

# Signals Are All You Need: A Deep Learning Architecture for ECG Lead Identification

Nathan Macdonald

*Department of Advanced Computing Sciences*

*Faculty of Science and Engineering*

*Maastricht University*

*Maastricht, The Netherlands*

**Abstract**—Although electrocardiographic imaging (ECGI) provides detailed cardiac activation maps, its clinical accessibility remains limited—one key barrier being the accurate localisation of electrodes on the torso to construct a body surface potential map (BSPM). As a step towards eliminating this overhead, this study develops a deep learning model for automated 12-lead ECG lead identification from signal data alone. A permutation equivariant architecture is introduced that couples a branched convolutional network with a GRU encoder and utilises a series of Set Attention Blocks (SABs) to integrate inter-lead information.

Trained using the PTB-XL database, the model achieved an overall lead-level accuracy of 98.5% [95% CI 98.2–98.7], and demonstrated moderate robustness across cardiac conditions. Principal component analysis (PCA) showed that the model internalises the precordial V1-V6 arc as well as the horizontal and vertical electrical planes. Gradient-based attribution analysis revealed the Q and R waves to be the key morphological discriminators. Recall for lead aVF was the highest (99.8%), and lowest for lead V3 (96.7%). In general, recall for limb/augmented leads (99.2%) was higher than for precordial leads (97.5%).

To the authors' knowledge, this research presents the first model for signal-based lead identification in the 12-lead ECG, paving the way for continuous electrode localisation and truly imageless ECGI.

**Index Terms**—Electrocardiography, electrocardiographic imaging, deep learning, lead identification, electrode localisation

## I. INTRODUCTION

Electrocardiography faces a dichotomy between spatial resolution and clinical accessibility [1], [2]. On the one hand, the electrocardiogram (ECG) enables rapid, non-invasive assessments of cardiac function [3]. On the other hand, electrocardiographic imaging (ECGI) yields detailed cardiac potential maps at the cost of substantial clinical overheads in the form of patient-specific data acquisition [4]–[6], which limits its widespread clinical adoption. To make ECGI more convenient for clinicians and patients alike, recent progress towards "imageless" ECGI removes the need for patient-specific heart models [7]–[9]. However, this modality still requires knowing where on the torso each waveform was recorded. This study thereby develops and evaluates an automated lead-labelling model for the standard 12-lead ECG, thereby laying the groundwork for more general signal-based electrode localisation.

This thesis was prepared in partial fulfilment of the requirements for the Degree of Bachelor of Science in Data Science and Artificial Intelligence, Maastricht University. Supervisors: Joël Karel, Ralf Peeters, Pietro Bonizzi.

## A. Background

The introduction of the ECG by physiologist Willem Einthoven in 1902 [10] revolutionised the study and diagnosis of heart disease [11]. Electrodes are placed in specific locations on a patient's body to measure body surface potentials (BSPs), which are subsequently interpreted as non-invasive indicators of cardiac electrical activity [3]. The number and position of these electrodes vary between ECG setups; the most popular being the standard 12-lead ECG, which employs three bipolar limb leads (I-III), three augmented limb leads (aVR, aVL, aVF), and six precordial leads (V1-V6). This multi-lead arrangement enhances diagnostic power by providing multiple spatial perspectives of cardiac electrical activity. Standard ECG, however, lacks fine spatial resolution: each lead reflects a weighted average of potentials generated by the entire myocardium, making local information difficult to obtain [12]. As a result, the modality struggles with tasks that require high spatial fidelity—for example, its sensitivity for left ventricular hypertrophy is very low (6.9%) [13], [14], and its performance on patients with structural heart disease is poor [2], [15], [16].

To overcome these shortcomings, body surface potential mapping (BSPM) utilises up to hundreds of electrodes to generate a dense potential map of the whole torso [6]. Increasing the coverage of sampled BSPs reduces the impact of the averaging effect, allowing post-processing algorithms to disentangle overlapping source contributions and to reconstruct a far more detailed view of the heart's electrical activity. Building on this approach, recent advancements in the inverse problem of electrocardiography [17], [18] have led to the development of ECGI. This non-invasive, high-resolution modality uses BSPMs to reconstruct detailed maps of the surface potentials on the epicardium [4]. ECGI has demonstrated success in situations where 12-lead ECG falls short, such as localising outflow tract ventricular arrhythmia [19] and mapping atrial fibrillation and tachycardias [20]. Nevertheless, ECGI remains mostly limited to laboratory use since it introduces significant clinical overheads: a patient-specific torso-heart geometry needs to be acquired, usually through computed tomography (CT) or magnetic resonance imaging (MRI) [1], [4]. In addition, generating the BSPM requires knowledge of each electrode's location, which is currently obtained via the same CT or MRI procedure [21], or a computer vision system [22]–[24]. There

remains a need for localisation methods that do not use any patient anatomical data.

### B. This Study

Recent work in imageless ECGI shows that cardiac activation maps can be obtained without patient-specific heart models, thereby removing the need for CT or MRI [9]; however, there remains a need for a method of electrode localisation uncoupled from an explicit representation of the patient's torso geometry. This study aims to explore the feasibility of signal-based electrode localisation by designing and validating a deep learning architecture suited for the sub-problem of 12-lead ECG lead identification. Specifically, each lead's raw waveform is passed separately through a combined Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) encoder. The resulting features are then combined using a set-attention mechanism to capture inter-lead relationships and global context. This is a novel area of research, aiming to ultimately allow ECGI to be performed with a level of convenience comparable to that of traditional ECGs.

The 12-lead ECG setting is attractive since large public datasets, such as PTB-XL [25], provide thousands of annotated ECGs, enabling deep learning models to be trained and evaluated effectively on a wide range of patient demographics and heart conditions. In addition, misplaced or swapped electrodes are common sources of diagnostic errors in routine ECG acquisition [26], [27]. Therefore, an electrode localisation model would be a practical quality control tool. In this 12-lead setting, assigning each waveform to its correct lead effectively stands in for electrode localisation — especially for precordial leads, where the lead labels serve as proxies to their physical locations.

In summary, there exists a need for an ECGI workflow with minimal clinical overheads. Current research trends show that imageless ECGI is a practical enhancement to the modality. In light of these developments, this paper aims to contribute to signal-based electrode localisation research by proposing and validating a deep learning architecture for 12-lead ECG lead identification.

### C. Research Questions

- 1 Can deep learning successfully learn spatial relationships** between ECG leads from their signal data alone?
- 2 To what extent is inter-lead information necessary** when localising electrodes?
- 3 How robust is a lead labelling model** when presented with abnormal heart rhythms?
- 4 Which morphological features** of an ECG signal are most relevant when distinguishing between leads?
- 5 How do the 12 standard ECG leads rank** in identifiability under the proposed model?

## II. RELEVANT WORK

Since there is little to no published research on ECG electrode localisation using signal data alone, this section

provides a brief overview of current deep learning approaches in ECG.

This study focuses on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), as these have emerged as the dominant architectures in the ECG deep learning field [28], [29]. Auto-Encoders (AEs) also appear somewhat frequently in the literature [30], [31]. In principle, AEs could be used for lead labelling; however, they introduce many design complexities (reconstruction loss, unique training process, etc.) which are beyond the scope of this paper. Therefore, this paper focuses on CNNs and RNNs.

### A. Beat Classification

Most of the deep learning work applied to ECG revolves around classifying beat types. While this classification task differs from the one in this study, the techniques used for learning latent feature representations of ECG signals inform the design of the deep learning architecture proposed here.

1D CNNs are well-suited for learning local features of time series data by applying sliding filters of various sizes to extract features at different levels of granularity [32]. In 2019, a 33-layer residual 1D CNN achieved cardiologist-level AUC/F1 on single-lead ECG [33].

Alternatively, the sequential ECG data can be transformed into a 2D representation using various techniques, such as the Short-time Fourier transform [34]. These methods will not be covered in this study, as they add complexity without clear benefits, and in 2024, *Narotamo et al.* showed that a standalone Gated Recurrent Unit (GRU) outperforms such 2D methods on beat classification tasks [29].

GRU networks are a type of RNN [35], designed as a modification to the popular Long Short-Term Memory (LSTM) networks [36] with fewer trainable parameters. The architecture's impressive performance on beat classification [29], [37] and its relatively lightweight design make it appropriate for this study.

Combined CNN + RNN hybrids have become a popular architectural composition [38], aiming to combine the local feature extraction capabilities of CNNs with the ability of modern RNNs to handle long time dependencies in sequential data. These hybrids achieve > 98% sensitivity/specificity on atrial fibrillation and multi-beat tasks [37], [39]. In 2019, *Guo et al.* [37] achieved state-of-the-art accuracy on arrhythmia classification by coupling GRU with DenseNet [40], a fully-connected CNN variant that allows low- and high-level features to flow unaltered to the GRU network. The use of DenseNet inspires this paper's adoption of an Inception-V1 style CNN [41], which also preserves multi-scale temporal features while maintaining a lighter-weight design.

### B. Signal Based Detection of Electrode Misplacement

Recent work on electrode misplacement detection has shifted from traditional machine learning algorithms (such as decision trees) to deep learning models [42]. In 2021, *Rjoob et al.* showed a bidirectional LSTM and a 1D CNN outcompeting

machine learning models and far surpassing cardiologists when spotting one or two intercostal V1/V2 shifts [43].

In 2024, *Huang et al.* developed a precordial lead swap detection model that uses a 3-layer DenseNet CNN [44]. What makes this paper unique is how inter-lead information is shared within the model. At each layer, the feature vectors of each lead are combined with the correlation coefficients between those same feature vectors. The worst performance was seen in V2-V3 reversals and V5-V6 reversals, potentially due to the similarity in waveforms and amplitudes of these adjacent leads.

