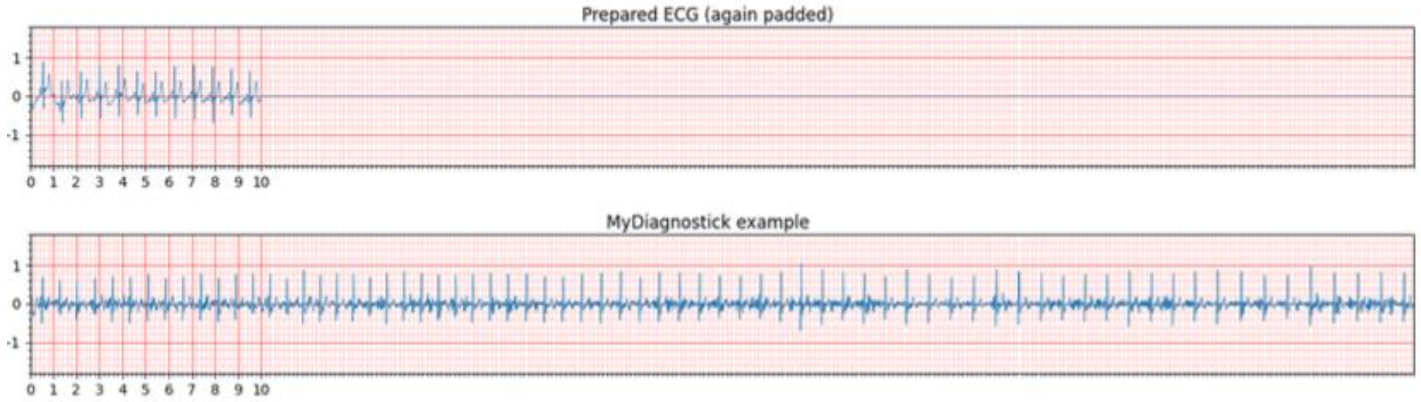


Figure 3.7: Multi-scale residual Convolutional Network combined with three Transformer encoder blocks

Methologies: Modification for multi-lead ECGs



Methologies: Model modifications for transfer to MyDiagnostick database



Methologies: Fine-tuning to Physionet 2021 challenge metrics

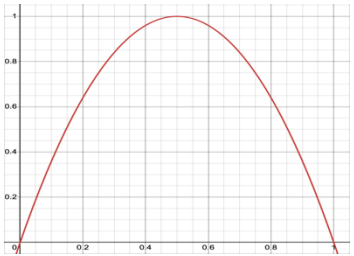
$$\text{Loss} = \sum^{batch} \left(\text{BCE}(T, P) - \text{CL}(T, P) + \text{SL}(P) \right) \tag{3.5}$$

$$\text{CL}\,(T, P) = \sum_{i=1}^{class\ n} \sum_{j=1}^{class\ n} \omega_{ij}\, a_{ij} \tag{3.6}$$

$$\text{SL}\,(P) = \text{average}\,(\,-4P(P-1)\,) \tag{3.7}$$

$$A = T^T \left(\frac{P}{N} \right) \tag{3.8}$$

$$N = ((T + P - T \odot P)\,\mathbf{1}_{c\times 1})\;\mathbf{1}_{1\times c} \tag{3.9}$$



