DeepMI: Deep Multi-lead ECG Fusion for Identifying Myocardial Infarction and its Occurrence-time

Girmaw Abebe Tadesse[a,b,*], Hamza Javed[a], Komminist Weldemariam[b], Yong Liu[c], Jin Liu[c], Jiyan Chen[c], Tingting Zhu[a]

[a]*Department of Engineering, University of Oxford, Oxford, United Kingdom*
[b]*IBM Research, Kenya, Africa*
[c]*Department of Cardiology, Guangdong Cardiovascular Institute, Guangdong Provincial Key Laboratory of Coronary Disease, Guangdong Provincial People's Hospital, University of Technology, Guangzhou, China*

Abstract

Myocardial Infarction (MI) has the highest mortality of all cardiovascular diseases (CVDs). Detection of MI and information regarding its occurrence-time in particular, would enable timely interventions that may improve patient outcomes, thereby reducing the global rise in CVD deaths. Electrocardiogram (ECG) recordings are currently used to screen MI patients. However, manual inspection of ECGs is time-consuming and prone to subjective bias. Machine learning methods have been adopted for automated ECG diagnosis, but most approaches require extraction of ECG beats or consider leads independently of one another. We propose an end-to-end deep learning approach, DeepMI, to classify MI from normal cases as well as identifying the time-occurrence of MI (defined as acute, recent and old), using a collection of fusion strategies on 12 ECG leads at data-, feature-, and decision-level. In order to minimise computational overhead, we employ transfer learning using existing computer vision networks. Moreover, we use recurrent neural networks to encode the longitudinal information inherent in ECGs. We validated DeepMI on a dataset collected from 17,381 patients, in which over 323,000 samples were extracted per ECG lead. We were able to classify normal cases as well as acute, recent and old onset cases of MI, with AUROCs of 96.7%, 82.9%, 68.6% and 73.8%, respectively. We have demonstrated a multi-lead fusion approach to detect the presence and occurrence-time of MI. Our end-to-end framework provides flexibility for different levels of multi-lead ECG fusion and performs feature extraction via transfer learning.

1. Introduction

[1]Cardiovascular diseases (CVDs) are the leading cause of death globally, and 85% of CVD deaths are due to heart attacks (i.e., Myocardial Infarction) and strokes[1]. Traditional diagnosis of heart attack mainly employ interpretation of ECG recordings, which requires precise acquisition devices and highly trained clinicians (i.e., cardiologists), both of which are in limited supply in resource-constrained areas. Cardiologists visually inspect the conventional 12-lead ECG waveforms as images when making diagnosis. However, such a process is tedious and can be highly subjective [2]. ECG readings are also sensitive to mounting position and prone to movement artefacts [3], resulting in noisy readings that add to the difficultly of making reliable diagnoses from them. This problem becomes more visible in low-resource settings where there are a limited number of cardiologists, or none at all. Numerous studies have also shown that it is not always pos-

*Corresponding author:
Email address: girmaw.abebebe.tadesse@ibm.com (Girmaw Abebe Tadesse)

[1]Declaration of Interest: None

sible to detect cardiovascular abnormalities from a visual inspection of the ECG trace alone, given the small amplitude and short PQRST durations involved [4]. In this context, computer aided diagnosis methods present a promising solution to the problem of analysing and identifying CVDs from ECG readings.

Automatic approaches using computer algorithms have been proposed to extract domain-specific handcrafted ECG features both in time- and frequency-domains for heart disease diagnosis [5, 6, 7, 8]. Manually engineered features (e.g., identifying complexes in the ECG trace) could be tractable compared to automatically learned features; however, they might be susceptible to noise/motion artefacts. Moreover, these approaches often require extensive pre-processing steps that are prone to error. Feature engineering also requires careful consideration and is associated with a time-consuming and labour-intensive model development process, which often fails to take into account variations in patient characteristics, mounting positions and device specifications. By contrast, end-to-end deep learning methods can eliminate the need for explicit feature engineering, by learning optimal representations directly from the raw data for the task at hand. Additionally, there are currently no studies that have investigated the prediction of occurrence-time in MI (i.e., the age of a MI), information that is crucial for preoperative risk assessment [9], where patients with acute (i.e., within seven days) and recent (i.e., less than 30 days but longer than seven days) MIs are considered to be at higher risk of a perioperative cardiac event, and patients with old MIs (i.e., more than 30 days) are at higher risk of perioperative cardiac morbidity. Such systems become more impactful, in low-resource settings, where there is no patient data archival practice; or in emergencies when a patient history might not be available on the spot.

In this paper, we present an end-to-end deep learning system for predicting the occurrence time of MI using 12-lead ECG waveforms (see Fig. 1). The contributions of this work are as follows:

1. spectral-based data representation for onset time detection of heart attack, which offers a more general representation that overcome issues of variability in sampling rates and device specifications across ECG manufactures;
2. application of cross-domain transfer learning between natural images classification and ECG waveform detection, to extract spectral features to reduce the dimension of the spectrograms and minimise redundant information;
3. utilising joint spectral-temporal modelling to encode the spatial and temporal information from multi-channel ECG waveforms;
4. proposing the use of a variety of fusion approaches to combine information at different levels of data representation (i.e., data-, feature- and decision-level fusions) aiming to utilise distinctive characteristics offered by each ECG lead; and
5. predicting the occurrence time of MI, utilising a large scale dataset containing $> 15,000$ patients and $> 323,000$ data samples, to provide timely intervention that can potentially improve patient outcomes.

The paper is organised as follows: Section 2 surveys the research literature on the topic of automated ECG analysis for CVD diagnosis, with a focus on MI detection. Section 3 presents the proposed framework, including different multi-lead fusion strategies and diagnosis modelling techniques. Section 4 details the experimental setup including descriptions of the dataset used for validation, the set up of network architectures and parameters in the proposed approach, in addition to methods considered for comparison. Section 5 presents results and discusses important findings, whilst concluding remarks are provided in Section 6.

2. Related work

The potential diagnostic value provided by ECG signals, particularly for the diagnosis of CVDs such as MI, has long been recognised by researchers [4]. To date, the majority of automated methods proposed for detecting MI from ECG

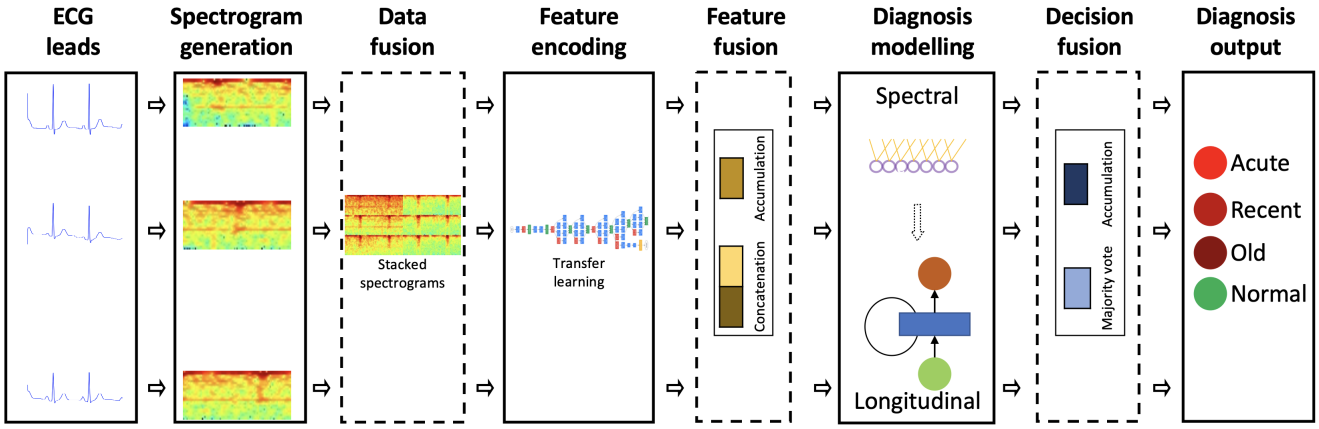


Figure 1: Overview of the proposed framework: multi-lead ECG waveforms are provided as input, and spectrogram generation is used to encode the frequency-time characteristics. Transfer learning is applied to encode deep features using existing computer vision networks. Spectral and temporal models were employed for diagnosis modelling.

traces, have focused on identifying abnormalities after beat-by-beat segmentation and manually extracting statistical and morphology-based features from the PQRST complexes of the segmented beats. These approaches often entail error-prone pre-processing steps (such as identifying ECG complexes), which due to the small amplitudes and short duration of segmented beats are particularly sensitive to noise and motion artefacts. This is followed by a handcrafting of features which requires domain expertise and careful consideration and is associated with a time-consuming and laborious model development process [10, 11, 12, 13, 14].

