

Cardiovascular disease diagnosis using cross-domain transfer learning

Girmaw Abebe Tadesse^{1*}, Tingting Zhu^{1*}, Yong Liu², Yingling Zhou²,
Jiyan Chen², Maoyi Tian³, David Clifton¹

¹ Department of Engineering and Science, University of Oxford, UK

² Guangdong Provincial People's Hospital, China

³ The George Institute for Global Health, China

Abstract—While cardiovascular diseases (CVDs) are commonly diagnosed by cardiologists via inspecting electrocardiogram (ECG) waveforms, these decisions can be supported by a data-driven approach, which may automate this process. An automatic diagnostic approach often employs hand-crafted features extracted from ECG waveforms. These features, however, do not generalise well, challenged by variation in acquisition settings such as sampling rate and mounting points. Existing deep learning (DL) approaches, on the other hand, extract features from ECG automatically but require construction of dedicated networks that require huge data and computational resource if trained from scratch. Here we propose an end-to-end trainable cross-domain transfer learning for CVD classification from ECG waveforms, by utilising existing vision-based CNN frameworks as feature extractors, followed by ECG feature learning layers. Because these frameworks are designed for image inputs, we employ a stacked spectrogram representation of multi-lead ECG waveforms as a preprocessing step. We also proposed a fusion of multiple ECG leads, using plausible stacking arrangements of the spectrograms, to encode their spatial relations. The proposed approach is validated on multiple ECG datasets and competitive performance is achieved.

Index Terms—Transfer Learning, Cardiovascular Disease, Deep Learning, Health Informatics

I. INTRODUCTION

Cardiovascular diseases (CVDs) cause millions of deaths annually and more than three quarters of the world's deaths from CVDs occur in low- and middle-income countries [1]. Traditional diagnosis of CVDs mainly employs interpretation of ECG recordings, which requires precise acquisition devices and high expertise of the clinicians (i.e. cardiologists) that are limited in deprived areas. Cardiologists visually inspect the conventional 12-lead ECG waveforms as images when making diagnosis. However, such process is tedious and can be highly subjective. Automatic approaches have been proposed to extract domain-specific handcrafted ECG features both in time- and frequency-domains for CVD diagnosis [2], [3], [4], [5]. However, these methods lack robustness and do not generalise across variations in patient characteristics and device specifications.

Recently, deep learning (DL) methods have been studied, mainly for atrial fibrillation (AF) detection from ECG ([6], [7], [8], [9], [10]) and photoplethysmogram ([11], [12]) waveforms. However, mostly a dedicated network is designed

and trained from scratch with only problem-specific time-series ECG waveforms [6], [8], [9], [10], without exploiting existing networks from other domains such as computer vision. Spectro-temporal analysis method proposed in [7] is shown to improve AF detection; however, its requirement of prior QRS complex detection makes the proposed approach complex. In addition, existing works mainly utilise single-lead ECG waveform, and thus the fusion of multiple-leads, in the conventional 12-lead ECG acquisition, is left unexploited.

In this paper, we present a cross-domain transfer learning that exploits (existing) vision-based deep neural networks for automatic diagnosis of CVDs from 12-lead ECG recordings. To do so, each ECG lead is analysed in frequency domain and its spectrogram is computed separately followed by normalization and stacking of the spectrograms of the remaining leads. The proposed framework is an end-to-end trainable and enables us (i) to provide cross-domain knowledge transfer from images in computer vision to vital signs in clinical setting; (ii) to avoid training of a dedicated (and sometime complex) network from scratch; and (iii) to encode the intrinsic relationships among ECG leads. Importantly, we analysed and verified the effectiveness of different stacking arrangements of multiple ECG leads on the diagnostic performance, which is the first of its kind to the best of authors' knowledge. The proposed framework is validated on both public (ICBEB) and private (GGH) datasets, each containing thousands of CVD patients.

The paper is organized as follows: Section II presents the proposed framework including the stacked spectrogram computation. Section III provides the experiments, i.e. datasets, parameter setup and results and discussion. Finally, Section IV concludes the paper.