weights

	164889003	164890007	6374002	426627000	733534002	164909002	713427006	59118001	270492004	713426002	39732003	445118002	164947007	251146004	111975006	698252002	426783006	284470004	63593006	10370003	365413008	427172004	17338001	164917005	47865007	427393009	426177001	427984000	164934002	59931005
164889003	1.0		0.5		475	0.3			475	0.4		0.3		0.3	0.35	0.35	0.3		425	0.45	0.35		0.25	0.3375		375		425		375	0.4		0.35		0.3		0.3		375	0.5		0.5	
164890007	0.5		1.0		475	0.3			475	0.4		0.3		0.3	0.35	0.35	0.3		425	0.45	0.35		0.25	0.3375		375		425		375	0.4		0.35		0.3		0.3		375	0.5		0.5	
6374002		475		475	1.0		325			475			425		325		325		375		375		475		375		275	0.3625		0.4		0.45		0.4		375		375		325		325	0.4		475		475
426627000	0.3		0.3		325	1.0			325	0.4		0.5		0.5	0.45	0.45	0.5		375	0.35	0.45		0.45	0.4625		425		375		425	0.2		0.45		0.5		0.5		425	0.3		0.3	
733534002		475		475		475		325		325		325		375		375		325	0.45		475		375		275	0.3625		0.4		0.45		0.4		375		375		325		325	0.4		475		475
164909002		475		475		475	325	1.0		325		325		375		375		325	0.45		475		375		275	0.3625		0.4		0.45		0.4		375		375		325		325	0.4		475		475
713427006	0.4		0.4		425	0.4		425	1.0		0.4		0.4	0.45	0.45	0.4		475	0.45	0.45		0.35	0.4375		475		475		475	0.3		0.45		0.4		0.4		475	0.4		0.4	
59118001	0.3		0.3		325	0.5			325	0.4		1.0		0.5	0.45	0.45	0.5		375	0.35	0.45		0.45	0.4625		425		375		425	0.2		0.45		0.5		0.5		425	0.3		0.3	
270492004	0.3		0.3		325	0.5			325	0.4		0.5		1.0	0.45	0.45	0.5		375	0.35	0.45		0.45	0.4625		425		375		425	0.2		0.45		0.5		0.5		425	0.3		0.3	
713426002	0.3		0.3		325	0.5			325	0.4		0.5		1.0	0.45	0.45	0.5		375	0.35	0.45		0.45	0.4625		425		375		425	0.2		0.45		0.5		0.5		425	0.3		0.3	
39732003	0.35		0.35		375	0.45			375	0.45		0.45		0.45	1.0	0.5	0.45		425	0.4	0.5		0.4	0.4875		475		425		475	0.25		0.5		0.45		0.45		475	0.35		0.35	
445118002	0.35		0.35		375	0.45			375	0.45		0.45		0.45	0.5	1.0	0.45		425	0.4	0.5		0.4	0.4875		475		425		475	0.25		0.5		0.45		0.45		475	0.35		0.35	
164947007	0.3		0.3		325	0.5			325	0.4		0.5		0.5	0.45	0.45	1.0		375	0.35	0.45		0.45	0.4625		425		375		425	0.2		0.45		0.5		0.5		425	0.3		0.3	
251146004		425		425	0.45		375	0.45		475		375		375	425		425		375	1.0		475		425		325	0.4125		0.45		475	0.45		325		425		375		375	0.45		425		425
111975006	0.45		0.45		475	0.35			475	0.45		0.35		0.35	0.4	0.4	0.35		475	1.0	0.4		0.3	0.3875		425		475		425	0.35		0.4		0.35		0.35		425	0.45		0.45	
698252002	0.35		0.35		375	0.45			375	0.45		0.45		0.45	0.5	0.5	0.45		425	0.4	1.0		0.4	0.4875		475		425		475	0.25		0.5		0.45		0.45		475	0.35		0.35	
426783006	0.25		0.25		275	0.45			275	0.35		0.45		0.45	0.4	0.4	0.45		325	0.3	0.4		1.0	0.4125		375		325		375	0.15		0.4		0.45		0.45		375	0.25		0.25	
284470004	0.3375		0.3375		0.3625		0.4625		0.3625		0.4375		0.4625		0.4625		0.4875		0.4875		0.4625		0.4125		0.3875		0.4875		0.4125		1.0		0.4625		0.4125		0.4625			0.2375		0.4875		0.4625		0.4625		0.4625		0.3375		0.3375
63593006		375		375	0.4		425	0.4				475		425		425		475		475		425	0.45		425		475		375	0.4625		1.0	0.45	0.5			275	475		425		425	0.5		375		375
10370003		425		425	0.45		375	0.45				475		375		375		425		425		325	0.4125		0.45		1.0		0.45		0.5				325		425		375		375	0.45		425		425
365413008		375		375	0.4		425	0.4				475		425		425		475		475		425	0.4625		0.5		0.45		1.0				275		475		425		425	0.5		375		375
427172004		375		375	0.4		425	0.4				475		425		425		475		475		425	0.4625		0.5		0.45		1.0				275		475		425		425	0.5		375		375
17338001																																															
164917005	0.4		0.4		375	0.2			375	0.3		0.2		0.2	0.25	0.25	0.2		325	0.35	0.25		0.15	0.2375		275		325		275	1.0		0.25		0.2		0.2		275	0.4		0.4	
47865007	0.35		0.35		375	0.45			375	0.45		0.45		0.45	0.5	0.5	0.45		425	0.4	0.5		0.4	0.4875		475		425		475	0.25		1.0		0.45		0.45		475	0.35		0.35	
427393009	0.3		0.3		325	0.5			325	0.4		0.5		0.5	0.45	0.45	0.5		375	0.35	0.45		0.45	0.4625		425		375		425	0.2		0.45		1.0		0.5		425	0.3		0.3	
426177001	0.3		0.3		325	0.5			325	0.4		0.5		0.5	0.45	0.45	0.5		375	0.35	0.45		0.45	0.4625		425		375		425	0.2		0.45		0.5		1.0		425	0.3		0.3	
427984000		375		375	0.4		425	0.4		475		425		475		475		425	0.45		475		375	0.4625		0.5		0.45		0.5				425		475		425	1.0		375		375
164934002	0.5		0.5		475	0.3			475	0.4		0.3		0.3	0.35	0.35	0.3		425	0.45	0.35		0.25	0.3375		375		425		375	0.4		0.35		0.3		0.3		375	1.0		0.5	
59931005	0.5		0.5		475	0.3			475	0.4		0.3		0.3	0.35	0.35	0.3		425	0.45	0.35		0.25	0.3375		375		425		375	0.4		0.35		0.3		0.3		375	0.5		1.0	