Crucially, many of the models cited above operate on a single lead and thereby cannot exploit inter-lead relationships. The models that do utilise multiple leads [29], [31], [44] typically do this through convolutional channels or RNN hidden states, which depend on a prescribed lead order. Such architectures are not permutation equivariant; reordering the leads has an unpredictable effect on the output. A review of the literature reveals no previously published ECG model that combines multi-lead input with permutation equivariant processing. The architecture introduced in this paper is novel in this regard.

### III. METHODOLOGY

#### A. Data

This study uses the PhysioNet PTB-XL database [25], which consists of 21 799 clinical 12-lead ECGs from 18 885 patients. The dataset is organised into five overlapping *diagnostic superclass* subsets, shown in Fig. 1. Each ECG is 10 s long and is available in both 100 Hz and 500 Hz sampling frequencies. 100 Hz samples are used to accelerate the forward pass of the model, while remaining above a suggested lower limit of 90 Hz [45]. To remove baseline drift, a fourth-order high-pass Butterworth filter with a cut-off frequency of 0.5 Hz is applied to each signal. The database is stratified into 10 folds. Per the suggestions attached to the database, folds one through eight are used for training, and the last two folds are for validation and testing, respectively.

To mitigate overfitting, the training data is jittered with Gaussian noise, which is equivalent to applying L2 (Tikhonov) regularisation to the model weights [46]. This improves generalisation by smoothing the learned decision boundaries. For each recording, a single standard deviation value  $\sigma$  is sampled from the interval (0.001, 0.015) to allow some samples to be relatively unaffected. Zero-mean Gaussian noise  $N(0, \sigma)$  is then added using this standard deviation.

#### B. Problem Definition

Let  $\mathcal{L} = \{L_1, L_2, \dots, L_M\}$  be the ordered set of lead identifiers (labels), where  $M$  is the number of leads under the given ECG recording protocol. Let  $x_i \in \mathbb{R}^T$  be a single recording of length  $T$  from an unknown lead. The set  $X = \{x_1, x_2, \dots, x_K\}$  is then the *unordered* set of  $K \leq M$  simultaneously recorded ECG signals.

The naive approach to lead labelling would be to construct the following classifier:

$$f_\theta : \mathbb{R}^T \rightarrow \mathcal{L} \quad (1)$$

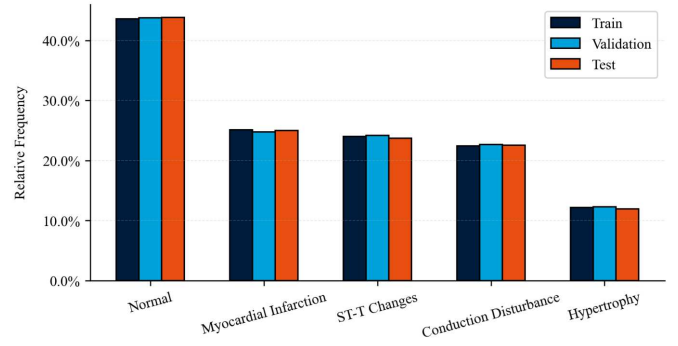


Fig. 1. Prevalence of diagnostic superclasses in the PTB-XL dataset, shown for each of the dataset splits. Diagnostic superclasses are not mutually exclusive.

parameterised by  $\theta$  such that

$$f_\theta(x_i) = L \implies "x_i \text{ originated from lead } L"$$

A more informed approach is to have the classifier predict all leads at once, forming an *injective* mapping  $\pi : X \rightarrow \mathcal{L}$ , such that  $\pi(x_i) \neq \pi(x_j)$  ( $i \neq j$ ).

Alternatively, if  $K = M$ , then the problem can be formulated as a permutation inference problem, in which an ordering of  $X$  needs to be found such that  $X_i$  originates from lead  $L_i$ . The permutation  $\pi^*$  which orders  $X$  in such a way is then the *ground-truth* permutation. A model which outputs a probability distribution over  $\mathcal{L}$  for each  $x_i$ , providing that the *arg max* of these distributions is unique across the lead set, can then be used to predict  $\pi^*$  as follows:

$$\hat{\pi} = \arg \max_{\pi} \sum_{i=1}^K \log p_\theta(L_{\pi(i)} | x_i) \quad (2)$$

where  $p(L_j | x_i)$  is the model's posterior probability that waveform  $x_i$  originated from lead  $L_j$ . For this problem, a classifier should utilise information from all signals in  $X$  to make its decisions.

To ensure injectivity, a matching algorithm, such as the Hungarian method [47], can be used that assigns unique lead labels to each signal such that the likelihood of the resulting assignment is maximised.

The above problem definition can be used to define two accuracy metrics over a dataset of size  $N$ : **lead accuracy**,

$$\text{LeadAcc} = \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K \delta_{i,k}, \quad (3)$$

$$\text{with } \delta_{i,k} = \begin{cases} 1 & \text{if } \hat{\pi}^{(i)}(k) = \pi^{*(i)}(k) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

which considers accuracy on a lead-by-lead basis, and **set accuracy**,

$$\text{SetAcc} = \frac{1}{N} \sum_{i=1}^N \min_{1 \leq k \leq K} \delta_{i,k} \quad (5)$$

which reuses  $\delta_{i,k}$  from (4) to define the proportion of times all  $K$  leads were identified correctly.

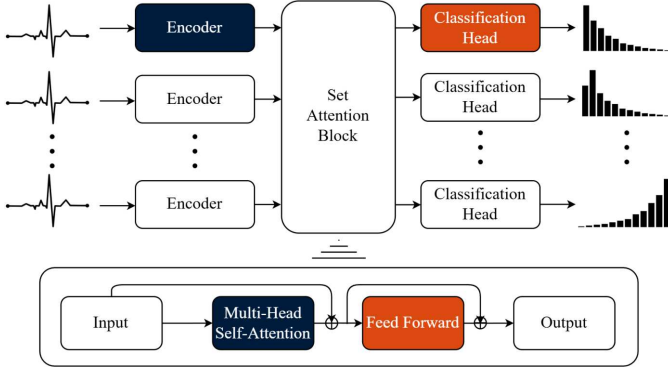


Fig. 2. High-level overview of the proposed architecture. The ECG signals are encoded to produce individual feature vectors, which pass through a series of Set Attention Blocks (only one shown above), allowing inter-lead information to be shared across the feature representations.

### C. Proposed Architecture

This paper introduces a deep learning architecture specifically designed to solve the lead identification problem.

At a high level, the model begins by processing each lead individually with an ECG signal  $\rightarrow$  feature space encoder. These feature representations are then enriched with inter-lead information through a sequence of Set Attention Blocks (SABs) [48]. The final representations are classified individually by a feed-forward, softmax head.

The task of the encoder is to project the length  $T$  ECG signals into feature representations of size  $D < T$ , which contain enough information about each signal for accurate lead identification. Feeding the feature vectors into a classification head at this stage satisfies the classifier described in (1), but would result in extremely informative inter-lead information being left on the table.

This is where the SAB comes in; each signal is projected into query, key, and value spaces, and multi-head self-attention is performed over the entire lead set. This allows signals to integrate information from all others, yielding a contextually-enriched set of feature vectors. Because the SAB is permutation equivariant and agnostic to the input size, the model can process any number of leads in any order. The model presented here uses five SABs stacked sequentially, a depth chosen to balance representational capacity with computational efficiency.

To answer research questions 1 and 3, the accuracy of the model on the whole test set, as well as on the diagnostic superclass subsets, is calculated. In addition, principal component analysis (PCA) is performed on the learnt feature vectors of each signal in order to determine if the spatial relationships between electrodes are reflected in the representations.

#### Encoding Individual Leads

This paper implements an Inception V1-style CNN [41] ("branched" CNN) routed to a GRU block that uses an attention mechanism (GRUAttn). The branched CNN applies 1x1, 5x1, and 11x1 convolutions in parallel alongside a max-pool branch to extract features at different temporal scales. In order to reduce the number of learnable parameters, a 1x1 convolutional

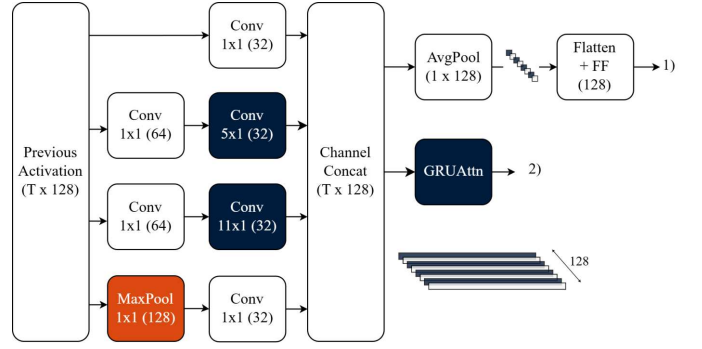


Fig. 3. The branched CNN encoder, using the *Inception V1* design [41]. The output of each branch is concatenated along the channel dimension, which can then be either 1) pooled along the time dimension and flattened to produce a 1D feature vector, or 2) sent through a GRUAttn encoder.

bottleneck is inserted before the 5x1 and 11x1 kernels, which halves the number of channels they receive. For sampling frequencies other than 100 Hz, the kernel sizes should be rescaled to have a similar receptive field.

After each inception block, the outputs of the branches are concatenated along the channel dimension, and a max-pooling layer compresses the signal to half of its length, speeding up the forward pass through the GRUAttn layers. The GRU treats the channels as features of the signal at each time point, capturing long-range dependencies between these features. The attention mechanism projects each GRU hidden state into an "importance" score, soft-maxes those scores to get attention weights, and forms the final feature representation as the weighted sum of all the hidden states.