II. METHOD

The proposed method for CVD classification from a 12-lead ECG is shown in Fig. 1. Each lead is segmented into multiple windows, and its frequency-time (spectrogram) representation is obtained by applying a fast Fourier transform (FFT), $\mathcal{F}(\cdot)$. The spectrogram contains the frequency response magnitude at different frequency bins. Let E_n^i be the n^{th} window of the i^{th} ECG lead, the spectrograms can be presented as $\bar{E}_n^i = \mathcal{F}(E_n^i)$ after the FFT. We further

* The first two authors contributed equally to this work.

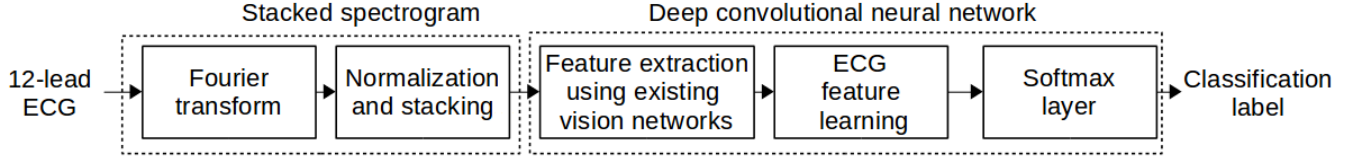


Fig. 1: Overview of the proposed approach

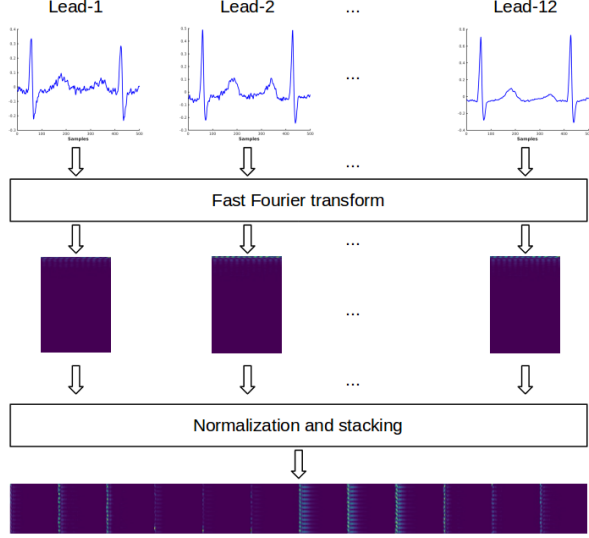


Fig. 2: Steps of generating a stacked spectrogram from a 12-lead ECG.

normalize \bar{E}_n^i as follows

$$\dot{E}_n^i = \frac{\bar{E}_n^i}{\max(\bar{E}_n^i)} \times 255.$$

\dot{E}_n^i exhibits image-like characteristics as the normalization bounds its values to $[0, 255]$ (Fig. 2). This representation enables transfer learning from existing vision models pre-trained on large image datasets, e.g. ImageNet [13]. Spectrogram representation also provides robustness against variation in sampling rate and mounting positions [14]. This approach also helps to reduce the amount of data required for training. A pre-trained GoogLeNet [15] is used to extract hidden-layer CNN features from stacked spectrograms. Later, we build a new hidden layer inside the GoogLeNet pipeline to learn ECG features followed by a Softmax layer for classification.

We also studied different arrangements of stacking the spectrograms from multiple leads and experimented their effects on CVD diagnosis performance. A conventional 12-lead ECG contains the following leads $I, II, III, V1, V2, V3, V4, V5, V6, aVF, aVL$ and aVR , and the corresponding electrodes are often mounted as shown in Fig. 3 (a). The leads are categorized into 4 subgroups, each subgroup represents a vertical stacking of three leads: $G_1 = (I, II, III)$, $G_2 = (V1, V2, V3)$, $G_3 = (V4, V5, V6)$ and $G_4 = (aVL, aVR, aVF)$. We studied three arrangements for stacking the spectrograms (Order-I, Order-II and

TABLE I: Summary of ICBEb dataset. SD: standard deviation; Min: minimum duration; %: percentage of records in each class.