Methologies: Pre-processing

- Normalising to -1 and 1
- Zero-padding / truncation to 10s (5000, 1)
- Resampling from 500 to 200 Hz (2000, 1)
- Butterworth bandpass filtering (0.3 to 21 Hz)

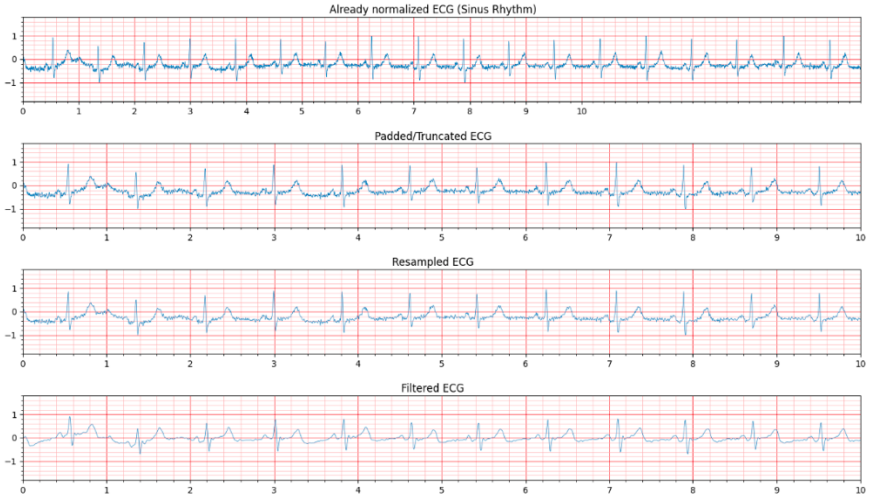
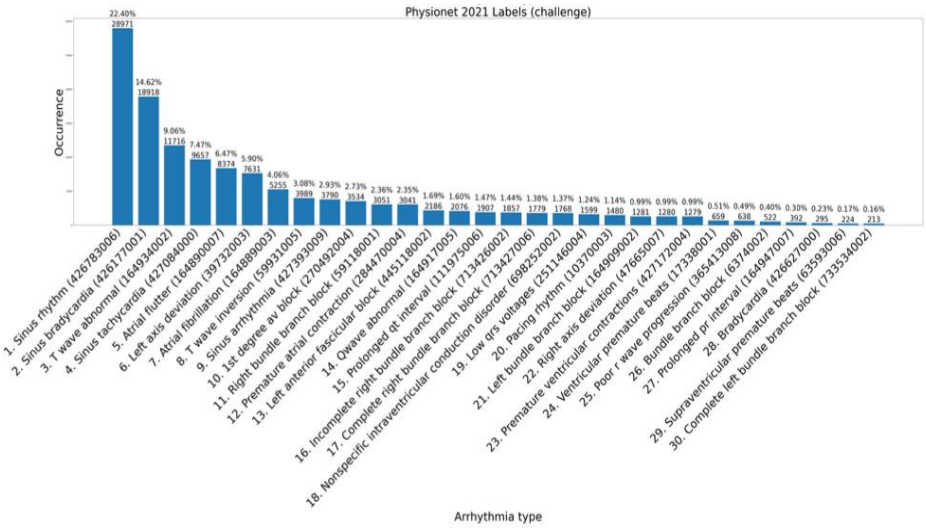