The handcrafted features extracted from the ECG signals can then be used to make classification decisions based either on simple rule-based thresholding [15], or by training supervised ML models [16], including deep learning models [17, 18]. Feature engineering requires careful consideration and is associated with a time-consuming and laborious model development process [12]. For example, in the work of [12], a separate feature selection step is first performed to identify the 15 most informative discrete wavelet transform (DWT) features from the set considered. Moreover, in the numerous studies published, highly varied feature sets have been proposed for MI classification. The absence of an established consensus on informative features may indicate the lack of generalisabilty provided by the handcrafted features proposed to date. Deep learning methods by contrast, remove the need for beat segmentation and manual feature engineering through their ability to automatically learn optimal representations directly from the raw data.

2.1. Deep learning methods

Given the success of deep learning in a wide range of application domains, a growing number of studies have investigated the use of end-to-end deep networks for ECG based CVD detection [19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. The most widely employed and often best performing architectures have been CNNs, e.g. [20, 23, 27, 28]. A smaller number of these studies considered RNNs [26, 29], whilst [25, 29] demonstrated that a cascaded CNN-LSTM model architecture achieved superior performance to either individual model architectures on their own. This provides supporting evidence to the broad fact that capturing local and long-term temporal characteristics of ECG signals through joint modelling, upon which this study is built on.

More recently, the following are examples of studies that have employed end-to-end deep learning approaches for ECG based MI detection specifically [23, 20, 21, 27, 29, 28]. For the majority of these works, MI detection was framed solely as a binary classification problem (non-MI or MI), whilst in others MI localisation was

the focus through a multi-class problem formulation. The latter concerned identifying different MI types characterised by the location of the blocked artery (e.g. posterior lateral MI), as an example, 10-classes of MI location are predicted in the work of [27]. The utility of discriminating between such MI types is clear, however prediction of MI by time-onset, as we propose in this work, has not yet been researched. Characterising MI diagnosis by estimated occurrence time has clear practical implications, as this additional information would enable proactive care management and appropriate intervention strategies to be enacted.

2.2. Transfer learning

Whilst deep learning methods have the potential to achieve state-of-the-art performance, they often require copious amounts of training data. This is particularly true when a sophisticated architecture is trained from scratch, which contains many parameters that must be learnt. To overcome these challenges in data and time-consuming hyper-parameter setup, transfer learning can be exploited. Transfer learning aims to leverage models that are pre-trained, namely models that have been carefully developed for a different task and/or on a different dataset. For the task of CVD detection using ECGs, the works of [22, 30] showed the effectiveness of cross-domain transfer learning through their use of Google's Inception image recognition model [31]. In these works, raw ECG signal traces were treated as images to be analysed by the network, thereby directly emulating the analysis cardiologists perform. Our work differs from these, and other similar studies, in that we first map 1D ECG signals to 2D spectrogram representations, before treating the latter as an image to be analysed by a pre-trained computer vision network through transfer learning. We consider the spectrogram representation as input, rather than the image of the ECG signal itself, as the former allows spectral time-varying information present in the signal to be better visualised. Thereby potentially enabling more clinically relevant information for diagnosis to be captured and learnt by an ML model. In previous work by the authors, we have shown the effectiveness of spectrogram image representations of physiological signals, such as PPG, for ML based disease diagnosis [32, 33].

2.3. Single vs. multiple ECG leads

Published research in CVD detection from ECG signals, has primarily restricted itself to using a single lead [19, 20]. In the studies where multiple ECG leads were considered, individual models were generally developed and evaluated for each lead separately [27]. Fusion strategies at both the data and result level were investigate extensively in this work. Whilst the use of 12 leads ECG may usually be limited outside clinical settings, and single lead analysis is advantageous in certain contexts (for example remote monitoring of patients), the work of [34] highlights the prevalence and seriousness of in-hospital MI. In these contexts, patients can be consistently and conveniently monitored using the standard and inexpensive 12 lead ECG setup, which clinicians use in their entirety to make diagnoses. Importantly, research has shown that different ECG traces contain different predictive importance for MI diagnosis [23], highlighting the value of considering multiple lead fusion for CVD prediction in general [35, 23, 25, 36].

2.4. Validation datasets

Research methods proposed to date for automatic CVD detection, have primarily developed and evaluated their methods on the open-source PhysioBank Physikalisch-Technische Bundesanstalt (PTB) ECG dataset [37], or on proprietary datasets (e.g. [25]). High performance achieved on the detection of MI or MI types in the PTB dataset, may indicate that to assess the generalisability of developed methods, validation on new datasets is vital. To date, studies that have employed alternative datasets have often been small. For example in [25] models were trained on ECG records 10 seconds long from 362 patients with cardiovascular abnormalities. By contrast, we have developed our models using a dataset with over $17,000$ patients and $323,000$ ECG sample measurements, which is an order of magnitude higher than those used

in the literature. Moreover, the dataset used in this work contain additional labels regarding the eventual time occurrence of MI, allowing us to investigate the possibility of predicting time-onset of heart attack from ECG readings alone.

3. Proposed framework: DeepMI

In this section, we outline the system aspects of the proposed framework for identifying MI and its occurrence time.

3.1. Overview

Figure 1 shows the overview of our proposed approach. Given a dataset, $\mathcal{D}$, which contains η patients diagnosed for MI, i.e., $\mathcal{D} = \{\mathcal{S}_i\}_{i=1}^{\eta}$, where $\mathcal{S}_i$ represents the i^{th} patient represented by 12 ECG leads, $\mathcal{S}_i = \{\mathbf{l}_i^j\}$, $j \in [I, II, III, aVR, aVL, aVF, V_1, V_2, V_3, V_4, V_5, V_6]$. Each patient is diagnosed and the onset time of MI cases is determined clinically. Thus, the ground truth of $\mathcal{S}_i$ is $g_i \in \{acute, recent, old, normal\}$. Thus, our approach focuses on predicting the diagnose label using an end-to-end trainable framework designed to operate with limited training data and in low computational resource settings. To this end, we, first, extract the frequency-time characteristics, i.e., spectrogram, of the ECG waveforms. Then feature encoding is done on these spectrograms via transfer learning using pre-trained computer vision networks. Diagnosis modelling is performed with spectral and longitudinal models, which encodes spatial and temporal information, respectively. To integrate information from different ECG leads, we employ different fusion techniques. Next we describe the details of each step in the proposed framework.