Class	#records	Time unit (s)				%
		Mean	SD	Min	Median	
Normal	918	15.43	7.61	10.00	13.00	13
Atrial fibrillation (AF)	1098	15.01	8.39	9.00	11.00	16
First-degree atrioventricular block (I-AVB)	704	14.32	7.21	10.00	11.27	10
Left bundle branch block (LBBB)	207	14.92	8.09	9.00	12.00	3
Right bundle branch block (RBBB)	1695	14.42	7.60	10.00	11.19	25
Premature atrial contraction (PAC)	556	19.46	12.36	9.00	14.00	8
Premature ventricular contraction (PVC)	672	20.21	12.85	6.00	15.00	10
ST-segment depression (STD)	825	15.13	6.82	8.00	12.78	12
ST-segment elevated (STE)	202	17.15	10.72	10.00	11.89	3
Total	6877	15.79	9.04	6.00	12.00	100

Order-III). Order-I = (G_1, G_2, G_3, G_4) and Order-II = (G_1, G_4, G_2, G_3) further stack the subgroups vertically as shown in Fig. 3 (b) and Fig. 3 (c), respectively. Finally, Order-III stacks the subgroups in similar order as Order-II but horizontally as shown in Fig. 3 (d).

III. EXPERIMENTS

A. Datasets

The ICBEb dataset is obtained from the China physiological signal challenge at the 7th International Conference on Biomedical Engineering and Biotechnology (ICBEb)¹. ICBEb contains 12-lead ECG records from 6,877 subjects, among which 5,959 subjects are diagnosed with one (or more) of eight cardiovascular abnormalities and the remaining 918 subjects are diagnosed with normal conditions (see Table I). Among 5,959 abnormal cases, 477 are diagnosed with two or more CVDs. The ECGs are sampled at 500 Hz and their temporal duration ranges between 6 and 60 seconds. The abnormalities are *atrial fibrillation (AF)*, *first-degree atrioventricular block (I-AVB)*, *left bundle branch block (LBBB)*, *right bundle branch block (RBBB)*, *premature atrial contraction (PAC)*, *premature ventricular contraction (PVC)*, *ST-segment depression (STD)* and *ST-segment elevated (STE)*. Though the majority of the ICBEb classes have balanced number of samples, a few suffer from an imbalance problem. Particularly, each of LBBB and STE contains only 3% of the whole samples in the dataset. In addition to the abnormality label, the meta data also contains age and gender of the subject that could be exploited in future validations.

The GGH dataset is collected in southern areas of China, and contains 12-lead ECG waveforms from 21,241 anonymised patients, i.e. 15,578 myocardial infarction (MI) and 5,663 normal cases, sampled at 500 Hz. MI cases could be either ST-elevation myocardial infarction (STEMI) and non-STEMI. In addition to the conventional leads, GGH also

¹<http://www.icbeb.org/Challenge.html>

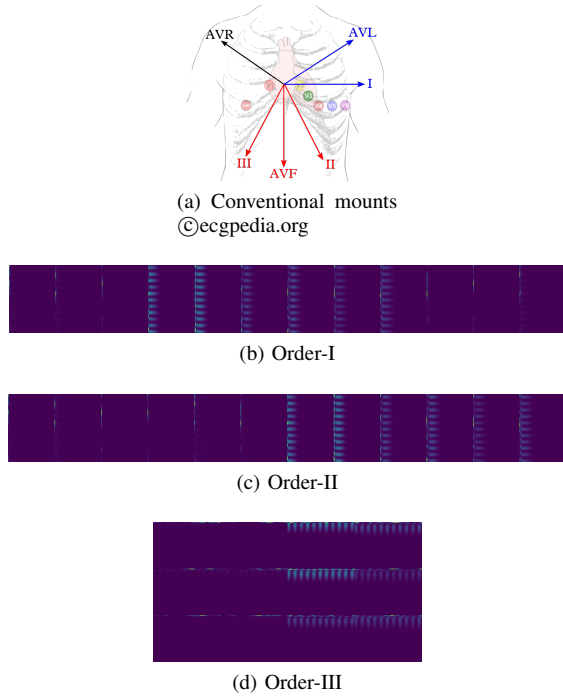


Fig. 3: Visualisations of different arranging orders experimented to stack spectrograms from a conventional 12-lead ECG. Note Order-I and Order-II are rotated by $+90^\circ$ for better visualisation.

contains $V3r$, $V4r$ and $V5r$ leads. After preprocessing, we found that 3,722 myocardial infarction (MI) cases contain multiple ECG waveforms with zero values (i.e. flat) and hence were discarded from the upcoming experiments. Only the common 12 leads are considered and additional 3 cases that do not contain all of the 12 conventional leads are avoided. 136 normal cases are further discarded with similar procedures. In total, data from 11,853 MI and 5,528 normal patients in GGH dataset are used in the experiments. Note that there is imbalance among classes in GGH dataset as the MI cases are at least twice more than the normal cases.