The number of layers, kernel sizes, and other hyperparameters (including a feature dimension of  $D = 128$ ) were chosen through a combination of manual tuning and grid search. The model implemented in this study uses two branched CNN blocks and two GRUAttn layers. A detailed description of the model can be found in Appendix F.

#### Inter Lead Relevance

Set attention is this study's primary mechanism for incorporating global context into each signal's feature representation. Beyond enabling inter-lead relationships to inform the classification head, the SAB also encourages the network to assign unique lead labels to each signal. This behaviour arises because the SAB couples the latent representations  $z_i$  of each signal  $i$  with the representations of all  $K$  signals, so that an appropriate loss function averaged over the set propagates conflicting gradients across all elements whenever duplicate predictions occur (predicting V1 twice, for instance). This is an important property of the model, since (2) requires that the inferred permutation  $\hat{\pi}$  be injective. While one could enforce injectivity through a post-hoc matching algorithm, set attention generally alleviates the need for such post-processing.

A DeepSets [49] architecture is also considered for its simplistic way of capturing set-level information. DeepSets generates global features across all leads through a permutation-

invariant pooling operation (e.g., sum pooling), which has the benefit of being very efficient, but is limited in its expressiveness for inter-lead relationships. Nevertheless, this paper includes results from a DeepSets architecture for comparison purposes. DeepSets is integrated into the proposed architecture by concatenating the output of the DeepSets block with the feature representations of each signal.

To further assess inter-lead relevance, the effect of training on different lead subsets is examined. Since each precordial trace corresponds to a unique chest electrode—whereas limb and augmented leads are mathematically derived from multiple [3]—precordial-lead labelling more directly supports the downstream goal of electrode localisation. Accordingly, a "precordial model" is trained exclusively on the six precordial leads, and its accuracy is reported as "precordial lead accuracy".

When training the precordial model, several limb leads are included alongside the precordial leads. This is to determine if additional limb and augmented leads provide useful context. In practice, they may introduce more redundancy and noise than useful information, since the augmented leads are linear combinations of the limb leads, and the limb leads themselves satisfy Einthoven's law ( $II = I + III$ ) [3]. Several lead subsets are trained on, while the training loss is optimised exclusively on the precordial leads.

#### D. Morphology

For an input signal  $x \in \mathbb{R}^{K \times T}$  ( $K$  input leads,  $T$  samples), the model outputs a logit matrix  $f(x) \in \mathbb{R}^{K \times M}$ . The goal is to explain the single logit  $f_{k, \hat{\pi}(k)}$  corresponding to the predicted label  $\hat{\pi}(k)$  of  $k$ . Let  $l = (k, \hat{\pi}(k))$  be the index into this logit matrix; by assigning a SHapley Additive exPlanation (SHAP) [50]  $\phi_{i,t}^l$  to time index  $t$  of input signal  $i$ , the contribution of this particular input to the value of the target logit can be analysed. In other words, these SHAP values indicate how relevant each input signal is, over time, towards the predicted label of a specific lead. This paper uses the GradientShap algorithm [50] to estimate the SHAPs, details of which can be found in appendix C.

To assess the contributions of different morphological features, global attributions are calculated for each sample in a batch  $B$ . Each attribution map is heartbeat-aligned, stacked and averaged to give

$$\bar{\phi}_{i,t}^l = \frac{1}{|B|} \sum_{x \in B} \phi_{i,t}^l(x) \quad .$$

Overlaying  $\bar{\phi}_{i,t}^{(l)}$  on the mean cardiac cycle reveals the specific morphological features that drive lead discrimination, answering Research Question 4.

Additionally, these attributions should indicate inter-lead relevance. For example, under the lead ordering  $\mathcal{L} = \{V1, \dots, V6\}$  and assuming the input signals are ordered in the same way,  $\phi_{1,t}^{(1,1)}$  captures the attributions of  $V1$ 's trace when identifying  $V1$ , while  $\phi_{2,t}^{(1,1)}$  indicates the contribution of  $V2$ 's signal to that same  $V1$  prediction.

#### E. Lead Recall

To answer Research Question 5, the model is trained on the full 12-lead set, and the proportion of times each lead  $k$  is identified correctly is calculated. Borrowing  $\delta_{i,j}$  from (4), the following formula for recall over  $N$  ECGs can be used:

$$\text{Recall}_k = \frac{1}{N} \sum_{i=1}^N \delta_{i,k} \quad . \quad (6)$$

Micro-averaging the recalls over all  $k$  is equivalent to lead accuracy, as defined in (3).

#### F. Training

Given a matrix of model predictions

$$P = [p_\theta(\mathcal{L}_j | x_i)]_{1 \leq i \leq K, 1 \leq j \leq M} \quad (7)$$

and a matrix  $P^* \in \{0, 1\}^{K \times M}$  containing one-hot encoded ground-truth labels for each signal  $i$ , the cross-entropy loss is calculated as follows,

$$\text{Loss} = -\frac{1}{K} \sum_{i=1}^K \sum_{j=1}^M P_{ij}^* \log(P_{ij}) \quad . \quad (8)$$

The Adam optimiser [51] minimises this with an initial learning rate of  $\eta = 7 \times 10^{-3}$ , which decays according to a cosine annealing scheduler (see appendix B).

The models were trained in mini-batches of size 38, and training parameters were determined through a combination of manual tuning, grid search, and Bayesian optimisation. After each run, the model checkpoint with the highest validation accuracy was used as the representative model. The models were finally validated on the PTB-XL test set.

#### G. Statistical Analysis

The per-lead recalls defined in (6) are reported alongside 95% Wilson confidence intervals (CIs) [52]. The Wilson interval was chosen instead of alternatives like the Wald interval because it does not suffer from the same overshooting and zero-width interval issues [52]. The independence assumption of the underlying binomial model holds for metrics that operate at the ECG level or are not aggregated across leads. As such, Wilson intervals are used for per-lead recall and set accuracy, whereas CIs for lead accuracy are generated using bootstrapping ( $B = 10\,000$ ), where resampling occurs at an ECG-level. Details can be found in Appendix D.

Additionally, McNemar's test [53] is applied to the pairwise predictions between the precordial model and those trained with additional leads. The test evaluates the null hypothesis that their error rate is identical; only the samples on which the models disagree enter the test statistic. A  $p$ -value below 0.05 is taken as evidence of a statistically significant difference in error rate, implying that one model performs better than the other.



Lead	Recall (%)					
	All [95% CI]	NORM	MI	CD	STTC	HYP
aVF	99.8 [99.5–99.9]	100.0	99.5	99.2	99.8	99.6
aVR	99.6 [99.2–99.8]	99.9	99.3	99.2	100.0	100.0
I	99.6 [99.2–99.8]	99.9	99.5	99.2	99.6	99.2
aVL	99.4 [98.9–99.6]	99.9	98.7	98.5	99.6	100.0
III	99.3 [98.9–99.6]	99.8	98.4	98.8	99.6	98.5
V1	99.2 [98.7–99.5]	99.9	98.6	99.2	99.2	98.9
II	99.0 [98.5–99.4]	99.8	97.8	97.8	99.2	99.6
V6	98.0 [97.2–98.4]	99.1	96.0	96.4	96.9	96.6
V4	97.1 [96.4–97.8]	99.1	96.0	95.2	96.9	96.6
V5	97.1 [96.4–97.8]	99.0	95.3	95.6	97.1	96.0
V2	96.9 [96.1–97.6]	99.2	94.4	94.4	96.0	96.2
V3	96.7 [95.8–97.4]	99.0	95.1	94.2	96.0	95.8
Overall	98.5 [98.2–98.7]*	99.5	97.3	97.3	98.3	98.2

Table I. Per-lead recall with 95 % Wilson confidence intervals. (ALL = full dataset, NORM = normal ECG, MI = myocardial infarction, CD = conduction disturbance, STTC = ST/T-change, HYP = hypertrophy). \*Overall CI calculated using bootstrapping.

Lead set	Precordial lead acc. (%) [95 % CI]	Precordial set acc. (%) [95 % CI]	$\Delta$ / p
$\mathcal{L}_{pre} = \{V1, \dots, V6\}$	97.7 [97.5–98.2]	93.6 [92.5–94.6]	-
$\mathcal{L}_{pre} \cup \{II\}$	97.8 [97.4–98.2]	93.8 [92.7–94.7]	+0.1 / 0.36
$\mathcal{L}_{pre} \cup \{I, II, III\}$	97.9 [97.5–98.2]	93.6 [92.5–94.6]	+0.2 / 0.24
$\mathcal{L}_{pre} \cup \{aVF\}$	97.7 [97.3–98.1]	93.4 [92.3–94.4]	0.0 / 0.95
All 12 leads	97.6 [97.2–98.0]	92.4 [91.3–93.5]	-0.1 / 0.36

Table II. Precordial lead and precordial set accuracies when training on different lead sets. Values are given with 95 % Wilson score confidence intervals. The right-most column reports the lead accuracy difference with respect to the baseline set  $\mathcal{L}_{pre}$ , together with the corresponding McNemar p-values.

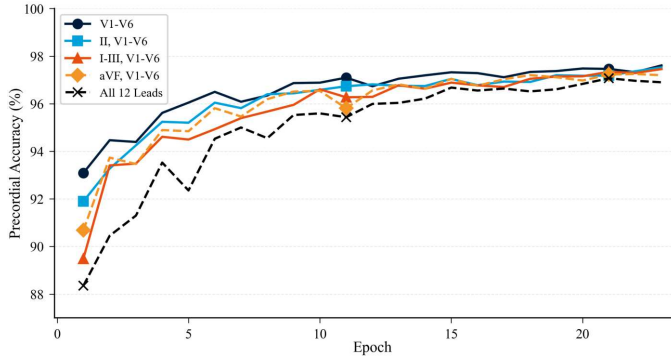


Fig. 4. Precordial validation loss over the training runs (seed 42) of models trained on different lead sets.