Given the raw ECG waveforms, a preprocesing is applied to clean up the data, which includes filtering of patients with erratic ECG readings in one or more of the ECG leads. Particularly, missing readings which could occur due to loose mounting of the electrodes during ECG acquisition. To discard some of the noise/motion artefacts, often characterized by low and high frequency characteristics, we employed a band-pass filter (a high pass filter followed by a Gaussian filter) on all ECG recordings. Other pre-processing steps include the subsequent sampling of ECG windows per lead, i.e., $\mathbf{l}_i^j = [\mathbf{w}_{in}^j]_{n=1}^{\gamma}$, where $\mathbf{w}_{in}^j$ represents the n^{th} window segmented from the j^{th} lead of the i^{th} patient and γ is the total number of windows from a patient lead, $\mathbf{l}_i^j$.

3.2. Spectrogram generation

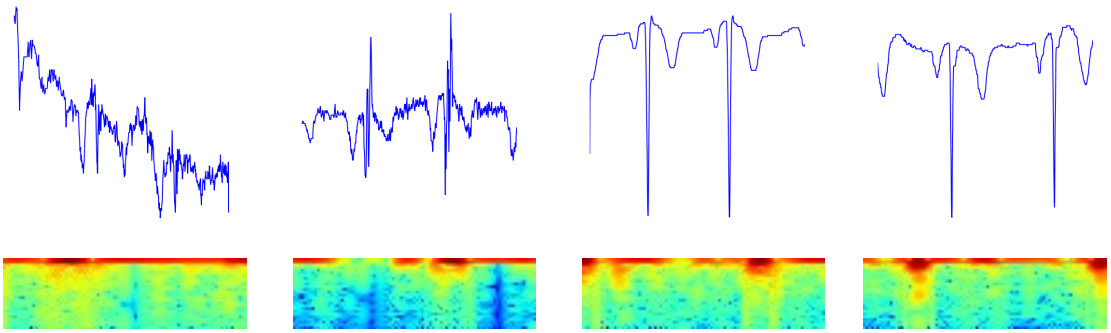
Rather than using the raw ECG time-series to encode features for the MI diagnosis, the proposed framework employs a frequency-time (spectrogram) representation that captures the time-varying characteristics of the waveforms, and it is also robust across variations in device specifications such as sampling rate [38]. The spectrogram generation is computed from each $\mathbf{w}_{in}^j$ by applying a Fourier transform, that results in a 2D frequency-time representation F_{in}^j. The subscript i is dropped henceforth to improve readability. We also normalise spectrograms by their maximum value, followed by a logarithmic scale to smooth the representation as

$$\hat{F}_n^j = \log\left(\frac{F_n^j}{max(F_n^j)} * 255\right). \quad (1)$$

The smoothed $\hat{F}_n^j$ is then stored as image using the JPEG image format so that it resembles natural images for the transfer learning step. While other colormaps could also be used, we employed the 'viridis' colormap to convert the spectrogram array to an image after normalization in Eq. (1). The JPEG format is selected due to its efficiency but other formats could also be used.

3.3. Feature encoding

The spectrogram generation provides an image-like representation that suits convolution-based deep learning encoding. However, designing a dedicated deep network and encoding features by training it from scratch poses an enormous training data requirement, computational resources to train and a prolonged hyper-parameter tuning process. As a result, we opt to



(a) Acute (b) Recent (c) Old (d) Normal

Figure 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. Note: only 2-seconds duration of each waveform is shown to allow detailed visualisation of the patterns.

utilise existing computer vision networks, such as GoogLeNet [31] and MnasNet [39], to encode features via a transfer learning approach. Thus, a hidden layer of these existing CNNs could be extracted from each normalised spectrogram input, $\hat{F}_n$, resulting a high-dimensional feature vector, $\mathbf{d}_n \in \mathbb{R}^{\tau}$. The use of such cross-domain learning exploits powerful existing architectures and relaxes the training data requirement.

3.4. Multi-lead fusion

The proposed framework offers a variety of fusion techniques to utilise the discriminant characteristics offered by the conventional 12 ECG leads [23]. To this end, *data*, *feature* and *decision* fusions are experimented with, details of which are provided below.

3.4.1. Data fusion

The spectrograms from different ECG leads are fused together at data level in the form of stacked spectrograms. Given the set of normalised spectrograms of the leads, $\{\hat{F}_n^j\}$ for $j \in [I, II, III, aVR, aVL, aVF, V_1, V_2, V_3, V_4, V_5, V_6]$, the data fusion is applied as in Equation (2), which mimics the visualisation of the ECG leads by cardiologists in clinical practice [32]. This results in a stacked spectrogram, Φ_n, as the output of the data fusion.

$$\Phi_n = \begin{bmatrix} \hat{F}_n^{I}, & \hat{F}_n^{aVR}, & \hat{F}_n^{V1}, & \hat{F}_n^{V4} \\ \hat{F}_n^{II}, & \hat{F}_n^{aVL}, & \hat{F}_n^{V2}, & \hat{F}_n^{V5} \\ \hat{F}_n^{III}, & \hat{F}_n^{aVF}, & \hat{F}_n^{V3}, & \hat{F}_n^{V6} \end{bmatrix}. \quad (2)$$

In case of data fusion, transfer learning is applied to encode a single deep feature vector, $\mathbf{d}_n$, from the stacked spectrogram, Φ_n.

3.4.2. Feature Fusion

Another information fusion strategy from the multiple ECG leads could be implemented at the feature level. Namely, after deep features are extracted from the spectrogram of each lead, i.e., $\mathbf{d}_n^j$ for $j \in [I, II, III, aVR, aVL, aVF, V_1, V_2, V_3, V_4, V_5, V_6]$, where each $\mathbf{d}_n^j$ is extracted from the corresponding $\hat{F}_n^j$. We have experimented with two feature fusion approaches: *concatenation* and *accumulation*. Concatenation refers to the stacking of the feature vectors to a single vector, i.e. the input to the modelling step becomes $\bar{\mathbf{d}}_n = \mathcal{C}(\mathbf{d}_n^j)$, where $\mathcal{C}(\cdot)$ represents the concatenation operation and, as a result, $\bar{\mathbf{d}}_n \in \mathcal{R}^{N_l \times \tau}$, where N_l represents the number of leads (i.e., $N_l = 12$) and τ represents the dimension of each $\mathbf{d}_n^j$. On the other hand, accumulation-based feature fusion, $\mathcal{A}(\cdot)$, results in $\hat{\mathbf{d}}_n \in \mathcal{R}^{\tau}$, with equal dimension as the individual feature vectors due to the accumulation operation being defined as:

$$\hat{\mathbf{d}}_n = \mathcal{A}(\mathbf{d}_n^j) = \frac{\sum_{l=1}^{N_l} \mathbf{d}_n^j}{N_l}. \quad (3)$$

3.4.3. Decision Fusion

Decision fusion aims to exploit the distinctiveness in each lead by training a specific diagnosis model for each lead, after which the final output

is obtained from the fusion of predictions from these lead-specific models. Given 12-lead ECGs, the diagnosis prediction specific to each lead results in $\mathbf{p}_n^j \in \mathbb{R}^{N_d}$, where N_d is the number of diagnosis classes. To obtain the final prediction output from these different lead models, we propose two different decision fusion techniques: *accumulation* and *majority vote*. Decision accumulation is applied similarly to feature fusion as in Equation (3) and the average of lead-specific prediction $\mathbf{p}_n^j$ is computed. Majority vote, on the other hand, selects the most frequently predicted diagnosis class, across the different model leads, as the final predicted diagnosis.

3.5. Diagnosis modelling

After deep features are obtained in the transfer learning module, three diagnosis modelling techniques are developed in the proposed framework: *spectral* and *longitudinal* and joint *spectral-longitudinal*. These techniques are described more detail below.

