

A Spectral-longitudinal Model for Detection of Heart Attack from 12-lead Electrocardiogram Waveforms

Girmaw Abebe Tadesse^{1,2}, Hamza Javed², Komminist Weldemariam¹, Tingting Zhu²

¹IBM Research - Africa

²Department of Engineering and Science, University of Oxford, UK

Abstract—Cardiovascular diseases (CVDs) remain responsible for millions of deaths annually. Myocardial infarction (MI) is the most prevalent condition among CVDs. Although data-driven approaches have been applied to predict CVDs from ECG signals, comparatively little work has been done on the use of multiple-lead ECG traces and their efficient integration to diagnose CVDs. In this paper, we propose an end-to-end trainable and joint spectral-longitudinal model to predict heart attack using data-level fusion of multiple ECG leads. The spectral stage transforms the time-series waveforms to stacked spectrograms and encodes the frequency-time characteristics, whilst the longitudinal model helps to utilise the temporal dependency that exists in these waveforms using recurrent networks. We validate the proposed approach using a public MI dataset. Our results show that the proposed spectral-longitudinal model achieves the highest performance compared to the baseline methods.

Index Terms—Cardiovascular Disease, Deep Learning, Transfer Learning, Health Informatics, PTB ECG Database

I. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of deaths, and over 75% of these deaths occur in low- and middle- income countries. This is due to limited access to medical experts and precise diagnosis devices in these countries [1]. Among different types of CVDs, myocardial infarction (MI), commonly known as a heart attack, constitutes a significant portion of deaths caused by CVDs.

It has become common clinical practice to inspect ECG recordings for the diagnosis of CVDs such as MI. This procedure is often associated with a delay in timely decision making, particularly in resource-constrained settings, and hence degrades the quality of care delivered. To intelligently facilitate decision making in such circumstances, machine learning (ML) approaches can be applied throughout the end-to-end diagnosis process. ML could also provide additional insights to domain experts while reducing the subjectivity associated with traditional diagnosis practices.

Earlier works on automated CVD diagnosis require laborious handcrafting of domain-specific ECG features that are obtained through analysing signal morphologies such as QRS complexes, P-waves, R-peaks, T-waves and the PQ, QRS and ST segments [2], [3], [4], [5]. However, these methods lack robustness and do not generalise across variations in patient characteristics and device specifications. Generally, current limitations of existing works in CVDs diagnosis could be

characterised as follows. First, although MI is the leading cause of deaths among CVDs, it has received less attention compared to other CVDs. Secondly, most existing methods tend to utilise only a single or few leads of ECG waveforms, despite the fact that multiple ECG leads are understood to contain more heart-related information [6]. Moreover, those works that use multi-lead ECG still did not fully explore the fusion of these leads to maximise decision making [7]. Thirdly, most studies to date focused on classifying single ECG beats or short segments of longitudinal ECG waveforms without exploiting their temporal dependencies.

In this paper, we present a joint spectral-longitudinal model to predict MI from the conventional 12-lead ECG by exploiting the power of deep learning and transfer learning techniques. While deep learning is used to automate the feature encoding process and thus improve the generalisability of diagnosis [6], [8], [9], transfer learning is employed to ease the large data requirement for training of deep networks [10].

The key elements of our proposed approach are as follows (see Fig. 1): (1) a spectral representation is applied to transform 1-D time-series to 2-D frequency-time representations, with the goal of improving robustness across variations of device specifications and enables cross-domain transfer learning; (2) we perform early fusion of data from multiple leads by stacking the spectrograms. This helps to encode the spatial relationship among leads and mimics the clinical practice of arranging ECG leads; (3) we employ convolutional neural network (CNN) to further encode the spectral characteristics; (4) we employ a longitudinal model that consists of a long-short term memory (LSTM) network to exploit the temporal dependency among subsequent samples. We have validated our approach on the well known publicly available PTB Diagnostic ECG Database [11], [12].

II. METHOD

Our proposed approach predicts heart attack using data-level fusion of 12 ECG leads using two process blocks: *Data preparation and fusion* and *Modelling* as shown in Fig. 1. In this section, we first formulate the problem before discussing each block in greater detail.