## IV. RESULTS

### A. Architecture & Per-Lead Recall

The proposed 12-lead model achieved a lead accuracy of 98.5% [95% CI 98.2–98.7] and a set accuracy of 91.7% [90.4–92.8]. The model exhibited excellent performance on normal sinus-rhythm samples (99.5% [99.4–99.7] lead acc., 96.6% [95.2–97.5] set acc.), and showed the worst performance on samples with conduction disturbances (97.3% [96.7–98.0] lead

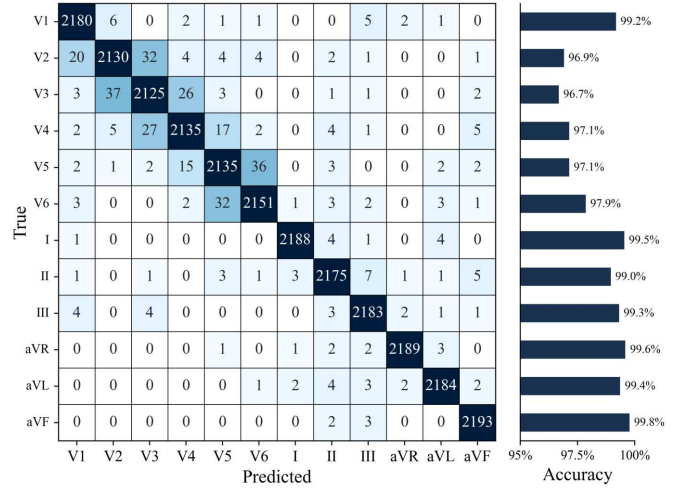


Fig. 5. A confusion matrix of the top-performing model on all 12 leads of the PTB-XL test set, with per-lead accuracies shown on the right.

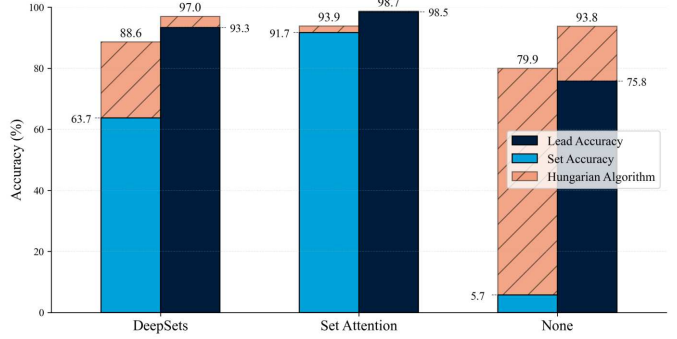


Fig. 6. Set accuracy and lead accuracy between the proposed model using different methods of encoding inter-lead information, with and without the Hungarian algorithm being applied post-hoc. Accuracy for a model without any inter-lead information mechanism is also included.

acc., 85.5% [82.1–88.3] set acc.) and myocardial infarction (97.3% [96.7–98.0] lead acc., 85.3% [82.1–88.0] set acc.).

Fig. 5 shows a confusion matrix of the model trained on all 12 leads, while Fig. 7 provides confusion matrices for the precordial model for each diagnostic superclass.

The per-lead recalls are shown in Table I. Limb and augmented leads had the highest micro-averaged recall of 99.4% [99.2–99.6], compared to precordial leads with 97.5% [97.1–97.9]. Out of the precordial leads, recall for V1 was the highest (99.2% [98.7–99.5]), followed by V6 (98.0% [97.2–98.4]), while lead aVF had the highest recall overall (99.8% [99.5–99.9]). The lowest recall was seen for leads V2 and V3, with 96.9% [96.1–97.6] and 96.7% [95.8–97.4], respectively.

### B. Context Encoding

DeepSets performed worse than set attention, yielding a set accuracy of 63.7% [95% CI 61.7–65.7] and a lead accuracy of 93.3% [92.8–93.8]. As a benchmark, a model with no context encoder (instead, a feed-forward network with three hidden layers) achieved 5.7% [4.8–6.8] and 75.8% [75.1–76.4] set and lead accuracy, respectively. Using the Hungarian algorithm, the

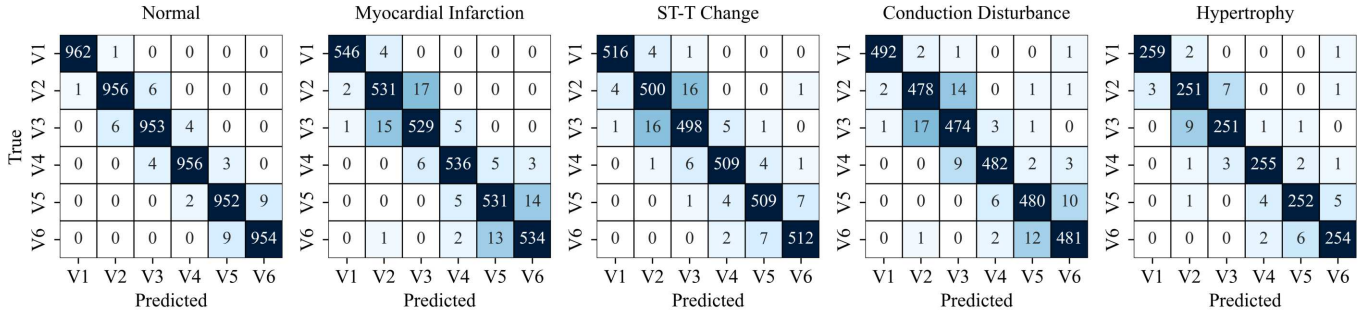


Fig. 7. Confusion matrices of the precordial model on different PTB-XL diagnostic superclasses.

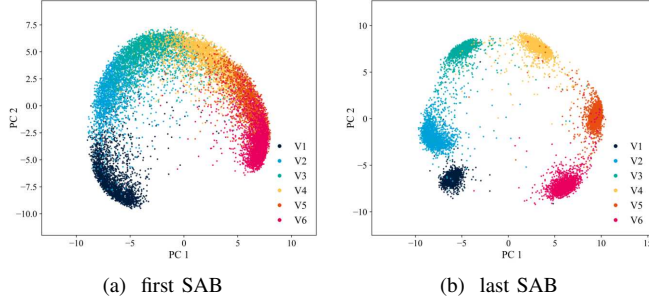


Fig. 8. Feature representations after the first and last Set Attention Blocks, visualised using PCA. Each dot represents the feature vector of a signal projected onto the first two principal components, coloured according to the lead the signal originated from.

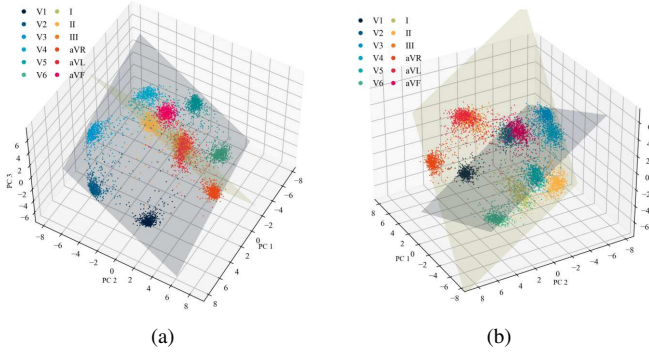


Fig. 9. 3D PCA plots of all 12 leads. Planes of best fit are visualised for the precordial and limb/augmented leads, respectively.

SAB model achieved slightly higher lead and set accuracy; 93.9% [92.8–94.8] and 98.7% [98.4–98.0]. Likewise, the results for DeepSets improved to 88.6% [87.2–89.9] and 97.0% [96.6–97.4], and the model with no context encoder exhibited a drastic accuracy increase to 79.9% [78.2–81.6] and 93.8% [93.1–94.4]. The results are shown in Fig. 6.

### C. Lead Sets

Table II reports the precordial lead and set accuracies when training on different lead sets. The results of the McNemar tests indicate that there is no significant difference in precordial lead accuracy when adding additional leads to the precordial

lead set. The validation loss after each epoch of the model corresponding to each lead set is shown in Fig. 4.

### D. Representations

Fig. 8 projects the precordial feature representations produced by the first and last SAB, respectively, onto the first two principal components (66% explained variance). Each point is one input signal, colour-coded by its ground truth lead. The representations after the first SAB form a smooth, arched manifold along which the leads form a spectrum ranging from V1 (left) to V6 (right), consistent with the actual order of lead placements on the chest. The final SAB yields representations which form much tighter clusters.

Including the third principal component as well as the limb and augmented leads reveals two distinct flat subspaces, along which the precordial and non-precordial leads are respectively arranged. Fig. 9 presents two views of this 3D embedding; the left panel (9a) emphasises the precordial plane, while the right (9b) highlights the limb/augmented plane.

### E. Morphology

The peaks of the R and S waves dominate the attribution maps in Fig. 10, whereas the Q, P, and T waves exhibit minimal attribution.

Evidence of inter-lead relevance is apparent. The signal from the lead corresponding with the labelling target is given the most attribution, followed by the directly adjacent leads. V6 appears to be the most important lead, with its R peak appearing relatively strongly in attributions for all labelling targets. On the other hand, V1 shows little relevance outside of identifying V2 (and V1 itself).