3.5.1. Spectral modelling

This modelling approach only uses the spectral information, encoded as deep features, for diagnosis modelling. To do so, a dense layer is applied that takes the deep feature vector as input; which means $\mathbf{d}_n$ during data fusion, $\bar{\mathbf{d}}_n$ during concatenated feature fusion, $\hat{\mathbf{d}}_n$ during accumulated feature fusion and $\mathbf{d}_n^j$ during decision fusion. The output of the dense layer is $\mathbf{r}_n \in \mathbb{R}^{\kappa}$, where κ is the dimension of the dense layer. The dense layer helps to refine the spectral features and reduce the feature dimension (since $\kappa < \tau$) as

$$\mathbf{r}_n = \sigma(W_{rd}\mathbf{d}_n + \mathbf{b}_r), \tag{4}$$

where σ is a non-linear activation function, $W_{rd} \in \mathbb{R}^{\kappa \times \tau}$ is the weight matrix linking the deep feature vector and the dense layer, and $\mathbf{b}$ is the bias vector. Note that during feature concatenation, τ becomes $N_l = 12$ longer. The prediction vector for the MI diagnosis, $\mathbf{p}_n$, is obtained by applying a softmax function on the dense layer output, $\mathbf{r}_n$, as

$$\mathbf{p}_n = \frac{e^{W_{sr}\mathbf{r}_n}}{e^{W_{sr}\mathbf{r}_n} + 1}, \tag{5}$$

where $W_{sr} \in \mathbb{R}^{N_c \times \kappa}$ is the weight matrix linking the dense layer and diagnosis classes and N_c is the number of class labels. The diagnosis class corresponding to the index of the highest element in the prediction vector, $\mathbf{p}_n$, becomes the final prediction output by the proposed framework.

3.5.2. Longitudinal modelling

The spectral modelling above only encodes the spectral information in each sample window, $\mathbf{w}_n$, segmented from a long ECG waveform. However, subsequent samples possess temporal dependency as γ samples were segmented from each ECG lead. We propose a recurrent neural network to exploit these long-term temporal dynamics. Long short-term memory (LSTM) networks are designed to encode temporal dependency, and they handle the vanishing and exploding gradient problems better than the vanilla recurrent neural networks (RNNs) via their *input*, *output* and *forget* gates that act as switches to control memory information about the past.

Given the previous cell information, $\mathbf{c}_{n-1}$, the output gate, $\mathbf{o}_n$, the forget gate, $\mathbf{f}_n$, the candidate cell information, $\bar{\mathbf{c}}_n$, and input gate, $\mathbf{i}_n$; the current LSTM hidden state, $\mathbf{h}_n$, can be computed as

$$\mathbf{h}_n \;=\; \mathbf{o}_n \odot \phi(\mathbf{f}_n \odot \mathbf{c}_{n-1} + \mathbf{i}_n \odot \bar{\mathbf{c}}_n)\;, \tag{6}$$

where $\odot$ is an element-wise multiplication and ϕ is a tanh activation function. The dimension of $\mathbf{i}_n$, $\mathbf{h}_n$, $\mathbf{c}_{n-1}$, $\mathbf{c}_n$, $\mathbf{o}_n$ and $\bar{\mathbf{c}}_n$ is $\mathbb{R}^{\nu}$, which is the number of neurons in the LSTM layer. Note that when only the longitudinal model is used, its input is the high dimensional deep features. Finally, a softmax function similarly to Equation (5) is applied on $\mathbf{h}_n$ to obtain the MI diagnosis prediction, $\mathbf{p}_n$, using the longitudinal model.

3.5.3. Spectral-longitudinal modelling

Separate use of the spectral and longitudinal modelling approaches have the following limitations. The spectral model fails to encode the long-term temporal dependency existing between subsequent samples segmented from long duration ECG waveforms. On the other hand, the longitudinal model takes as input the high dimensional

deep features (τ during data, feature accumulation and decision fusion, 12τ during feature concatenation) obtained from the transfer learning step, which might result in overfitting due to the curse of dimensionality, especially when feature concatenation is used. We therefore propose a joint spectral-longitudinal model to address these limitations and utilise the advantages offered by both models, i.e. refinement of deep features and dimensionality reduction by the dense layer of the spectral model and the temporal encoding using LSTM in the longitudinal model. Thus, the output of the dense layer in the spectral model, $\mathbf{r}_n$, is used as input to the longitudinal model. As a result, the hidden state, $\mathbf{h}_n$, in the longitudinal model encodes the temporal dependency among subsequent $\mathbf{r}_n$. Finally, the softmax layer is used to predict the MI diagnosis, $\mathbf{p}_n$, similarly to Equation (5).

4. Experimental Setup

In this section, we describe details of the dataset used for development and validation of onset time prediction, the set up associated with the different stages of the proposed framework. Namely, spectrogram generation, transfer learning, spectral and longitudinal model architectures and their corresponding hyper-parameters. We also describe the variety of methods used for comparative purposes.

4.1. Dataset

The ECGs from 17,381 patients (11,853 MI and 5,528 Normal cases) were collected in the Provincial Key Laboratory of Coronary Heart Disease, Guangdong Cardiovascular Institute (GCI), which is located in Guangdong province, China. The study has obtained ethics committee approval and informed patient consent. Each 12-lead ECG waveforms was anonymised, sampled at 500 Hz and was 10 s long. The ECG signals for each patient contain the standard 12 leads, which are *I, II, III, V1, V2, V3, V4, V5,V6, aVF, aVL,* and *aVR*. Cardiologists annotated the MI cases furthermore into three sub-groups: *Acute*, *Recent* and *Old* based on hospital records like patient history, in combination with ECGs. The final GCI dataset resulted $1,489$ Acute (MI occurred within 7 days), $5,377$ Recent (MI occurred in less than 30 days but longer than 7 days) and $4,613$ Old (MI occurred beyond 30 days) MI cases.

4.2. Parameter Setup

We set the duration of a sample, $\mathbf{w}_n$, to 1 s in order to achieve a balance between longer duration (which might have some degree of redundancy and also reduce the number of training samples) and shorter duration (which might lack enough information to do inference). We applied an overlapping percentage of 50%, between subsequent samples, which results a total of $323,133$ samples from each lead, of which: $28,291$ are Acute, $102,163$ are Recent, $87,647$ are Old, and the remaining $105,032$ are Normal samples. In the spectrogram generation step, we applied a Fourier transform, with a chunk of 0.1 s and an overlapping percentage of 90%, which results in a spectrogram representation with 91×26 resolution when stored as an image in JPEG format.

Among existing computer vision networks, we experimented with GoogLeNet's Inception-v3 [31] and MnasNet [39] networks that are trained on natural images from ImageNet [40]. GoogLeNet is known to be quite robust due to its novel inception module and the fact that it has been validated across multiple domains. MnasNet, on the other hand, is considered a mobile computing friendly architecture, designed with limited resource settings in mind. Thus, we extract deep features from the penultimate layers of these networks using the spectrograms as input, resulting $\tau = 2,048$ and $\tau = 1,056$ dimensional deep feature vectors from the Inception-V3 and MnasNet networks, respectively. In order to maximise learning of MI-specific features on the top of the transfer learning from existing networks, we inserted a new fully-connected dense layer with a dimension of $\kappa = 16$ neurons with a ReLu activation, which will be trained entirely on our dataset. We have aimed for an LSTM network which is simple and consists of a single layer of size $\nu = 16$. The number of time steps in the

Table 1: The descriptions of methods employed and compared against each other in this study. Different fusion techniques, such as *Data*, *Feature* and *Decision* were experimented. Data fusion involves stacking of the spectrograms from different leads; feature fusion refers to the aggregation of features extracted from multiple leads. Decision fusion refers to the aggregation of decisions from different ECG leads. Two types of modelling were also employed: *Spectral* and *Longitudinal* that encode spatial and temporal information, respectively.