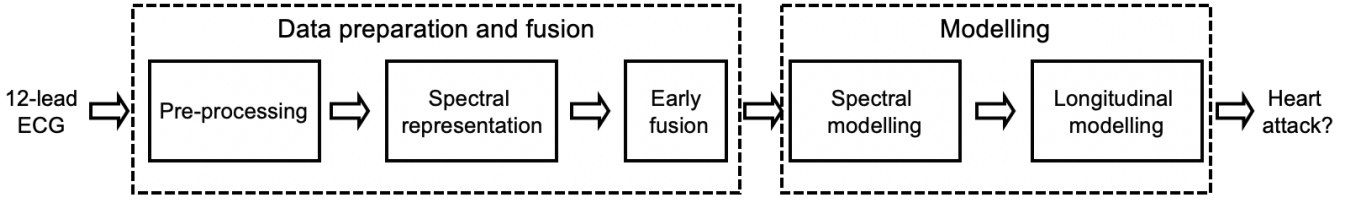


Fig. 1: Overview of the proposed approach.

A. Problem formulation

Given a dataset \mathcal{D} that contains the 12-lead ECG waveforms of N patients, i.e., $\mathcal{D} = \{R_i\}_{i=1}^N$, we aim to design a machine learning algorithm to identify patients with MI or not. A set of the 12 leads (U) is composed of six limb leads (I, II, III, aVR, aVL , and aVF) and six precordial leads (V_1, \dots, V_6). Thus, the ECG data of the i^{th} patient can be represented as $R_i = \{\mathbf{r}_i^l\}$ where \mathbf{r}_i^l represents the waveform of l^{th} ECG lead and $l \in U$. Our algorithms output c_i as the diagnosis output of R_i , i.e., $c_i \in \{Healthy, MI\}$. To do so, we extract multiple samples (windows) from the waveform of each patient's ECG lead $\mathbf{r}_i^l \in R_i$, i.e., $\mathbf{r}_i^l = (\mathbf{t}_{i1}^l, \mathbf{t}_{i2}^l, \dots, \mathbf{t}_{iW}^l, \dots, \mathbf{t}_{iW}^l)$, where W is the number of windows extracted from \mathbf{r}_i^l . Note that hereafter we drop the subscript i to improve readability. We then employ early fusion of information from multiple leads followed by spectral $\mathcal{S}(\cdot)$ and longitudinal $\mathcal{L}(\cdot)$ models that encode the spectral information on a sample window and the temporal dynamics across multiple windows, respectively.

B. Data preparation and fusion

ECGs are often susceptible to noise and movement artefacts, and so we employed a band-pass filter to mitigate these issues. We then employ a short-time Fourier transform (STFT), $\mathcal{T}(\cdot)$ to transform the ECG sample, \mathbf{t}_w^l , into a frequency-time representation (spectrogram), i.e. $\Gamma_w^l = \mathcal{T}(\mathbf{t}_w^l)$ (see Fig. 2). The spectral representation is an effective way to encode the temporal deviations expected on the ECG waveforms due to a heart attack. Moreover, spectral-based data representation achieves better robustness across variations in device specifications used for ECG collection. Furthermore, the 2-D representation of spectrograms enables the application of image-based CNNs to encode both frequency and time characteristics of the ECG signals. The 2-D representation also provides the flexibility to employ cross-domain transfer learning through the use of existing computer vision networks [13].

Different ECG leads can pick up the electrical activity of the heart from different angles. Thus, it is important to utilise the discriminative characteristics offered by these leads. For this purpose, we employ a data-level fusion scheme at the early stages of the pipeline as shown in Fig. 3. This avoids the need to model each lead separately. And, the stacking of spectrograms from the different leads in this early fusion scheme closely resembles the clinical visualisation of ECG readings performed by cardiologists [10]. The output of the

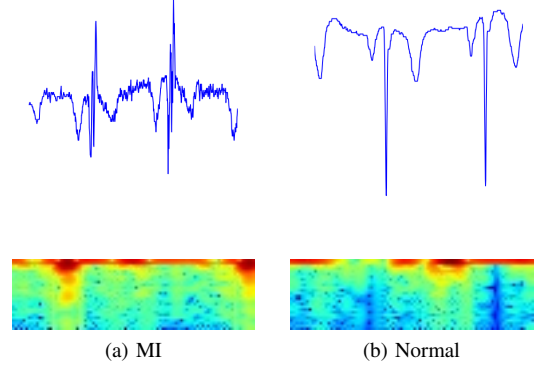


Fig. 2: Examples of ECG waveforms (top row) and their corresponding spectrograms (bottom row) from an aVR lead for a patient randomly selected from each class.