## V. DISCUSSION

### A. General Performance & Inter-Lead Information

The proposed architecture achieves 94.7% and 98.5% precordial set and lead accuracy on PTB-XL, confirming its strength for automatic lead labelling. During model validation, the branched CNN + GRU encoder slightly outcompeted other encoder varieties; however, the performance differences were small, and even a stand-alone branched CNN—with around 78% fewer parameters—achieved above 97% precordial lead accuracy (results shown in Appendix G). On the other hand, the accuracy difference when using different context

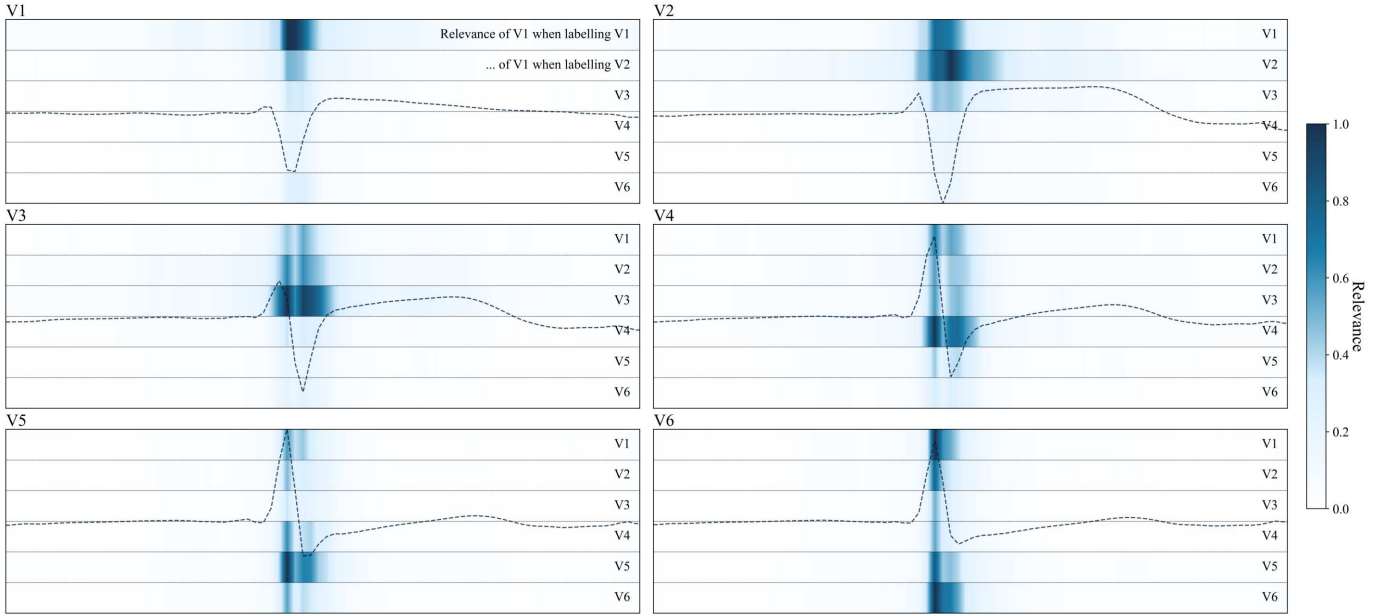


Fig. 10. The *global attributions* of each lead’s morphology when identifying different leads, averaged over a batch of ECGs, and normalised for each labelling target separately. An averaged ECG is overlaid. Higher relevance indicates that the model’s prediction (organised into the rows of each heatmap) was more sensitive to that region of the input. Inter-lead relationships can be interpreted by considering the same labelling target across various leads.

encoders (or none at all), shown in Fig. 6, is more significant. This suggests that the model’s ability to encode inter-lead relationships is a stronger determinant of accuracy than its expressiveness for individual lead features. However, Fig. 4 shows that when the model already has sufficient contextual information, adding extra leads increases training load without improving performance on a prioritised lead subset.

Post-hoc matching algorithms show a significant increase in accuracy when no inter-lead information is shared within the model itself. Applying the Hungarian algorithm to the context-agnostic model drastically improves both set and lead accuracy. However, the fact that the improved accuracy is still significantly less than that of the set attention model implies that the SABs are doing more than just finding an optimal assignment. The results mentioned thus far suggest that inter-lead information is not just helpful but critical for achieving high accuracy, answering Research Question 4.

### B. Robustness to Beat Types

The results of the model’s performance on the different diagnostic classes can be used to answer Research Question 3. On a per-sample basis, the model excels on normal sinus rhythms but exhibits its worst performance on cases of myocardial infarction and conduction disturbances. With that being said, the reduction in performance on these samples is modest, remaining above 97% accuracy.

### C. Lead-by-Lead Analysis

More insight can be gleaned by analysing the types of errors the model makes. The greatest confusion arises between leads V2 and V3, consistent with their known QRS similarity and high positional sensitivity [54]. In general, confusion arises

between adjacent leads (clearly visible as values hugging the leading diagonal in Fig. 5), and barely at all between those with one or more intermediate leads between them. Despite V2/V3 confusion being a source of poor accuracy, the fact that errors between morphologically similar leads dominate the error space is a positive finding. When reconstructing a BSPM, swapping two similar leads will introduce a significantly smaller error than swapping two dissimilar ones. Therefore, the model’s error profile is biased towards mistakes which have less impact on the intended downstream application of this study.

The results in Table I provide a clear indication of which leads are easier to identify than others, answering Research Question 5. In general, the model performs significantly better on the limb and augmented leads than the precordial leads. This is likely because the electrodes forming the limb leads are spatially distant, and the mathematical transformations that the augmented leads undergo make them unique. On the other hand, the precordial leads progress gradually along the chest, meaning neighbouring leads (V1-V2, V2-V3, etc.) can look very similar. This is the reason for most of the errors seen in the precordial leads, and is why their recalls are much worse. With this in mind, the fact that V1 and V6 are the easiest to label makes sense since they lie on the edges of the arrangement.

### D. Spatial Representations

A PCA of the feature representations (Fig. 8) reveals that the model learns a spatially ordered representation of the leads, forming a smooth arc from V1 to V6. After just one SAB, V1 stands out most distinctly, reflecting its unique placement on the right sternal edge. Furthermore, the two distinct planes presented in Fig. 9 stand out as corresponding to the vertical



and horizontal electrical planes from which the ECG views the heart.

This has promising implications for the model’s ability to handle continuous electrode localisation. If the model is indeed internalising chest-surface geometry from these leads, and not just learning arbitrary class boundaries, then regressing those representations onto actual chest-surface coordinates could enable the model to predict each electrode’s continuous position rather than its discrete label. Of course, training the model on a dense, distributed electrode configuration would equip it with a richer representation of the torso’s surface, improving its geometric priors.

The spatial arrangement of these representations, as well as the high lead-labelling accuracy in general, provides an affirmative answer to Research Question 1.

#### E. Interpreting Morphology Attributions

The attributions presented in Fig. 10 show that the R and S waves are the most relevant morphological features when discriminating between precordial leads. This aligns with standard electrocardiographic knowledge regarding the progression of R and S waves across the chest [55].

Dropping the input-baseline multiplier in (11) yields raw gradient saliencies, which can be interpreted as model sensitivities, instead of global attributions. These *local* attributions [56] are shown in Fig. 16 in the appendix, which highlights the P and Q waves. Since these waves typically have small amplitudes, the model must have higher sensitivity in these regions in order to detect meaningful changes. This provides extra information about the important morphology: while the R and S waves are the most *discriminating* features when identifying precordial leads, the model is most *sensitive* to small changes in the P and Q waves. This answers Research Question 4.

#### F. Modifications for BSPM

As it stands, the proposed architecture can be applied to the BSPM case with minimal modification. Due to input-size agnosticism, the same model can even be used with a variety of BSPM lead configurations. A standardised electrode coordinate system does need to be established, however, so that a regression head can output into this coordinate system. Wang *et al.* [9] have already proposed a method for converting a 3D torso geometry into a square representation. This approach has the benefit that, at inference time, the coordinate system is decoupled from the actual geometry of the patient’s torso. The resulting 2D BSPM can then be directly used to reconstruct an atrial surface potential map using the deep learning method proposed in [9].

One design consideration is that the implemented SAB has a computational complexity of  $O(n^2)$ , scaling quadratically with the number of leads. Instead, an *Induced* SAB (ISAB) can be used [48], which introduces  $m$  inducing points, reducing the complexity to  $O(nm)$ .

The model presented in this paper benefits from having a fixed-size label set; there are exactly as many labels as there are input signals. The correct identification of one lead reduces

the space of possible labels for the remaining signals. If the model is confident about  $K - 1$  of its predictions, the identity of signal  $K$  is effectively determined by elimination. While the model acts on all leads in parallel, the set-attention mechanism still allows this constraint to inform predictions. This benefit diminishes as the label space grows, and disappears entirely in the regression case. To translate this benefit to a continuous localisation task, the model should be informed by explicit spatial priors about the electrodes, such as the minimum inter-electrode distance.

#### G. Lead Swap Detection

The proposed model can be used to detect precordial lead swaps, as swapping these leads only affects the order of the signals presented to the model, not their amplitudes or timings. Such swaps simply show up as a permutation of the model’s output.

Passing the six precordial waveforms through the model yields the matrix  $P$  described in (7). The permutation  $\pi^*$  of the input signals can then be estimated by solving (2). If  $\pi^*(i) = i \quad \forall i$  then no swap took place. A swap can be detected if  $\exists i \neq j : (\pi^*(i) = j) \wedge (\pi^*(j) = i)$ , in which case the swapped leads are  $L_i$  and  $L_j$ .

#### H. Limitations and Future Work

The primary limitation of this study is that it solves a classification task as a proxy for the ultimate goal of localisation, which is a regression task. While the applicability of the model on the latter task has been discussed, the realisation of this remains as future work.

Additionally, a more comprehensive analysis of model hyperparameters, such as feature dimension size and the number of attention heads, should be conducted. Informal testing was conducted in this study to settle on reasonable choices, but there is room for optimisation. Additionally, the effect of signal duration (or the number of included beats) on the model’s performance was not investigated.

Furthermore, although this study reports accuracy across broad diagnostic categories, it does not examine the influence of specific conditions, arrhythmias, or patient demographics.