Figure 4.5: ECG pre-processing steps applied to the Physionet 2021 data

Experimental setup: Physionet 2021 challenge database

Dataset source	Recording length in seconds	Sampling frequency	ECG samples
Chapman-Shaoxing database & Ningbo database	10s	500 Hz	45,152
PTB database & PTB-XL database	10-120s	either 500 or 1000 Hz	22,353
Georgia 12-lead challenge database	5-10s	500 Hz	10,344
CPSC database & CPSC-extra database	6-144s	500 Hz	10,330
INCART 12-lead database	1800s (30 min)	257 Hz	74



Source: https://physionet.org/content/challenge-2021/1.0.3/

Experimental setup: MyDiagnostick database

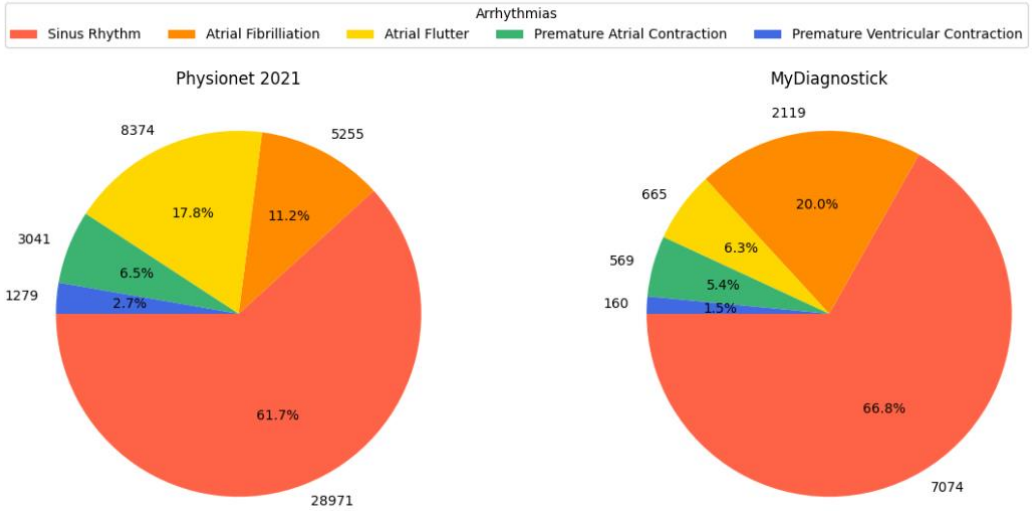


Figure 4.2: Physionet 2021 and MyDiagnostick class distribution - SR, AF, AFL, PAC and PVC

Pre-procssing: Multi-label class-balancing

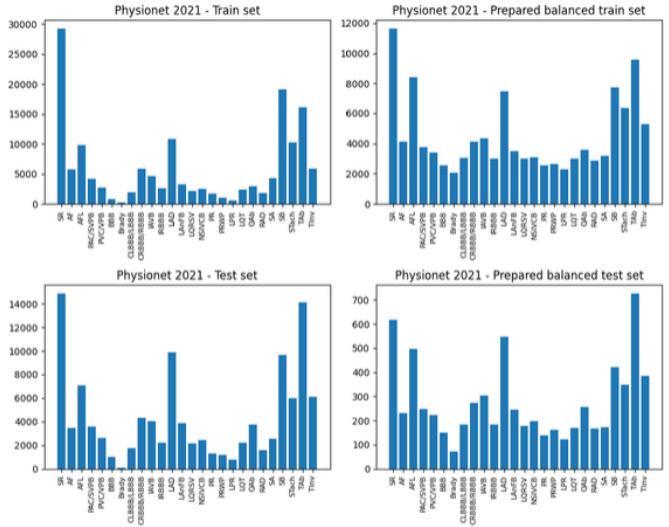


Figure 4.3: Balanced Physionet 2021 data (All scored classes)