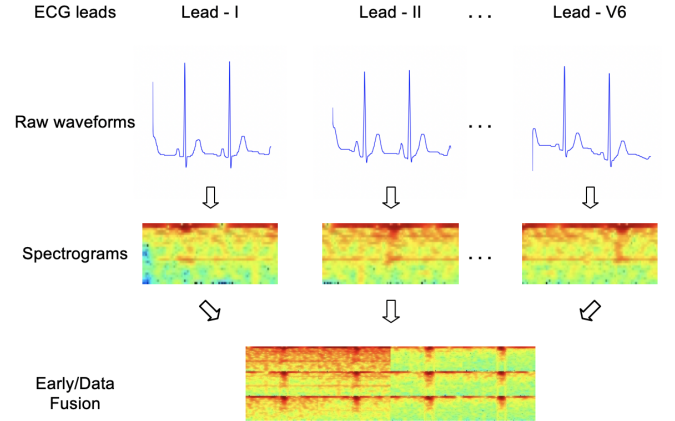


Fig. 3: Early data-fusion of multi-lead information from ECG waveforms.

early fusion is represented by E_w :

$$E_w = \begin{bmatrix} \Gamma_w^I, & \Gamma_w^{aVR}, & \Gamma_w^{V1}, & \Gamma_w^{V4} \\ \Gamma_w^{II}, & \Gamma_w^{aVL}, & \Gamma_w^{V2}, & \Gamma_w^{V5} \\ \Gamma_w^{III}, & \Gamma_w^{aVF}, & \Gamma_w^{V3}, & \Gamma_w^{V6} \end{bmatrix},$$

C. Modelling

After early fusion is done, spectral modelling ($\mathcal{S}(\cdot)$) is applied through the use of CNNs to encode frequency-time characteristics on E_w (see Fig. 4). We propose employing cross-domain transfer learning by using existing computer vision networks, e.g., GoogLeNet [14], to encode dense features taken from their hidden layers, i.e. $\mathbf{d}_w = \mathcal{S}(E_w)$.

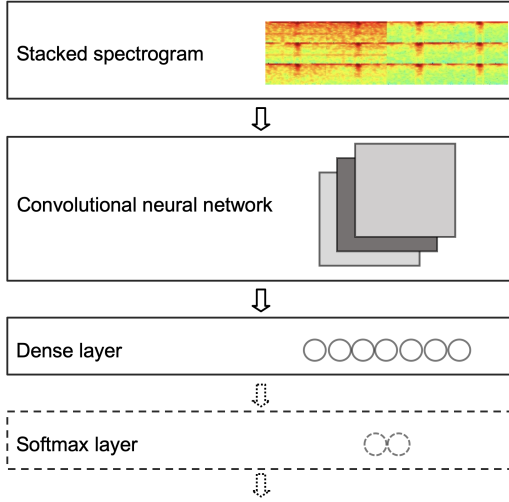


Fig. 4: Spectral modelling that includes a CNN block to encode spectral information.

This helps to avoid training a dedicated spectral-based CNN from scratch, a process that is associated with requiring enormous training data and computational resources. At this stage, the spectral model could be used on its own to predict heart attack, with a softmax layer added to the top of the dense layer. However, in the proposed spectral-recurrent modelling, the output of the dense layer, \mathbf{d}_w , is fed into the recurrent network-based longitudinal modelling step described next.

The longitudinal model, $\mathcal{L}(\cdot)$, is designed to encode the temporal dependency across subsequent windows extracted from a patient using the corresponding spectral dense features, \mathbf{d}_w , obtained from the spectral modelling stage. Thus, the input to $\mathcal{L}(\cdot)$ becomes the temporally sorted dense layer outputs from the spectral model, i.e., $(\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_w, \dots, \mathbf{d}_W)$. We then apply an LSTM-based RNN model to exploit the temporal dependency among subsequent samples extracted from longitudinal ECG data. Such a model is effective in handling the vanishing gradient problem commonly encountered in recurrent networks. Finally, we employed a softmax layer that wraps the output gate results into a heart attack prediction vector.