#### I. Implications

This study provides a foundation for future work on fully signal-based electrode localisation. Already, the model presented here can be used to identify leads in a standard ECG in cases where that information has been lost. With further development, the model could be extended to general electrode localisation, which, alongside current imageless ECGI research, would eliminate the need for CT, MRI, or any patient scanning. By lowering costs, reducing radiation exposure, and simplifying workflow complexity, ECGI could see widespread clinical adoption. Due to the high spatial resolution ECGI offers, cardiologists would be able to make more informed diagnoses, allowing life-threatening heart conditions to be treated appropriately.

## VI. CONCLUSION

This study introduces the first fully signal-based, permutation equivariant deep learning architecture for 12-lead ECG lead identification. On an unseen PTB-XL hold-out set, it achieves 98.5% lead-level accuracy and identifies all 12 ECG leads correctly 91.7% of the time. Limb and augmented leads prove easier to identify (with lead aVF having the highest recall), while the precordial leads prove most difficult. The model is fairly robust to abnormal heart rhythms, only dropping to 97.3% accuracy on samples with myocardial infarction and conduction disturbances. The integration of inter-lead information was shown to be crucial for highly accurate labelling, and gradient-based attributions show that the model relies chiefly on R and S wave morphology. In addition, PCA shows that the spatial ordering of the precordial leads is preserved in the learnt feature space, indicating that the model internalises spatial relationships between leads instead of generating arbitrary class boundaries.

Accuracy on precordial leads can be increased by training exclusively on the precordials instead of the full 12-lead set, suggesting that the limb and augmented leads do not provide extra useful contextual information for labelling this lead subset.

This study paves the way for signal-based electrode localisation to be used in imageless ECGI workflows. The lowered costs, reduced radiation exposure, and simplified procedures promise to make high-resolution cardiac mapping readily available in many clinics. Future work will extend the framework to regression-based electrode localisation and validate the performance of other deep learning networks on this task.

## ACKNOWLEDGMENT

Thank you to Joël Karel, Ralf Peeters, and Pietro Bonizzi for their support and advice over the course of this thesis.

## REFERENCES

- [1] M. Cluitmans, D. H. Brooks, R. MacLeod, O. Dössel, M. S. Guillem, P. M. Van Dam, J. Svehlikova, B. He, J. Sapp, L. Wang, and L. Bear, "Validation and opportunities of electrocardiographic imaging: From technical achievements to clinical applications," *Frontiers in Physiology*, vol. 9, p. 1305, Sept. 2018.
- [2] Y. Rudy and B. D. Lindsay, "Electrocardiographic imaging of heart rhythm disorders," *Cardiac Electrophysiology Clinics*, vol. 7, pp. 17–35, Mar. 2015.
- [3] P. Kligfield, L. S. Gettes, J. J. Bailey, R. Childers, B. J. Deal, E. W. Hancock, G. Van Herpen, J. A. Kors, P. Macfarlane, D. M. Mirvis, O. Pahlm, P. Rautaharju, and G. S. Wagner, "Recommendations for the standardization and interpretation of the electrocardiogram: Part i: The electrocardiogram and its technology: A scientific statement from the american heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the american college of cardiology foundation; and the heart rhythm society endorsed by the international society for computerized electrocardiology," *Circulation*, vol. 115, pp. 1306–1324, Mar. 2007.
- [4] M. Cluitmans, *Noninvasive reconstruction of cardiac electrical activity: Mathematical innovation, in vivo validation and human application*. PhD thesis, Maastricht University, Netherlands, 2016.
- [5] W. H. W. Schulze, Z. Chen, J. Relan, D. Potyagaylo, M. W. Krueger, R. Karim, M. Sohal, A. Shetty, Y. Ma, N. Ayache, M. Sermesant, H. Delingette, J. Bostock, R. Razavi, K. S. Rhode, C. A. Rinaldi, and O. Dössel, "ECG imaging of ventricular tachycardia: Evaluation against simultaneous non-contact mapping and CMR-derived grey zone," *Medical & Biological Engineering & Computing*, vol. 55, pp. 979–990, June 2017.
- [6] J. Bergquist, L. Rupp, B. Zenger, J. Brundage, A. Busatto, and R. S. MacLeod, "Body surface potential mapping: Contemporary applications and future perspectives," *Hearts*, vol. 2, pp. 514–542, Nov. 2021.
- [7] J. Reventos-Presmanes, E. Invers-Rubio, E. Ferró, I. Hernández-Romero, C. Herrero-Martín, J. Milagro, D. Lundback, E. Guasch, J. Tolosana, I. Roca-Luque, M. Guillem, L. Mont, J. Guichard, and A. Climent, "Validation of a novel imageless non-invasive electrocardiographic imaging for the characterization of atrial tachycardias," in *2022 Computing in Cardiology (CinC)*, vol. 498, pp. 1–4, 2022.
- [8] R. Molero, A. González-Ascaso, A. M. Climent, and M. S. Guillem, "Robustness of imageless electrocardiographic imaging against uncertainty in atrial morphology and location," *Journal of Electrocardiology*, vol. 77, pp. 58–61, Mar. 2023.
- [9] T. Wang, J. M. Karel, N. Osnabrugge, K. Driessens, J. Stoks, M. J. Cluitmans, P. G. Volders, P. Bonizzi, and R. L. Peeters, "Deep learning based estimation of heart surface potentials," *Artificial Intelligence in Medicine*, vol. 163, p. 103093, May 2025.
- [10] W. Einthoven, "Die galvanometrische registrirung des menschlichen elektrokardiogramms, zugleich eine beurtheilung der anwendung des capillarelektrometers in der physiologie," *Pflüger, Archiv für die Gesamte Physiologie des Menschen und der Thiere*, vol. 99, pp. 472–480, Nov. 1903.
- [11] W. Fye, "A history of the origin, evolution, and impact of electrocardiography," *The American Journal of Cardiology*, vol. 73, pp. 937–949, May 1994.
- [12] R. C. Barr, M. Ramsey, and M. S. Spach, "Relating epicardial to body surface potential distributions by means of transfer coefficients based on geometry measurements," *IEEE Transactions on Biomedical Engineering*, vol. BME-24, pp. 1–11, Jan. 1977.
- [13] D. Levy, S. B. Labib, K. M. Anderson, J. C. Christiansen, W. B. Kannel, and W. P. Castelli, "Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy," *Circulation*, vol. 81, pp. 815–820, Mar. 1990.
- [14] J. E. Burnes, B. Taccardi, R. S. MacLeod, and Y. Rudy, "Noninvasive ECG imaging of electrophysiologically abnormal substrates in infarcted hearts: A model study," *Circulation*, vol. 101, pp. 533–540, Feb. 2000.
- [15] M. Chyrchel, W. Silka, M. Wylaż, W. Wójcik, and A. Surdacki, "Electrocardiography versus echocardiography in severe aortic stenosis with the consideration of coexistent coronary artery disease," *Journal of Clinical Medicine*, vol. 13, p. 1013, Feb. 2024.
- [16] D. Pewsner, P. Jüni, M. Egger, M. Battaglia, J. Sundström, and L. M. Bachmann, "Accuracy of electrocardiography in diagnosis of left ventricular hypertrophy in arterial hypertension: Systematic review," *BMJ*, vol. 335, p. 711, Oct. 2007.
- [17] R. MacLeod and D. Brooks, "Recent progress in inverse problems in electrocardiology," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 1, pp. 73–83, Jan.-Feb. 1998.
- [18] A. M. Van Der Graaf, P. Bhagirath, H. Ramanna, V. J. Van Driel, J. De Hooge, N. M. De Groot, and M. J. Götte, "Noninvasive imaging of cardiac excitation: Current status and future perspective," *Annals of Noninvasive Electrocardiology*, vol. 19, pp. 105–113, Mar. 2014.
- [19] X. Zhou, L. Fang, Z. Wang, H. Liu, and W. Mao, "Comparative analysis of electrocardiographic imaging and ECG in predicting the origin of outflow tract ventricular arrhythmias," *Journal of International Medical Research*, vol. 48, p. 0300060520913132, Mar. 2020.
- [20] A. J. Shah, M. Hocini, O. Khaet, P. Pascale, L. Roten, S. B. Wilton, N. Linton, D. Scherr, S. Miyazaki, A. S. Jadidi, X. Liu, A. Forclaz, I. Nault, L. Rivard, M. E. Pedersen, N. Derval, F. Sacher, S. Knecht, P. Jais, R. Dubois, S. Eliaoutou, R. Bokan, M. Strom, C. Ramanathan, I. Cakulev, J. Sahadevan, B. Lindsay, A. L. Waldo, and M. Haissaguerre, "Validation of novel 3-dimensional electrocardiographic mapping of atrial tachycardias by invasive mapping and ablation," *Journal of the American College of Cardiology*, vol. 62, pp. 889–897, Sept. 2013.
- [21] Y. Ma, U. Mistry, A. Thorpe, R. J. Housden, Z. Chen, W. H. W. Schulze, C. A. Rinaldi, R. Razavi, and K. S. Rhode, "Automatic electrode and CT/MR image co-localisation for electrocardiographic imaging," in *Functional Imaging and Modeling of the Heart* (S. Ourselin, D. Rueckert, and N. Smith, eds.), (Berlin, Heidelberg), pp. 268–275, Springer Berlin Heidelberg, 2013.
- [22] W. H. Schulze, P. Mackens, D. Potyagaylo, K. Rhode, E. Tülümen, R. Schimpf, T. Papavassiliu, M. Borggreffe, and O. Dössel, "Automatic camera-based identification and 3-D reconstruction of electrode positions in electrocardiographic imaging," *Biomedical Engineering / Biomedizinische Technik*, vol. 59, Jan. 2014.