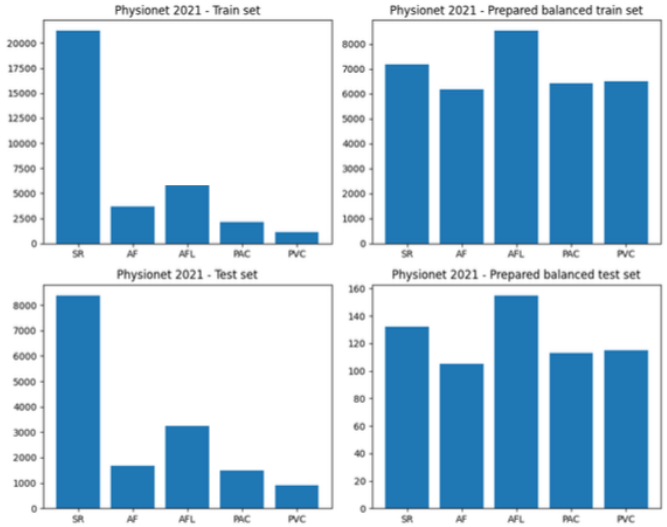


Figure 4.4: Balanced Physionet 2021 data (5 classes)

Experiments

Input shape	Pos. enc.	Enc. blocks	Heads	qkv dim	ff dim	Drop-out	Train. param.	Acc.	Prec.	Rec.	F1
(40, 50)	True	1	1	25	24	0.1	155.375	0.096	0.514	0.120	0.194
(40, 50)	True	1	1	25	24	0.4	155.375	0.065	0.418	0.083	0.139
(40, 50)	True	8	1	25	24	0.1	878.818	0.061	0.579	0.079	0.139
(40, 50)	True	8	1	25	24	0.4	878.818	0.098	0.592	0.122	0.203
(40, 50)	False	1	1	25	24	0.1	155.375	0.227	0.747	0.25	0.374
(40, 50)	False	1	1	25	24	0.4	155.375	0.224	0.742	0.253	0.378
(40, 50)	False	8	1	25	24	0.1	878.818	0.228	**0.765**	0.261	0.389
(40, 50)	False	8	1	25	24	0.4	878.818	0.226	0.737	0.255	0.379
(10, 200)	False	8	1	25	24	0.1	1.004.818	0.160	0.744	0.174	0.283
(10, 200)	False	8	8	25	24	0.1	2.129.018	0.177	0.709	0.197	0.308
(40, 50)	False	8	8	400	24	0.1	6.035.018	0.223	0.762	0.247	0.374
(40, 50)	False	8	8	25	2048	0.1	65.947.210	0.219	0.740	0.257	0.382
(40, 50)	False	8	8	400	2048	0.1	70.819.210	0.226	0.751	0.257	0.383
(40, 50)	False	8	8	400	2048	0.4	70.819.210	**0.245**	0.739	**0.286**	**0.413**

Table 4.2: Transformer encoder evaluated on the Physionet 2021 challenge data

Experimental results: All developed models on the Physionet 2021 challenge data (26 classes); the challenge metrics are on the right

Model	Parameters	Accuracy	Precision	Recall	F-measure
Random Forest (Biobss features)	5.556.212	**0.390**; /	**0.714**	0.382	0.406
Residual CNN without dilation	3.702.282	0.387; 0.310	0.593	0.692	**0.625**
Residual CNN with dilation	3.702.282	0.386; 0.311	0.590	**0.695**	0.624
Transformer encoder (8 heads, 8 blocks)	2.129.018	0.225; 0.100	0.380	0.619	0.432
Multi-scale Convolutional Network	53.202.522	0.358; 0.299	0.581	0.673	0.609
Modified Squeeze and Excitation Network	**152.210**	0.257; 0.184	0.463	0.622	0.501
Multi-scale CNN and Transformer encoder	11.105.058	0.371; 0.308	0.579	0.689	0.616

Table 4.3: Physionet 2021 train/test split evaluation

Model	AUROC	AUPRC	Accuracy	F-measure	Challenge metric
Random Forest (Biobss features)	0.560	0.142	**0.390**	0.153	0.147
Residual CNN without dilation	**0.874**	**0.439**	0.310	**0.482**	**0.556**
Residual CNN with dilation	0.842	0.404	0.311	0.442	0.554
Transformer encoder (8 heads, 8 blocks)	0.701	0.211	0.100	0.258	0.402
Multi-scale Convolutional Network	0.832	0.382	0.299	0.414	0.526
Modified Squeeze and Excitation Network	0.768	0.290	0.184	0.322	0.445
Multi-scale CNN and Transformer encoder	0.843	0.403	0.309	0.446	0.547