B. Parameter setup

In this section we describe the setup of parameters employed during validation. First, we set the window length of a sample to be 10 s and a spectrogram is computed for each window. No overlapping is applied between samples generated from a patient data. A Hamming window is employed for a short-time Fourier transform computation of each chunk (1 s) with 95% overlapping between subsequent chunks in order to obtain a smooth spectrogram representation. Since the energy is mainly accumulated in the low frequency band, only the first 25% of the frequency coefficients are used in order to avoid redundancy and reduce the spectrogram dimension to 125×181 per lead.

For CNN feature extraction, we used inception-v3 of GoogLeNet, particularly its next-to-last layer (i.e. ‘pool 3 : 0’), which provides a feature dimension of 2,048. To avoid overfitting, we keep the dimension of the

new hidden layer to 10 and a ReLU activation is employed on it. The dimension of the Softmax layer resembles the number of CVDs in each corresponding datasets, i.e. nine for ICBEb and two for GGH. For the end-to-end training, Adam optimizer is used with learning rate = 0.001, training steps = 10,000, training batch size = 128. 80%-train and 20%-test split is applied. The whole experiment is repeated 10 times and average accuracy is computed. Due to the class-imbalance in the datasets, the *sparse-softmax-cross-entropy* is utilised as a classification loss. In addition, batch normalization is applied on the new hidden layer. The framework is implemented in Tensorflow platform with Python 3.5. As a baseline, we validate the CNN features using SVM within MATLAB 2017a environment. Both linear and Gaussian kernels are experimented.

C. Results and discussion

We first experimented with the three stacking arrangements, i.e. Order-I, Order-II and Order-III, on the GGH dataset across different classification methods and results are shown in Fig. 4. The methods include SVM validation with linear and Gaussian kernels, end-to-end training with Softmax layer, and end-to-end (EE) training with a new hidden layer (HL) and a Softmax layer. Order-III achieves superior performance compared to Order-I and Order-II, consistently across the four methods. This suggests that Order-III better encodes the spatial relationships among the leads. In addition, Order-III benefits from its square-like representation compared to rectangular Order-I and Order-II representations (Fig. 3). This is because GoogLeNet framework automatically crops input images to 299×299 , as a result Order-I and Order-II lose much of their frequency information. Though the dimensions of the final spectrograms in Order-I and Order-II are exactly equal, the results show that Order-II consistently outperforms Order-I, which further strengthens the need to encode the spatial relationships among ECG leads.

For the remaining experiments, Order-III is employed. The proposed framework provides encouraging performance when validated on both ICBEb and GGH datasets (see Table II). As expected, the raw CNN features are less discriminant as demonstrated in the inferior baseline performance of SVM (linear) and (Gaussian). Gaussian kernel outperforms linear kernel in both datasets. Later, EE (HL+Softmax) is shown to improve the ICBEb baseline performance from 43.6% of SVM (Gaussian) to 49.9%. Similarity, EE (Softmax) improved the GGH performance from 79.1% to 85.8%. However, the new hidden layer tends to slightly degrade the performance for GGH as the model starts to overfit due to the high-dimension of CNN features from inception-v3.

The GGH classification task (binary) is less challenging compared to the nine-class classification task of ICBEb. As a result, lower diagnostic performance resulted on ICBEb, which is partly due to the lack of sufficient number of training data available as there are only ≈ 611 samples per class for training. Though about 8% of abnormal patients in ICBEb are diagnosed with multiple CVDs, we used only a