Method	Description
Data-Dense	Data fusion and spectral modelling using dense layers
Data-LSTM	Data fusion and longitudinal modelling using LSTM
Data-Dense-LSTM	Data fusion and spectral modelling using dense layers and longitudinal modelling using LSTM
Accumulation-Feature-Dense	Accumulation-based feature fusion and spectral modelling using dense layers
Accumulation-Feature-LSTM	Accumulation-based feature fusion and longitudinal modelling using LSTM
Accumulation-Feature-Dense-LSTM	Accumulation-based feature fusion and spectral modelling using dense layers and longitudinal modelling using LSTM
Concatenation-Feature-Dense	Concatenation-based feature fusion and spectral modelling using dense layers
Concatenation-Feature-LSTM	Concatenation-based feature fusion and longitudinal modelling using LSTM
Concatenation-Feature-Dense-LSTM	Concatenation-based feature fusion and spectral modelling using dense layers and longitudinal modelling using LSTM
Vote-Decision-Dense	Spectral modelling using dense layers and majority vote-based decision fusion
Vote-Decision-LSTM	Longitudinal modelling using LSTM and majority vote-based decision fusion
Vote-Decision-Dense-LSTM	Spectral modelling using dense layers, longitudinal modelling using LSTM, and majority vote-based decision fusion
Accumulation-Decision-Dense	Spectral modelling using dense layers and accumulation-based decision fusion
Accumulation-Decision-LSTM	Longitudinal modelling using LSTM and accumulation-based decision fusion
Accumulation-Decision-Dense-LSTM	Spectral modelling using dense layers, longitudinal modelling using LSTM, and accumulation-based decision fusion

LSTM is set to $\gamma = 19$ samples. Finally, the softmax layer consists a fully connected layer of size $N_c = 4$, which is equal to the number of diagnosis classes. During training a *sparse-softmax-cross-entropy* loss function is used with *Adam* optimiser and a learning rate of 0.01.

Both the GGH and PTB datasets used for the validation of the proposed framework exhibit imbalance of samples among classes. Though the number of MI cases is significantly higher than normal cases in GGH, the imbalance becomes smaller in the 4-class classification task OF MI onset detection since the MI cases further decomposed into Acute, Recent and Old classes. Considering the imbalance issue, we adopt the following strategies in our validation: a) we employ patient-level and five-fold stratified cross-validation, where the *patient-level* sampling makes sure samples extracted from one patient's ECG sequence cannot be evaluated against other samples from the same sequence, and the stratification enables fair representation of classes in each fold (a different validation strategy is also experimented and compared with the proposed five-fold stratified cross-validation in the Supplemental Material); b) every batch is designed to contain proportional number of samples from each class; c) sparse-softmax-cross-entropy is applied as our loss function that weighs the loss accordingly; d) Area under receiver operating characteristic (AUROC) is applied as the main metric to compute the classification performance, as this metric better handles the potential class imbalances. AUROC for each class is obtained using a one-vs-all strategy, and overall AUROC is computed from the average of AUROCs of all the four classes across the five cross-validations. Similarly, we also employed accuracy, precision, sensitivity, specificity and F-score to evaluate MI detection performance. The confusion matrices are also provided to obtain more insights on the potential misclassification errors among the classes.

The proposed framework consists of a variety of fusion techniques (data, feature and decision) and modelling approaches (spectral, longitudinal and spectral-longitudinal). These were experimented with for every possible combination of fusion and modelling approaches. A summary of these methods is provided in in Table 1.

5. Results and Discussion

In this section, we present the results obtained for the classification of MI cases from normal cases as well as the prediction of the onset time of MI (i.e. acute, recent and old), which was generally treated as a four-class classification problem. We also performed the comparison of the proposed framework with existing features and classifiers.

Table 2: Performance of the proposed framework validated to detect MI cases from Normal Patients in the GGH dataset. Acc.: Accuracy, Pre.: Precision, Sen: Sensitivity, Spe.: Specificity , F: F_1-score

	Performance metrics (%)					
Methods	Acc.	Pre.	Sen.	Spe.	F	AUROC
Spectral	70.3	86.7	66.8	77.8	75.5	80.6
Longitudinal	73.2	83.9	**75.3**	68.6	79.4	79.8
Spectral-longitudinal	72.5	**89.8**	67.6	**83.3**	77.1	**85.2**

5.1. MI Detection

We first evaluated the performance of our proposed approach to detect MI cases from Normal cases using MnasNet features on the GGH dataset. To this end, Spectral, Longitudinal and Joint Spectral-Longitudinal models were evaluated on the stacked spectrogram (data-fusion) of multiple ECG leads. The results are shown in Table 2, where an AUROC value of 85.2% is achieved using the Joint Spectral-Longitudinal model, higher than the independent performance of the Spectral and Longitudinal models.

5.2. Onset Time Detection of MI

Table 3 shows the AUROC (%) values per each diagnosis class across different fusion approaches: data, feature and decision. Overall, feature fusion was shown to outperform the other fusion approaches as the majority of the classes achieved their highest performance through feature fusion approaches. Using GoogLeNet features (top half of Table 3 the following highest AUROC values are reported for the classes: *Acute* (82.9% using *Concatenation-Feature-Dense-LSTM*), *Normal* (96.9% using *Concatenation-Feature-LSTM*) and *Old* (87.4% using *Concatenation-Feature-Dense-LSTM*). Using MnasNet features (bottom half of Table 3), all the four diagnosis classes achieved their highest AUROC values using feature fusion approaches, i.e., *Acute* (81.5% using *Concatenation-Feature-LSTM*), *Recent* (73.1% using *Concatenation-Feature-LSTM*)), *Normal* (98.2% using *Concatenation-Feature-Dense-LSTM*) and *Old* (75.1% using *Concatenation-Feature-Dense-LSTM*). Results suggest that *Normal* cases are, as expected, easier to detect compared to MI cases. On the other hand, MI cases are shown to be challenging to classify, particularly *Recent* cases as they are prone to being misclassified between the two extremes: *Recent* and *Old*. The results reported with MnasNet features in the bottom part of Table 3 are particularly encouraging, as competitive performance is achieved with GoogLeNet features. The MnasNet framework is relatively simple compared to GoogLeNet, with the lower dimensionality of MnasNet features ($\tau = 1,056$) proving to be an effective representation, particularly as input for the longitudinal models which otherwise were prone to overfitting using the $\tau = 2,048$-dimensional GoogleNet features. This is shown by the fact that the two MI cases achieved the best performance with the *Concatenation-Feature-LSTM* method using MnasNet features.

Generally, the superior performance obtained using feature fusion approaches reflects the unique discriminative characteristics of each ECG lead, which need to be exploited during modelling rather than merging early before modelling using data fusion or lately using decision fusion after modelling. Comparison between two feature fusion techniques: *accumulation* and *concatenation*, reveals both are effective but *concatenation* is superior as the information corresponding to the distinctiveness of each lead is fed into the model while *accumulation* might lose this information. Although it is the simplest and least resource demanding, data fusion is shown to be the worst performing fusion technique, partly due to the loss of discriminative characteristics of different leads too early in the pipeline. Another critical reason for inferior performance of data fusion is the information loss associated with the mandatory input resizing step of the existing frameworks, e.g., Inception-V3 resizes input spectrograms to 299×299.