Table 4.4: Physionet 2021 train/test challenge metric scores evaluation

Experimental results: Modification of proposed approach and multi-scale Convolutional Network for multi-lead ECG classification on Physionet data (26 classes)

Model	Parameters	Accuracy	Precision	Recall	F-measure
Multi-scale Convolutional Network (2-leads)	**1.808.794**	**0.362**; 0.216	0.555	**0.716**	0.598
Multi-scale CNN and Transformer encoder (2-leads)	2.676.706	0.352; 0.197	0.537	0.691	0.582
Multi-scale Convolutional Network (6-leads)	5.426.330	0.350; 0.213	0.549	0.697	0.584
Multi-scale CNN and Transformer encoder (6-leads)	6.444.770	0.353; 0.250	0.556	0.651	0.584
Multi-scale Convolutional Network (12-leads)	10.852.634	0.351; 0.185	0.543	0.709	0.595
Multi-scale CNN and Transformer encoder (12-leads)	12.096.866	0.357; **0.269**	**0.562**	0.714	**0.604**

Table 4.5: Physionet 2021 train/test split evaluation for multiple leads

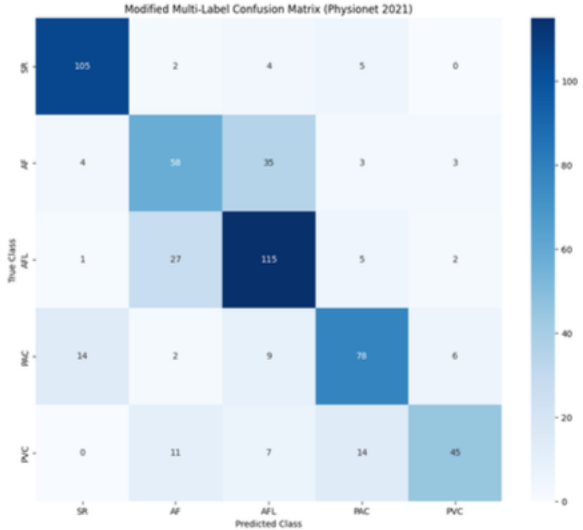
Model	AUROC	AUPRC	Accuracy	F-measure	Challenge metric
Multi-scale Convolutional Network (2-leads)	0.802	0.362	0.216	0.404	0.527
Multi-scale CNN and Transformer encoder (2-leads)	0.770	0.336	0.197	0.379	0.510
Multi-scale Convolutional Network (6-leads)	**0.805**	0.346	0.213	0.383	0.495
Multi-scale CNN and Transformer encoder (6-leads)	0.797	0.348	0.250	0.392	0.487
Multi-scale Convolutional Network (12-leads)	0.795	0.351	0.185	0.386	0.509
Multi-scale CNN and Transformer encoder (12-leads)	0.799	**0.367**	**0.269**	**0.409**	**0.552**

Table 4.6: Physionet 2021 challenge metric scores evaluation for multiple leads

Experimental results: 5-class models on Physionet data

Model	Accuracy	Precision	Recall	F-measure
Residual CNN without dilation (unbalanced test set)	**0.733**	**0.829**	**0.785**	**0.806**
Multi-scale CNN and Transformer encoder (unbalanced test set)	0.719	0.808	0.756	0.780
Residual CNN without dilation (balanced test set)	0.554	0.738	0.654	0.686
Multi-scale CNN and Transformer encoder (balanced test set)	**0.578**	**0.744**	**0.658**	**0.693**