III. EXPERIMENTS

In this section, we describe the dataset used for the validation of the proposed approach, the setup of parameters in the spectral and longitudinal models, followed by results obtained and their discussion.

A. Experimental Setup

Dataset: The PTB Diagnostic ECG database contains 12 ECG leads from patients diagnosed with multiple heart diseases, sampled at 1000 Hz [12], [11]. However, we consider only the subset of cases relating to MI and health, which amounts to a total of 148 and 52 subjects, respectively. The duration of ECG data may vary across subjects, and hence we only used the first 10 s segment of each patient’s data.

TABLE I: Performance of the proposed framework validated on the PTB dataset.

Methods	Precision	Sensitivity	Specificity	AUROC
Spectral	99%	60%	89%	88%
Longitudinal	96%	83%	85%	90%
Spectral-longitudinal	98%	66%	95%	94%

Setup: In the pre-processing step, we employed a window duration of 1 s to segment subsequent samples with 50% overlapping. Thus, a total of 19 samples are extracted from a 10 s ECG lead. For the spectrogram representation, the STFT is applied on frames of 0.1 s with 90% overlap to provide a smooth frequency-time representation. We use the GoogLeNet architecture [14] for the cross-domain transfer learning, and deep features were extracted from the penultimate hidden layer, which provides 2048-D hidden feature vector followed by a dense layer with 16 neurons. The longitudinal model is made up of a single-layered LSTM network of 8 neurons with a temporal duration of 19 samples. A learning rate of 0.001 is applied with a batch size of 76, which contains two patients per class in each iteration. A five-fold stratified cross-validation is applied to decompose the data into train and test sets. Precision, Sensitivity, Specificity and Area under the receiver operating characteristic (AUROC) are employed as performance metrics. The confusion matrices are also presented, when necessary, to assess the classification errors.

Baseline Methods: We designed two baseline methods, spectral and longitudinal, to assess the efficacy of our proposed joint spectral-longitudinal approach. The spectral method contains CNN features (e.g., extracted from GoogLeNet) followed by a dense and a softmax layer (see Fig. 4). In contrast, the longitudinal model directly takes the GoogLeNet features as input to the LSTM-based network.

B. Results and Discussion

Table I shows the AUROC values obtained for the prediction of MI in the test dataset. The above two baseline methods are used to compare against the proposed. As expected, the baseline spectral and longitudinal methods give inferior performance compared to the proposed spectral-longitudinal model (88%, 90% vs. 94% AUROC, respectively). The longitudinal model alone can not achieve significant improvement over the spectral model, i.e. 88% vs. 90% AUROC. This is partly due to the tendency of the longitudinal model to overfit when presented with a high dimensional (\mathbb{R}^{2048}) GoogLeNet feature input. This can also be observed within the confusion matrices shown in Fig. 5, which can be used to analyse misclassification errors. The longitudinal model achieves the highest detection sensitivity of the MI cases (83%). However, this detection sensitivity comes at the cost of lowest specificity (85%). Our joint spectral-longitudinal model achieves the highest AUROC with an effective trade-off between sensitivity and specificity, by using the dense layer outputs of the spectral model to reduce the high feature dimension of the GoogLeNet features from \mathbb{R}^{2048} to \mathbb{R}^8 . This approach reduces the level of overfitting likely affecting the

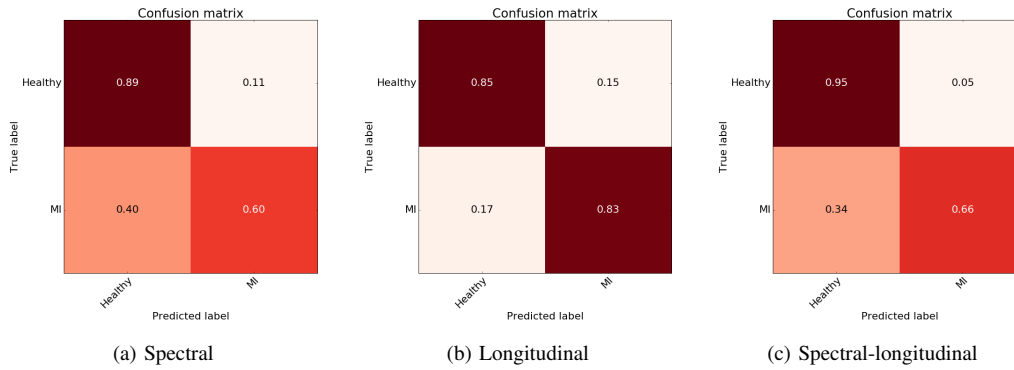


Fig. 5: Confusion matrices of spectral, longitudinal and the proposed spectral-longitudinal models.