The overall results shown in Table 3 provide further insights into the effectiveness of the proposed joint spectral-longitudinal modelling approach across the majority of the fusion techniques. The average AUROC performance show that joint spectral-longitudinal modelling achieved the highest in *Feature Accumulation* (77.4%), *Feature Concatenation* (80.5%),

Table 3: Performance of data, feature and decision fusion approaches in the proposed framework for the detection and classification of MI cases using spectral (Dense), longitudinal (LSTM) and spectral-longitudinal (Dense-LSTM) models, using: GoogLeNet (top table)and MnasNet features (bottom table).**A**: Acute, **R**:Recent, **N**:Normal, **O**: Old

GoogLeNet features

	AUROC per class (%)				Global (%)	
Method	**A**	**R**	**N**	**O**	**AUROC**	**Accuracy**
Data-Dense	70.0	62.2	83.8	63.5	69.9	69.9
Data-LSTM	75.7	66.1	92.7	69.2	75.9	73.4
Data-Dense-LSTM	78.1	53.9	85.9	67.9	71.4	70.3
Accumulation-Feature-Dense	76.4	64.4	86.4	64.1	72.8	71.3
Accumulation-Feature-LSTM	73.1	69.5	92.1	69.2	76.0	73.4
Accumulation-Feature-Dense-LSTM	77.8	65.7	93.9	72.2	77.4	75.6
Concatenation-Feature-Dense	80.8	66.6	95.0	**87.4**	77.8	75.0
Concatenation-Feature-LSTM	62.0	64.1	**96.9**	68.8	73.0	68.7
Concatenation-Feature-Dense-LSTM	**82.9**	68.6	96.7	73.8	**80.5**	**77.1**
Accumulation-Decision-Dense	74.0	65.2	85.3	64.0	72.1	71.0
Accumulation-Decision-LSTM	68.8	**70.6**	89.9	65.9	73.8	71.9
Accumulation-Decision-Dense-LSTM	74. 8	67. 2	93. 2	71. 9	76.8	75.8

MnasNet features

	AUROC per class (%)				Global (%)	
Method	**A**	**R**	**N**	**O**	**AUROC**	**Accuracy**
Data-Dense	68.1	59.8	78.9	61.6	67.1	68.9
Data-LSTM	71.3	67.9	90.8	68.4	74.6	73.3
Data-Dense-LSTM	73.7	54.8	77.6	59.8	66.4	67.2
Accumulation-Feature-Dense	73.8	62.6	85.5	64.3	71.5	70.9
Accumulation-Feature-LSTM	72.1	68.0	91.0	68.5	74.9	73.8
Accumulation-Feature-Dense-LSTM	75.5	66.3	92.9	71.8	76.6	73.7
Concatenation-Feature-Dense	78.9	65.0	94.3	69.3	76.9	74.6
Concatenation-Feature-LSTM	**81.5**	**73.1**	97.7	74.7	**81.8**	**78.3**
Concatenation-Feature-Dense-LSTM	77.7	68.1	**98.2**	**75.1**	79.8	77.8
Accumulation-Decision-Dense	71.4	63.8	81.4	63.3	70.0	69.4
Accumulation-Decision-LSTM	70.3	68.6	90.8	67.1	74.2	74.7
Accumulation-Decision-Dense-LSTM	73.1	66.9	91.9	70.2	75.5	72.9

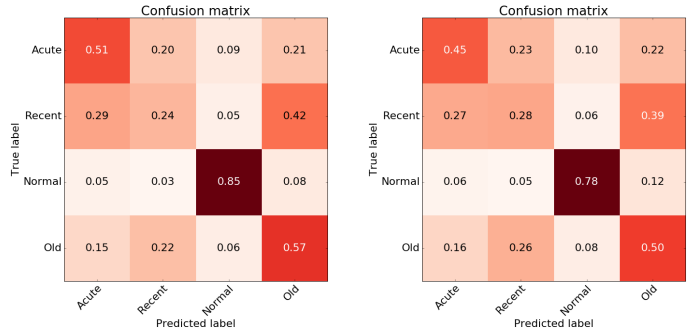
Decision Accumulation (76.8%) using GoogLeNet features. Similarly, superior performance by the spectral-longitudinal model is achieved using *Feature Accumulation* and *Decision Accumulation* using MnasNet features. Separately, longitudinal models achieved the highest in data fusion technique (75.9% using GoogLeNet and 74.6% using MnasNet features), where the feature dimension is not too long to lead to the curse of dimensionality and model overfitting. We also provided the accuracy metric in the last column, and the results showed similar behaviour as the AUROC, i.e., concatenation-based feature fusion superior performance, compared to the other fusion techniques. During the validation using MnasNet features, the LSTM-based longitudinal models tend to perform competitively with Dense-LSTM models, due to the smaller dimensionality of MnasNet features (1056) compared to GoogLeNet (2048), and hence less overfitting.

Among the decision fusion approaches, both majority vote and decision accumulation showed competitive performance. However, the former showed slight superiority over decision accumulation as shown in Fig. 3 (more in the Appendix), for the proposed joint spectral-longitudinal model. This is partly due to the loss of fine details resulting from the averaging performed in accumulation fusion techniques. Furthermore, the confusion matrices reveal misclassification errors observed in the classification of MI cases. It is clearly visible that the Normal class has been distinctively classified without significant misclassification with other MI cases. On the other hand, more misclassification errors occur among MI types, i.e. acute, recent and old. In summary, though data-fusion seems ineffective in leveraging the distinctiveness of the leads, care must still be taken when considering feature concatenation or decision fusion strategies. Namely, constraints imposed by training resource considerations due to the curse of dimensionality that might arise for feature concatenation, or the need of independent modelling for each lead in decision fusion.

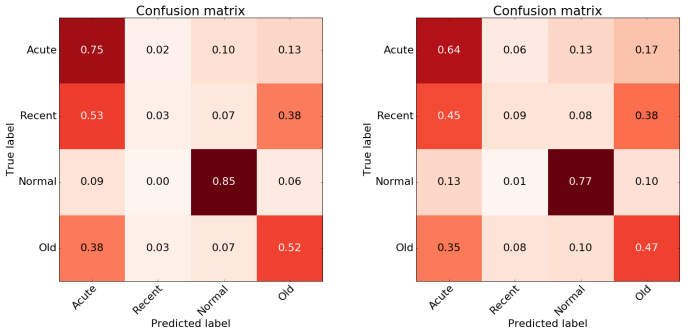
Our DeepMI framework not only demonstrates its feasibility in identifying an infarction from ECG readings alone, that is free from any clinical input, but it also indicates the predicted onset time of an infarction. Previous studies have shown that the occurrence time of an MI affects both the reinfarction rate, as well as mortality in surgery [41, 42]. In particular, the risk of developing an adverse cardiac event has been reported to decrease from 32.8% to 18.7% when surgery occurs within 30 days of MI, compared to surgeries occurring 31-60 days post-MI [42]. Utilising DeepMI could therefore provide the age of an MI from patient ECG readings only, without requiring previous patient histories or more involved laboratory test results. Our proposed framework could therefore be utilised to assist clinicians in remote settings, by providing additional information to assess patient risk.