- [23] N. Shenoy, M. Toloubidokhti, O. Gharbia, M. P. Khoshknab, S. Nazarian, J. L. Sapp, V. Singh, A. Kapoor, and L. Wang, "A novel 3d camera-based ecg-imaging system for electrode position discovery and heart-torso registration," *IEEE Journal of Biomedical and Health Informatics*, 2024.
- [24] J. Bayer, C. Hintermüller, H. Blessberger, and C. Steinwender, "ECG electrode localization: 3D DS camera system for use in diverse clinical environments," *Sensors*, vol. 23, p. 5552, June 2023.
- [25] P. Wagner, N. Strodthoff, R.-D. Boussejot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, p. 154, May 2020.
- [26] S. A. Medani, M. Hensey, N. Caples, and P. Owens, "Accuracy in precordial ECG lead placement: Improving performance through a peer-led educational intervention," *Journal of Electrocardiology*, vol. 51, pp. 50–54, Jan. 2018.
- [27] B. Walsh, "Misplacing V1 and V2 can have clinical consequences," *The American Journal of Emergency Medicine*, vol. 36, pp. 865–870, May 2018.
- [28] Z. Wu and C. Guo, "Deep learning and electrocardiography: Systematic review of current techniques in cardiovascular disease diagnosis and management," *BioMedical Engineering OnLine*, vol. 24, p. 23, Feb. 2025.
- [29] H. Narotamo, M. Dias, R. Santos, A. V. Carreiro, H. Gamboa, and M. Silveira, "Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches," *Biomedical Signal Processing and Control*, vol. 93, p. 106141, July 2024.
- [30] M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, June 2016.
- [31] D. Rubio Bizcaino, K. Bujnarowski, M. Matyschik, H. Mauranen, R. Zhao, P. Bonizzi, and J. Karel, "Neural network-based matrix completion for minimal configuration of body surface potential mapping," in *2019 Computing in Cardiology Conference*, Dec. 2019.
- [32] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, Apr. 2021.
- [33] A. Y. Hannun, P. Rajpurkar, M. Haghighpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, pp. 65–69, Jan. 2019.
- [34] J. Huang, B. Chen, B. Yao, and W. He, "Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019.
- [35] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, Nov. 1997.
- [37] L. Guo, G. Sim, and B. Matuszewski, "Inter-patient ECG classification with convolutional and recurrent neural networks," *Biocybernetics and Biomedical Engineering*, vol. 39, pp. 868–879, July 2019.
- [38] G. Keren and B. Schuller, "Convolutional RNN: an enhanced model for extracting features from sequential data," in *2016 International Joint Conference on Neural Networks (IJCNN)*, (Vancouver, BC, Canada), pp. 3412–3419, IEEE, July 2016.
- [39] G. Petmezas, K. Haris, L. Stefanopoulos, V. Kilintzis, A. Tzavelis, J. A. Rogers, A. K. Katsaggelos, and N. Maglaveras, "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets," *Biomedical Signal Processing and Control*, vol. 63, p. 102194, Jan. 2021.
- [40] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
- [41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [42] K. Rjoob, R. Bond, D. D. Finlay, V. McGilligan, S. J. Leslie, A. Rababah, D. Guldenring, A. Iftikhar, C. Knoery, A. McShane, and A. Peace, "Machine learning techniques for detecting electrode misplacement and interchanges when recording ECGs: A systematic review and meta-analysis," *Journal of Electrocardiology*, vol. 62, pp. 116–123, Sept. 2020.
- [43] K. Rjoob, R. Bond, D. Finlay, V. McGilligan, S. J. Leslie, A. Rababah, A. Iftikhar, D. Guldenring, C. Knoery, A. McShane, and A. Peace, "Reliable Deep Learning-Based Detection of Misplaced Chest Electrodes During Electrocardiogram Recording: Algorithm Development and Validation," *JMIR medical informatics*, vol. 9, p. e25347, Apr. 2021.
- [44] Y. Huang, M. Wang, Y.-G. Li, and W. Cai, "A lightweight deep learning approach for detecting electrocardiographic lead misplacement," *Physiological Measurement*, vol. 45, p. 055006, May 2024.
- [45] N.-T. Bui and G.-s. Byun, "The comparison features of ECG signal with different sampling frequencies and filter methods for real-time measurement," *Symmetry*, vol. 13, p. 1461, Aug. 2021.
- [46] C. M. Bishop, "Training with noise is equivalent to Tikhonov regularization," *Neural Computation*, vol. 7, pp. 108–116, Jan. 1995.
- [47] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, Mar. 1955.
- [48] J. Lee, Y. Lee, J. Kim, A. Kosiorek, S. Choi, and Y. W. Teh, "Set transformer: A framework for attention-based permutation-invariant neural networks," in *International conference on machine learning*, pp. 3744–3753, PMLR, 2019.
- [49] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, "Deep sets," *Advances in neural information processing systems*, vol. 30, 2017.
- [50] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.
- [51] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [52] S. Wallis, "Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods," *Journal of Quantitative Linguistics*, vol. 20, no. 3, pp. 178–208, 2013.
- [53] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.
- [54] M. Kania, H. Rix, M. Fereniec, H. Zavala-Fernandez, D. Janusek, T. Mroczka, G. Stix, and R. Maniewski, "The effect of precordial lead displacement on ECG morphology," *Medical & Biological Engineering & Computing*, vol. 52, pp. 109–119, Feb. 2014.
- [55] P. Davey, "ECG," *Medicine*, vol. 42, no. 8, pp. 423–432, 2014.
- [56] M. Ancona, E. Ceolini, C. Öztireli, and M. Gross, "Towards better understanding of gradient-based attribution methods for deep neural networks," *arXiv preprint arXiv:1711.06104*, 2017.

## APPENDIX A RESOURCES

The source code containing the preprocessed PTB-XL data and training scripts for the model can be found in the following GitHub repository: [https://github.com/Nataml/bachelors\\_thesis](https://github.com/Nataml/bachelors_thesis).

## APPENDIX B COSINE ANNEALING SCHEDULER

The learning rate of the optimiser is decayed according to the following equation:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min}) \left( 1 + \cos \left( \frac{T_{cur}}{T_{max}} \pi \right) \right) \quad (9)$$

where  $\eta_t$  is the learning rate at time step  $t$  (calculated as the current epoch + current batch ÷ total batches),  $\eta_{min} = 1 \times 10^{-7}$ ;  $\eta_{max} = 1 \times 10^{-3}$ ;  $T_{cur}$  is the current time step and  $T_{max} = 30$ .

APPENDIX C  
GRADIENTSHAP ATTRIBUTION EQUATIONS

SHAP values satisfy *completeness*

$$f_l(x) - f_l(x') = \sum_{i=1}^K \sum_{t=1}^T \phi_{i,t}^l(X) \quad (10)$$

where  $x'$  is a baseline signal.

GradientShap [50] estimates  $\phi_{i,t}^l$  by using *expected gradients*

$$\hat{\phi}_{i,t}^l(x) = \mathbb{E}_{b \sim p_B, \alpha \sim (0,1)} \left[ (x_{i,t} - b_{i,t}) \frac{\partial f_l(b + \alpha(x - b))}{\partial x_{i,t}} \right] \quad (11)$$

Where  $p_b$  is a distribution of baselines (in this study, chosen to be other samples from the dataset with the lead order shuffled).

APPENDIX D

WILSON CONFIDENCE INTERVALS AND BOOTSTRAPPING

A. Wilson Score Interval

The Wilson interval arises from inverting the score test for a binomial:

$$CI_{\text{Wilson}} : \frac{\hat{p} + \frac{z^2}{2N} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}}, \quad z = 1.96.$$

where  $\hat{p}$  is the approximate success probability (in this study, the recall of a specific lead on an unseen test set) and  $N$  is the number of trials (similarly defined in this study). The 95% confidence interval is computed using  $z = 1.96$ .

The Wilson interval provides a convenient, analytic way to compute confidence bounds on the performance metrics in this paper. It does not suffer from overshoot and zero-width intervals like other binomial proportion intervals do, which occur when  $p$  is close to 1 (as is often the case in this study, see Table I) [52].

B. Bootstrapping

The data used in this study is effectively clustered by ECG sample, within which the model's predictions are highly dependent. This implies that bootstrapping needs to occur at the level of the ECG, rather than at the lead level.

Due to this independence violation, Wilson CIs break down when applied to lead accuracy (equivalent to micro-averaged lead recall). The alternative bootstrapping method used to generate 95% CIs for these metrics is described below.

- 1) For bootstrap sample  $b$ :
  - a) Sample  $N$  ECGs with replacement.
  - b) Compute **LeadAcc** (3) on bootstrap sample  $b$ .
- 2) Obtain 95% CI by measuring the 2.5th and 97.5th percentiles of the **LeadAcc** distribution generated over all bootstrap samples  $b = 1, 2, \dots, B$ .

APPENDIX E  
DEEPSSETS

The DeepSets architecture is shown in Fig. 11.