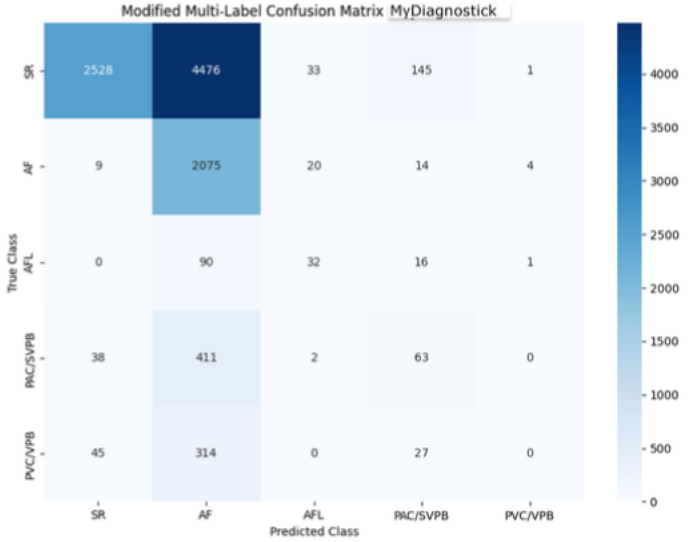
Table 4.7: Results of the residual CNN and the proposed approach for the Physionet 2021 challenge data (5 classes)



Experimental results: 5-class models on MyDiagnostick data

Model	Accuracy	Precision	Recall	F-measure
Residual CNN without dilation (5 classes)	**0.402**	0.744	**0.414**	**0.432**
Multi-scale CNN and Transformer encoder (5 classes)	0.090	0.575	0.136	0.189
Residual CNN without dilation (pre-traind on 26 classes)	0.111	**0.899**	0.210	0.240

Table 4.8: Results of the residual CNN and the proposed approach for the MyDiagnostick database (5 classes)



Discussion: Transformer

- Using a Transformer model without inputted features shows poor performances
- Parameter tuning does not help (e.g. embedding segments sizes, attention size, dropout …), however a deeper and larger network shows slight improvements, but should be evaluated critically duo to model sizes
- Adding temporal information unexpectedly randomizes performance
- Feeding convolutional features, as in the proposed approach, improves noticable the performances

Discussion: Transformer

- Using a Transformer model without inputted features shows poor performances
- Parameter tuning does not help (e.g. embedding segments sizes, attention size, dropout …), however a deeper and larger network shows slight improvements, but should be evaluated critically duo to model sizes
- Adding temporal information unexpectedly randomizes performance
- Feeding convolutional features, as in the proposed approach, improves noticable the performances

Discussion: Transformer

- Using a Transformer model without inputted features shows poor performances
- Parameter tuning does not show any improvements (e.g. embedding segments sizes, attention size, dropout …), however a deeper and larger network shows slight improvements, but should be evaluated critically duo to model sizes
- Adding temporal information unexpectedly randomizes performance
- Feeding convolutional features, as the proposed approach, improves noticable the performances

Conclusions

1. How well does a Transformer-based model perform on the Physionet 2021 challenge data compared to a feature-based model or a Convolutional Network?

Answer: Based on the experimental results is a conventional Transformer coder model trained on raw ECGs without any modifications, i.e. input features, not able to show good performances for ECG classification. Even by adjusting the segment sizes or adding positional information unexpectedly reduced the performances. The model shows slightly better performance with more parameters. The reason why the Multi-head Attention mech- anism is not able to extract meaningful features from raw ECGs could be related to the characteristics of the ECG signals or the input preparation, i.e. that the wave segments are not aligned. In comparison, the combination of a Convolutional Network and the Multi- Head Attention mechanism, as in the proposed approach, shows competing performances, while the question remains whether the Multi-Head Attention mechanism adds performance improvements. The highest accuracy is provided by the residual Convolutional Network and the feature-based Random Forest classifier, while within the Physionet 2021 challenge metrics the residual Convolutional Network performs best.

Conclusions

2. Can an ensemble Transformer model and Convolutional Network effectively capture spatio-temporal information from multi-lead ECGs and improve accuracy?

Answer: Based on the experiments, the modified proposed approach for multi-lead classifi- cation shows equivalent performance to the multi-scale Convolutional Network on different lead setups. While the proposed model architecture is designed such that the Multi-head Attention block can learn relations between leads using separate globally averaged feature map vectors from each lead as input, the model does not show remarkable performance improvements over the similarly designed multi-scale Convolutional Network without the attention mechanism on different lead experiments and therefore the research question can be answered as no.

Conclusions

3. How is the performance of the proposed approach at discriminating SR, AF, AFL, PAC and PVC?

Answer: Both models, the proposed approach and the residual Convolutional Network, show acceptable but improvable performance on the Physionet 2021 data. The models are both confident in classifying Sinus Rhythm. In terms of addressing confusion between Atrial Fibrillation and Atrial Flutter or improving data pre-processing steps, the models could potentially be improved in most cases.