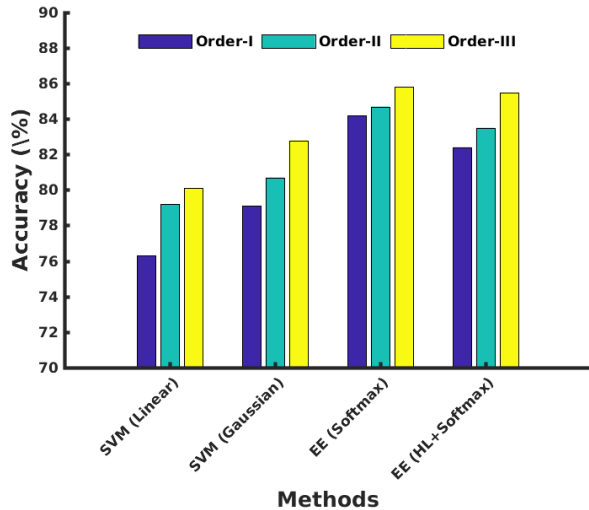


Fig. 4: Effect of different stacking orders on the classification performance validated on the GGH dataset.

TABLE II: Results of the proposed framework compared against SVM-based baseline and validated on both ICBE and GGH datasets. Note that Order-III is used during stacking. EE: end-to-end training; HL: new hidden layer.

Methods	Accuracy (%)	
	ICBEB	GGH
SVM (Linear)	36.2 ± 0.2	80.1 ± 0.0
SVM (Gaussian)	43.6 ± 0.4	82.8 ± 0.1
EE (Softmax)	49.8 ± 1.2	85.8 ± 0.9
EE (HL+Softmax)	49.9 ± 0.7	85.5 ± 0.4

single label ground truth for each case. As a result, we might mistakenly treat it as incorrect when the inference provides one of the other labels in multi-label cases. Furthermore, our proposed model considered all ECG leads as images while techniques reported to achieve higher accuracy in the challenge required extraction of each ECG cycle, treated each lead independently, and exploited additional temporal encoding frameworks such as long short-term memory (LSTM) recurrent networks. Generally, compared to a random guess of 11%, it is encouraging to obtain $\approx 50\%$ accuracy on ICBE using cross-domain knowledge transfer learning, without explicitly training a dedicated CNN network. The classification performance on GGH dataset could be further improved by sub-labelling the MI cases into to STEMI and non-STEMI categories and apply three-class classification.

IV. CONCLUSIONS

We proposed an end-to-end trainable cross-domain transfer learning for data-driven diagnosis of CVDs (from ECG waveforms) that potentially supports decision-making and hence improves patient care. Our stacking spectrogram representation of multi-lead ECG enables us to utilise existing pre-trained vision-based deep networks as feature extractors. We also analysed the fusion of multiple ECG leads using a plausible spectrogram stacking arrangements. The proposed

framework is validated on two datasets and results show that competitive performance is achieved using transfer learning without training a dedicated network from scratch. Future work includes reducing the high-dimension of CNN features using an auto-encoder that can optimises the reconstruction and classification errors simultaneously. Furthermore, cross-modal knowledge transfer could be adopted to exploit multi-sensor information beyond ECG waveforms.

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 23 April 2019.
- [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] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.
- [7] Z. Zhao, S. Sarkka, and A. B. Rad, "Spectro-temporal ECG analysis for atrial fibrillation detection," in *In Proc. of IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, 2018, pp. 1–6.
- [8] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscale 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] Bahareh Pourbabaee, Mehrsan Javan Roshkhar, and Khashayar Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, , no. 99, pp. 1–10, 2017.
- [10] Zhenjie Yao, Zhiyong Zhu, and Yixin Chen, "Atrial fibrillation detection by multi-scale convolutional neural networks," in *Proc. of IEEE International Conference on Information Fusion (Fusion)*, 2017, pp. 1–6.
- [11] Ming-Zher Poh, Yukkee Cheung Poh, Pak-Hei Chan, Chun-Ka Wong, Louise Pun, Wangie Wan-Chiu Leung, Yu-Fai Wong, Michelle Man-Ying Wong, Daniel Wai-Sing Chu, and Chung-Wah Siu, "Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms," *Heart*, 2018.
- [12] Supreeth Prajwal Shashikumar, Amit J Shah, Qiao Li, Gari D Clifford, and Shamim Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *Proc. of IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 2017, pp. 141–144.
- [13] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.
- [14] Girmaw Abebe and Andrea Cavallaro, "Inertial-vision: cross-domain knowledge transfer for wearable sensors," in *Proc. of International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, 2017, vol. 7.
- [15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.