5.3. Comparison with existing works and validation using a public database

Among existing works, we selected handcrafted features from ECG waveforms validated



(a) GoogLeNet features



(b) MnasNet features

Figure 3: Comparison of decision fusion approaches on (a) GoogLeNet and (b) MnasNet features. The first column contains *Majority vote* results and the second column *Decision accumulation*

with common classifiers and dedicated deep neural networks [43]. The handcrafted features were obtained from autoregressive parameter estimation using Burg's method, which was developed for spectral estimation [44], and it estimates the autoregressive coefficients optimised for the input signal, by minimizing loss in forward and backward prediction errors. Such representation was shown to outperform the fast-Fourier transform method in encoding dynamics of time-series signals, including ECG waveforms [45]. Common classifiers were used to validate on the autoregressive representation, which include Random Forest, Logistic Regression, K-nearest Neighbors, Gradient Boosted Trees and Multi-layer Perceptron. In addition to the handcrafted feature representation, simple deep networks, such as convolutional neural networks (CNNs) and Recurrent Neural Networks (RNNs) were implemented on the raw ECG waveforms. The CNN architecture adopted a four-layer architecture, and subsequent 1-D convolution was applied. The RNN architecture was made of a single-layer LSTM component, with a 1024 hidden units (using LSTM cells). Similar train-test validation was employed as of [43], i.e., 10-fold cross validation, and the AUROC performance of both the existing works and the proposed (Dense, LSTM and Dense-LSTM) framework with the data-fusion approach were shown in Table 4.

To further validate our proposed spectral-longitudinal model on a publicly available dataset, we used PhysioBank Physikalisch-Technische Bundesanstalt (PTB) [37]: The PTB ECG dataset comprises 15-lead ECG records of patients diagnosed with multiple heart diseases, sampled at 1000 Hz [46, 47]. In our evaluation, we only used 12-lead ECG data related to 148 (MI) and 52 (Healthy) subjects, with a total of 200 subjects. As the duration of ECG data may vary across subjects, we used only the first 10-second segment of each patient.

The results showed that handcrafted features performed inferior to our approach while Gradient Boosted Tree achieved higher performance compared to other non-deep learning based classifiers. The deep networks, i.e., CNNs and RNNs, that were trained from scratch with the raw ECG waveforms shown to underperform in their MI detection power, partly due to the limited amount of data available for training. This validates our approach which aims to utilize existing networks via transfer learning rather than designing a new network for each problem and train it extensively with a given dataset.

6. Conclusion

Myocardial infarction (MI) is globally the leading of cause of death among cardiovascular diseases. Due to the delay and resource expenditure associated with diagnosis from laboratory based blood sample tests, it is common clinical practice to inspect the electrocardiogram (ECG) records of patients instead. However, such an approach is still time consuming and subject to interpretation bias. In most developing countries, the number of cardiologists to interpret the ECG

Table 4: AUROC (%), results from the publicly available PTB database [37], using existing works that employed hand-crafted features with traditional classifiers, and raw ECG waveforms with dedicated CNN and RNN [43] frameworks. A comparison is conducted with data-fusion strategy of our proposed framework to detect MI cases from Normal cases.

PTB Dataset			
	Method	Classifier	AUROC
Existing	Handcrafted	Random Forest	68%
		Gradient Boosted Tree	70%
		Multi-layer Perceptron	61%
		Logistic Regression	66%
		K-nearest Neighbor	64%
	Raw ECG time-series	CNN	49%
		LSTM	49%
Proposed	Spectral	Dense	88%
	Longitudinal	LSTM	90%
	Spectral-Longitudinal	**Dense-LSTM**	**94%**

records is far lower than demand requires. In this context, data-driven approaches can assist the diagnosis process by providing decision support to the domain experts on site. Particularly, understanding the onset time of MI is crucially important in providing effective and early intervention, thereby improving patient outcomes. To this end, we have proposed an end-to-end trainable deep learning framework that takes raw ECG records and detects MI cases from normal (or non-MI cases), whilst also inferring the occurrence time of the heart attack as being either acute (within 7 days), recent (less than 30 days but longer than 7 days) and old (beyond 30 days). To do so, we employ a transfer learning technique that aims to exploit existing convolutional neural networks by representing the raw ECG time series with an image-like spectrogram representation.

The advantages offered by the proposed framework include: 1) the ability to detect heart attack (and its occurence time) with relatively simpler framework that employs transfer learning capabilities to avoid designing a dedicated network and train from scratch; 2) thereby reducing the development process and requirement of large training data; 3) moreover, the spectrogram representation of raw ECG waveforms encodes spectral-temporal dynamics that are valuable to detect MI in addition to achieving robustness across device specifications and small variations due to noise, motion artefacts and mounting positions; 4) specifically, the use of spectral-longitudinal model enables to encode short- and long-term dynamics available in multi-lead ECG waveforms 4) unlike other existing frameworks, the proposed framework is advantageous in its fusion strategies that aim to utilise multiple ECG leads; 5) the proposed framework is more impactful in low resource settings with limited data and lack of precise acquisition devices and domain experts; 6) the robustness of the proposed framework is validated via experimentation under different settings (e.g., different existing networks, comparison with multiple baselines and existing frameworks, and data representations (raw time series, ECG). The disadvantages of the proposed approach involves: 1) it is still a supervised learning framework that requires the labelling of the patient sample by the domain experts; 2) particularly, the data-fusion approach in accordance with the proposed framework may suffer from information loss due to the resizing pre-processing step available in the existing frameworks; 3) though transfer learning improves prediction performance with minimal effort and manageable training data requirement, the interpretability of automated features learned could be relatively hard compared to learning features from a dedicated network without transfer learning. Clearly the advantages offered by this work outweighs the limitations and a prediction performance of as high as 81.8% (AUROC) is obtained when we validated the proposed framework on a cohort of $> 17,000$ Chinese patients. Though the onset-time detection performance could sometimes be less optimal, e.g., for *Recent* MI cases, the proposed platform could still provide useful insights, such as perioperative risks, particularly in low-resource settings where there is very minimal cardiologists-to-patients and/or rich patient history is not common/available. Future work aims to expand the validation of the proposed framework to multiple recent networks, such as ResNeSt [48] and to utilize attention-based mechanisms in the longitudinal modelling stage. Furthermore, more layers other than the output layers in existing frameworks could be fine-tuned with the use of our large GGH dataset in order

to further improve the onset-time prediction performance.

Acknowledgements

This project was supported by the EPSRC "FAST" Healthcare NetworkPlus initiative. TZ was supported by the RAEng Engineering for Development Research Fellowship. The research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Declaration of Interest: None

References

[1] W. H. Organization, "Cardiovascular diseases (CVDs)," *URL link: www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), Last accessed on 08 June 2020.*

[2] T. Zhu, A. E. Johnson, J. Behar, and G. D. Clifford, "Crowd-sourced annotation of ecg signals using contextual information," *Annals of biomedical engineering*, vol. 42, no. 4, pp. 871–884, 2014.

[3] R. J. Martis, U. R. Acharya, and H. Adeli, "Current methods in electrocardiogram characterization," *Computers in biology and medicine*, vol. 48, pp. 133–149, 2014.

[4] V. Jahmunah, S. L. Oh, J. K. E. Wei, E. J. Ciaccio, K. Chua, T. R. San, and U. R. Acharya, "Computer-aided diagnosis of congestive heart failure using ecg signals–a review," *Physica Medica*, vol. 62, pp. 95–104, 2019.

[5] H. R. 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.

[6] S. A. Guidera and J. 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.

[7] S. Mehta, N. Lingayat, and S. Sanghvi, "Detection and delineation of P and T waves in 12-lead electrocardiograms," *Expert Systems*, vol. 26, no. 1, pp. 125–143, 2009.

[8] S. Dash, K. Chon, S. Lu, and E. Raeder, "Automatic real time detection of atrial fibrillation," *Annals of biomedical engineering*, vol. 37, no. 9, pp. 1701–1709, 2009.