Conclusions

4. What are the challenges in transferring the pre-trained models from the Physionet 2021 challenge data to the MyDiagnostick database? Do the models generalise well, even though different ECG devices were used?

Answer: Several pre-processing steps need to be considered, including type of lead, padding / truncating to equal sizes, resampling, filtering and normalisation. The biggest challenge comes from the different lengths of the recordings. The models cannot be transferred from short-term (10s) to long-term (60s) ECGs by simply zero-padding the training data. It is crucial to show examples of long-term ECGs in the pre-training phase to correctly train the models. As the latter criterion was not fully met, the models are not able to generalise to the MyDiagnostick data, but rather show biased performances.

Summary

- Development of various (deep-learning) models on the Physionet 2021 challenge database
- Proposed Transformer model do not show superior performances

Outlook: Improve on architecture design, features or exploit ensemble models

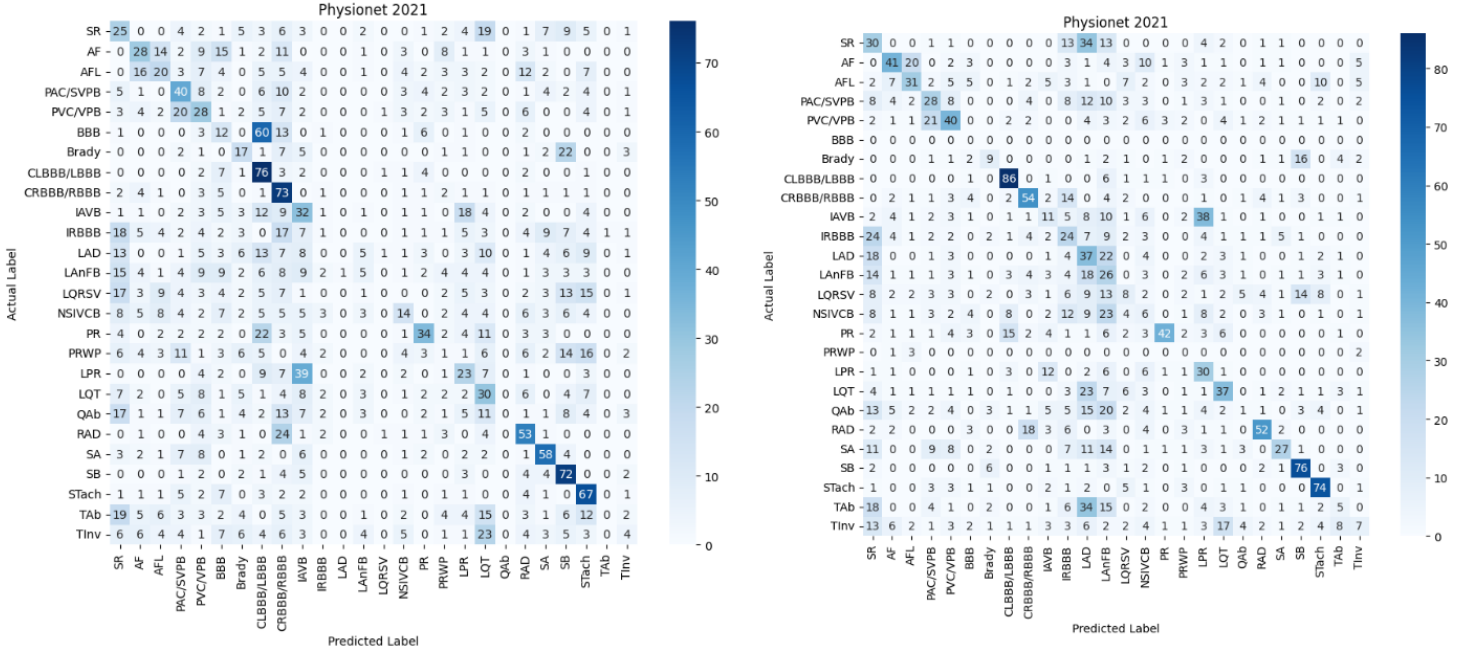


Figure 5.1: Confusion matrices from a residual Convolutional Network based on different training and tests on the Physionet 2021 data

Questions / Discussion