longitudinal model, and consequently the highest specificity values are achieved using our proposed model (95%). Generally, our proposed spectral-longitudinal model is able to achieve the lowest misclassification of healthy cases to MI cases (with an error of 5%).

IV. CONCLUSIONS

Cardiologists often inspect patient ECG readings for the diagnosis of CVDs, where MI is the leading cause of mortality amongst CVDs. In this paper, we have proposed a trainable end-to-end joint spectral-longitudinal model to predict heart attack from 12-lead ECG waveforms. We employed data-level integration of these multiple leads, by stacking their spectrograms in a similar arrangement to those used in clinical practice for the visualisation of ECG traces. The spectral model allowed us to transform the time-series ECG signals into a frequency-time representation while enabling cross-domain transfer learning by using existing convolutional neural networks. We encoded the temporal dependency that exists between subsequent samples generated from a patient's data using the longitudinal model. We validated the proposed approach using data from the public PTB dataset. Our experimental results indicated that the joint spectral-longitudinal modelling showed the highest performance (94% AUROC) with an optimal trade-off between sensitivity and specificity. Future work includes the exploration of more efficient integration of multiple-leads, as well as predicting the occurrence time of a heart attack.

ACKNOWLEDGEMENTS

TZ was supported by a Royal Academy of Engineering Research Fellowship.

REFERENCES

- [1] World Health Organization (WHO), "Cardiovascular diseases (CVDs)," URL link: [www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](http://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), Last accessed on 06 May 2020.
- [2] Heart Rate Variability, "Standards of measurement, physiological interpretation, and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.
- [3] Steven A Guidera and Jonathan S Steinberg, "The signal-averaged P wave duration: a rapid and noninvasive marker of risk of atrial fibrillation," *Journal of the American College of Cardiology*, vol. 21, no. 7, pp. 1645–1651, 1993.
- [4] Sarabjeet Mehta, Nitin Lingayat, and Sanjeev Sanghvi, "Detection and delineation of P and T waves in 12-lead electrocardiograms," *Expert Systems*, vol. 26, no. 1, pp. 125–143, 2009.
- [5] S Dash, KH Chon, S Lu, and EA Raeder, "Automatic real time detection of atrial fibrillation," *Annals of biomedical engineering*, vol. 37, no. 9, pp. 1701–1709, 2009.
- [6] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65, 2019.
- [7] Sardar Ansari, Negar Farzaneh, Marlana Duda, Kelsey Horan, Hedvig B Andersson, Zachary D Goldberger, Brahmajee K Nallamothu, and Kayvan Najarian, "A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health records," *IEEE Reviews in Biomedical Engineering*, vol. 10, pp. 264–298, 2017.
- [8] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ecg recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1744–1753, 2018.
- [9] Shensheng Sean Xu, Man-Wai Mak, and Chi-Chung Cheung, "Towards end-to-end ECG classification with raw signal extraction and deep neural networks," *IEEE Journal of Biomedical and Health Informatics*, 2018.
- [10] Girmaw Abebe Tadesse, Tingting Zhu, Yong Liu, Yingling Zhou, Jiyan Chen, Maoyi Tian, and David Clifton, "Cardiovascular disease diagnosis using cross-domain transfer learning," in *Proc. of Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 4262–4265.
- [11] R Bousseljot, D Kreiseler, and A Schnabel, "Nutzung der ekg-signalndatenbank cardiodat der ptb über das internet," *Biomedizinische Technik/Biomedical Engineering*, vol. 40, no. s1, pp. 317–318, 1995.
- [12] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [13] Girmaw Abebe Tadesse, Hamza Javed, N Le Nguyen Thanh, H Duong Ha Thi, Louise Thwaites, Tan Le Van, DA Clifton, and Tingting Zhu, "Multi-modal diagnosis of infectious diseases in the developing world," *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.