[9] L. A. Fleisher, K. E. Fleischmann, A. D. Auerbach, S. A. Barnason, J. A. Beckman, B. Bozkurt, V. G. Davila-Roman, M. D. Gerhard-Herman, T. A. Holly, G. C. Kane *et al.*, "2014 acc/aha guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the american college of cardiology/american heart association task force on practice guidelines," *Journal of the American College of Cardiology*, vol. 64, no. 22, pp. e77–e137, 2014.

[10] Z. Dokur, T. Olmez, and E. Yazgan, "Comparison of discrete wavelet and fourier transforms for ecg beat classification," *Electronics Letters*, vol. 35, no. 18, pp. 1502–1504, 1999.

[11] L. Sharma, R. Tripathy, and S. Dandapat, "Multiscale energy and eigenspace approach to detection and localization of myocardial infarction," *IEEE transactions on biomedical engineering*, vol. 62, no. 7, pp. 1827–1837, 2015.

[12] M. Adam, S. L. Oh, V. K. Sudarshan, J. E. Koh, Y. Hagiwara, J. H. Tan, R. San Tan, and U. R. Acharya, "Automated characterization of cardiovascular diseases using relative wavelet nonlinear features extracted from ecg signals," *Computer methods and programs in biomedicine*, vol. 161, pp. 133–143, 2018.

[13] L. D. Sharma and R. K. Sunkaria, "Inferior myocardial infarction detection using stationary wavelet transform and machine learning approach," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 199–206, 2018.

[14] S. S. Swain and D. Patra, "Multiscale energy based suitable wavelet selection for detection of myocardial infarction in ecg," *Healthcare Technology Letters*, vol. 6, no. 1, pp. 1–7, 2018.

[15] S. M. Abubakar, W. Saadeh, and M. A. B. Altaf, "A wearable long-term single-lead ecg processor for early detection of cardiac arrhythmia," in *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2018, pp. 961–966.

[16] M. Kumar, R. Pachori, and U. Acharya, "Automated diagnosis of myocardial infarction ecg signals using sample entropy in flexible analytic wavelet transform framework," *Entropy*, vol. 19, no. 9, p. 488, 2017.

[17] R. K. Tripathy, A. Bhattacharyya, and R. B. Pachori, "Localization of myocardial infarction from multi-lead ecg signals using multiscale analysis and convolutional neural network," *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11 437–11 448, 2019.

[18] ——, "A novel approach for detection of myocardial infarction from ecg signals of multiple electrodes," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4509–4517, 2019.

[19] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram

signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.

[20] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals," *Information Sciences*, vol. 415, pp. 190–198, 2017.

[21] T. Reasat and C. Shahnaz, "Detection of inferior myocardial infarction using shallow convolutional neural networks," in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE, 2017, pp. 718–721.

[22] R. Xiao, Y. Xu, M. M. Pelter, D. W. Mortara, and X. Hu, "A deep learning approach to examine ischemic st changes in ambulatory ecg recordings," *AMIA Summits on Translational Science Proceedings*, vol. 2018, p. 256, 2018.

[23] N. Strodthoff and C. Strodthoff, "Detecting and interpreting myocardial infarction using fully convolutional neural networks," *Physiological measurement*, 2018.

[24] S. Raghunath, A. E. U. Cerna, L. Jing, D. P. vanMaanen, J. Stough, D. N. Hartzel, J. B. Leader, H. L. Kirchner, C. W. Good, A. A. Patel *et al.*, "Deep neural networks can predict mortality from 12-lead electrocardiogram voltage data," *arXiv preprint arXiv:1904.07032*, 2019.

[25] S. Goto, M. Kimura, Y. Katsumata, S. Goto, T. Kamatani, G. Ichihara, S. Ko, J. Sasaki, K. Fukuda, and M. Sano, "Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients," *PloS one*, vol. 14, no. 1, p. e0210103, 2019.

[26] A. Darmawahyuni, S. Nurmaini, W. Caesarendra, V. Bhayyu, M. N. Rachmatullah *et al.*, "Deep learning with a recurrent network structure in the sequence modeling of imbalanced data for ecg-rhythm classifier," *Algorithms*, vol. 12, no. 6, p. 118, 2019.

[27] U. B. Baloglu, M. Talo, O. Yildirim, R. San Tan, and U. R. Acharya, "Classification of myocardial infarction with multi-lead ecg signals and deep cnn," *Pattern Recognition Letters*, vol. 122, pp. 23–30, 2019.

[28] C. Han and L. Shi, "Ml–resnet: A novel network to detect and locate myocardial infarction using 12 leads ecg," *Computer methods and programs in biomedicine*, vol. 185, p. 105138, 2020.

[29] K. Feng, X. Pi, H. Liu, and K. Sun, "Myocardial infarction classification based on convolutional neural network and recurrent neural network," *Applied Sciences*, vol. 9, no. 9, p. 1879, 2019.

[30] R. Xiao, Y. Xu, M. M. Pelter, R. Fidler, F. Badilini, D. W. Mortara, and X. Hu, "Monitoring significant st changes through deep learning." *Journal of electrocardiology*, vol. 51, no. 6S, pp. S78–S82, 2018.

[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[32] G. A. Tadesse, T. Zhu, Y. Liu, Y. Zhou, J. Chen, M. Tian, and D. Clifton, "Cardiovascular disease diagnosis using cross-domain transfer learning," in *Proc. of 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2019, pp. 4262–4265.

[33] G. Abebe Tadesse, H. Javed, N. Le Nguyen Thanh, H. Duong Ha Thi, L. Thwaites, T. Le Van, D. Clifton, and T. Zhu, "Multi-modal diagnosis of infectious diseases in the developing world," *IEEE Journal of Biomedical and Health Informatics*, 2020.

[34] S. M. Bradley, J. A. Borgerding, G. B. Wood, C. Maynard, and S. D. Fihn, "Incidence, risk factors, and outcomes associated with in-hospital acute myocardial infarction," *JAMA network open*, vol. 2, no. 1, pp. e187 348–e187 348, 2019.

[35] W. Liu, M. Zhang, Y. Zhang, Y. Liao, Q. Huang, S. Chang, H. Wang, and J. He, "Real-time multilead convolutional neural network for myocardial infarction detection," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1434–1444, 2017.

[36] G. Chen, Z. Hong, Y. Guo, and C. Pang, "A cascaded classifier for multi-lead ecg based on feature fusion," *Computer methods and programs in biomedicine*, vol. 178, pp. 135–143, 2019.

[37] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[38] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *Proc. of The IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2016, pp. 71–76.

[39] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "Mnasnet: Platform-aware neural architecture search for mobile," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2820–2828.

[40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. 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.

[41] M. Livhits, C. Y. Ko, and M. J. Leonardi, "Risk of surgery following recent myocardial infarction," *Journal of Vascular Surgery*, vol. 54, no. 3, p. 912, 2011.

[42] P. K. Garg, "Preoperative cardiovascular evaluation in patients undergoing vascular surgery," *Cardiology clinics*, vol. 33, no. 1, pp. 139–150, 2015.

[43] S. Rane, "A sparse, data-efficient ecg representation is predictive of myocardial infarction without expert knowledge," in *Neural Information Processing Systems (NEURIPS) Workshop on AI for Social Good (AISG)*, 2019.

[44] R. Bos, S. De Waele, and P. M. Broersen, "Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 6, pp. 1289–1294, 2002.

[45] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains," *ISRN*, vol. 2014, 2014.

[46] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[47] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet," *Biomedizinische Technik/Biomedical Engineering*, vol. 40, no. s1, pp. 317–318, 1995.

[48] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, Z. Zhang, H. Lin, Y. Sun, T. He, J. Mueller, R. Manmatha *et al.*, "Resnest: Split-attention networks," *arXiv preprint arXiv:2004.08955*, 2020.