Table III. Branched CNN + GRUAttn Encoder Architecture

Layer / Block	Type	Parameters / Shape
<b>Part 1: CNN Feature Extractor</b>		
Input Signal	-	Shape: (1, T)
<b>Inception Block 1</b>	-	<b>Output Channels: 128</b>
- Branch 1	Conv1d	Out: 32, Kernel: 1
- Branch 2	Conv1d (bottleneck) Conv1d	Out: 1, Kernel: 1 Out: 32, Kernel: 5, Padding: 2
- Branch 3	Conv1d (bottleneck) Conv1d	Out: 1, Kernel: 1 Out: 32, Kernel: 11, Padding: 5
- Branch 4	MaxPool1d Conv1d	Kernel: 3, Stride: 1 Out: 32, Kernel: 1
Concat	Concatenate Branches	Shape: (128, T)
MaxPool1d	Max Pooling	Kernel: 2, Stride: 2
<b>Inception Block 2</b>	-	<b>Output Channels: 128</b>
- Branch 1	Conv1d	Out: 32, Kernel: 1
- Branch 2	Conv1d (bottleneck) Conv1d	Out: 64, Kernel: 1 Out: 32, Kernel: 5, Padding: 2
- Branch 3	Conv1d (bottleneck) Conv1d	Out: 64, Kernel: 1 Out: 32, Kernel: 11, Padding: 5
- Branch 4	MaxPool1d Conv1d	Kernel: 3, Stride: 1 Out: 32, Kernel: 1
Concat	Concatenate Branches	Shape: (128, T/2)
MaxPool1d	Max Pooling	Kernel: 2, Stride: 2
CNN Output	-	Shape: (128, T/4)
<b>Part 2: GRU with Attention</b>		
GRU Input	Permute CNN Output	Shape: (T/4, 128)
GRU	Gated Recurrent Unit	Input: 128, Hidden: 128, Layers: 2
GRU Hidden States	-	Shape: (T/4, 128)
Attention	Attention Mechanism	Weighted sum over hidden states
- AttnProj	Linear + tanh	In: (T/4, 128), Out: (T/4, 128)
- AttnVec	Linear	In: (T/4, 128), Out: (T/4, 1)
- Softmax	Applied to AttnVec output	(T/4,)
- Matrix Multiplication	(1, T/4) × (T/4, 128)	(128,)
FFN	Feed-Forward Network	-
- Linear 1	Linear + RELU	In: 128, Out: 128
<b>Encoder Output</b>	<b>Feature Vector</b>	<b>Shape: (128,)</b>

APPENDIX F  
ARCHITECTURE

Specifics about the design of the proposed model can be found in tables III, IV, and V.

APPENDIX G  
ENCODERS

Table VI compares the lead and set accuracy of different encoders on the validation set.

Table IV. Set Attention Block (SAB) Architecture ( $D = 128$ )

Layer / Block	Type	Parameters / Shape
Input	Set of K Vectors	Shape: (K, D)
MHA	Multi-Head Attention	Embed: D, Heads: 8
LayerNorm1	Layer Normalization	Acts on: Input + MHA
FFN	Feed-Forward	-
- Linear 1	Linear + GELU	In: D, Out: $4 * D$ (512)
- Linear 2	Linear	In: $4 * D$ , Out: D
LayerNorm2	Layer Normalization	Acts on: Input + FFN
<b>SAB Output</b>	<b>Set of K Vectors</b>	<b>Shape: (K, D)</b>

Table V. Classifier Head Architecture

Layer	Type	Parameters / Shape
Input	Feature Vector	Shape: (128,)
Linear 1	Linear + ReLU	In: 128, Out: 256
Linear 2	Linear + ReLU	In: 256, Out: 128
Linear 3	Linear	In: 128, Out: K (num_leads)
<b>Output</b>	<b>Logits</b>	<b>Shape: (K,)</b>

## APPENDIX H INDIVIDUAL CASES

Below are the model's results on individual ECG samples. The `ecg_id` field is included as the unique identifier for the sample in the PTB-XL database. The model's predictions are given, where the ground truth ordering is [V1, V2, V3, V4, V5, V6]. The confidences associated with each prediction are also given to four decimal places. Certain metadata is attached to the sample, including the ECG report.

### A. ECG 9

- Predictions: [V1, V2, V3, V4, V5, V6]
- Confidences: [0.9999, 1.0, 0.9997, 0.9997, 0.9998, 0.9999]
- Superclass: Normal
- Report: *sinusrhythmus normales ekg*
- Fig. 12

### B. ECG 271

- Predictions: [V2, V1, V3, V4, V5, V6]
- Confidences: [0.9951, 0.9974, 0.9997, 0.9995, 0.9998, 0.9998]
- Superclass: Myocardial Infarction
- Report: *premature atrial contraction(s). sinus rhythm. left atrial enlargement. qs complexes in v2 and tiny r waves in v3. st segments are depressed and t waves inverted in i, avl, v5,6. consistent with ischaemic heart disease with old anteroseptal myocardial*
- Fig. 13

### C. ECG 172

- Predictions: [V1, V2, V3, V4, V5, V6]
- Confidences: [0.9999, 1.0, 0.9997, 0.9997, 0.9999, 0.9999]
- Superclass: Conduction Disturbance
- Report: *sinusrytm svÅrbestÅmmbar el-axel lÅga qrs amplituder i extremitetsavledningarna hÖgersidigt skÅnkelblock*
- Fig. 14

### D. ECG 483

- Predictions: [V1, V5, V3, V3, V4, V6]
- Confidences: [0.999, 0.9795, 0.9788, 0.991, 0.9744, 0.853]
- Superclass: Conduction Disturbance
- Report: *atrial fibrillation with rapid ventricular response. incomplete right bundle branch block. st segments are depressed in i, v2-6. non-specific but consistent with myocardial ischaemia.*
- Fig. 15

## APPENDIX I LOCAL ATTRIBUTIONS

The attributions map calculated without the input-baseline multiplier can be found in Fig. 16.



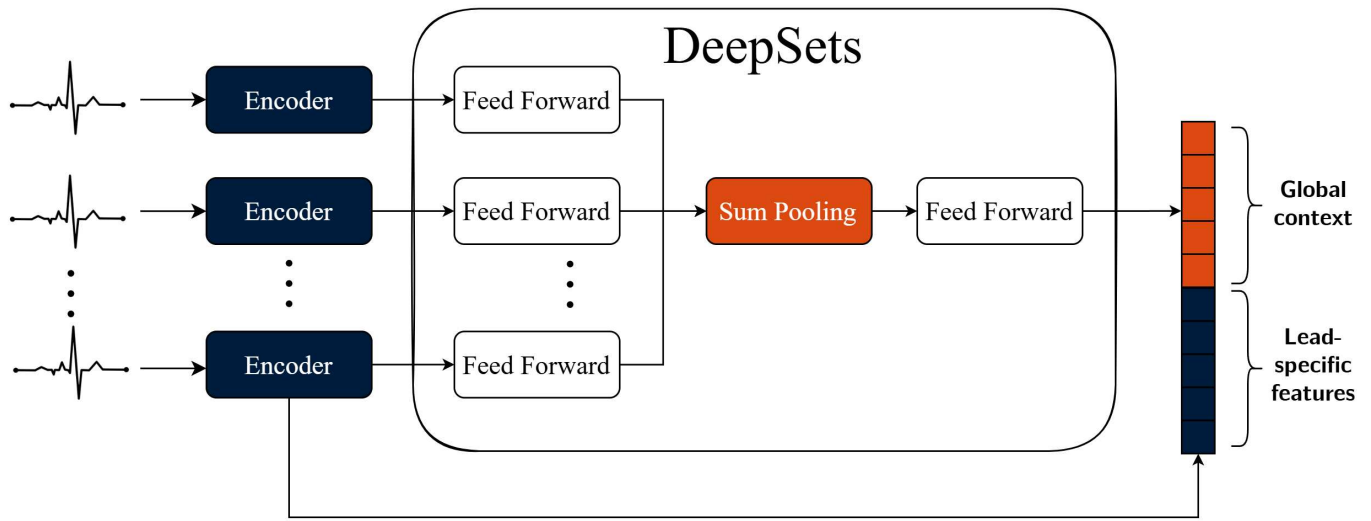


Fig. 11. DeepSets integrated into the proposed architecture.

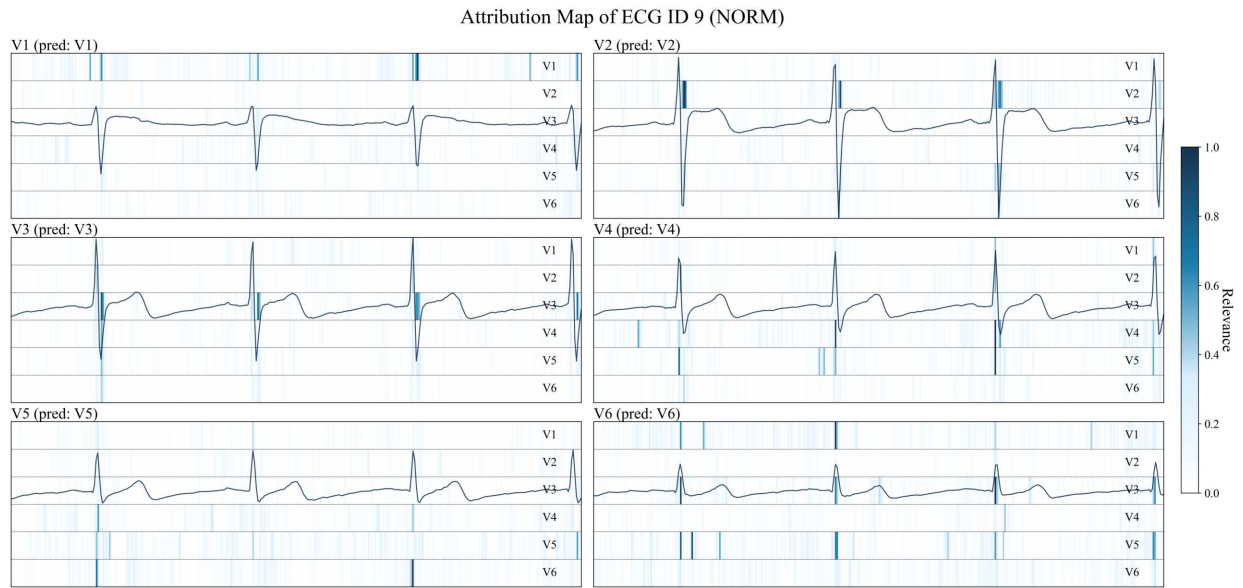


Fig. 12. Global attribution map for ECG ID 9. Normal sinus rhythm.

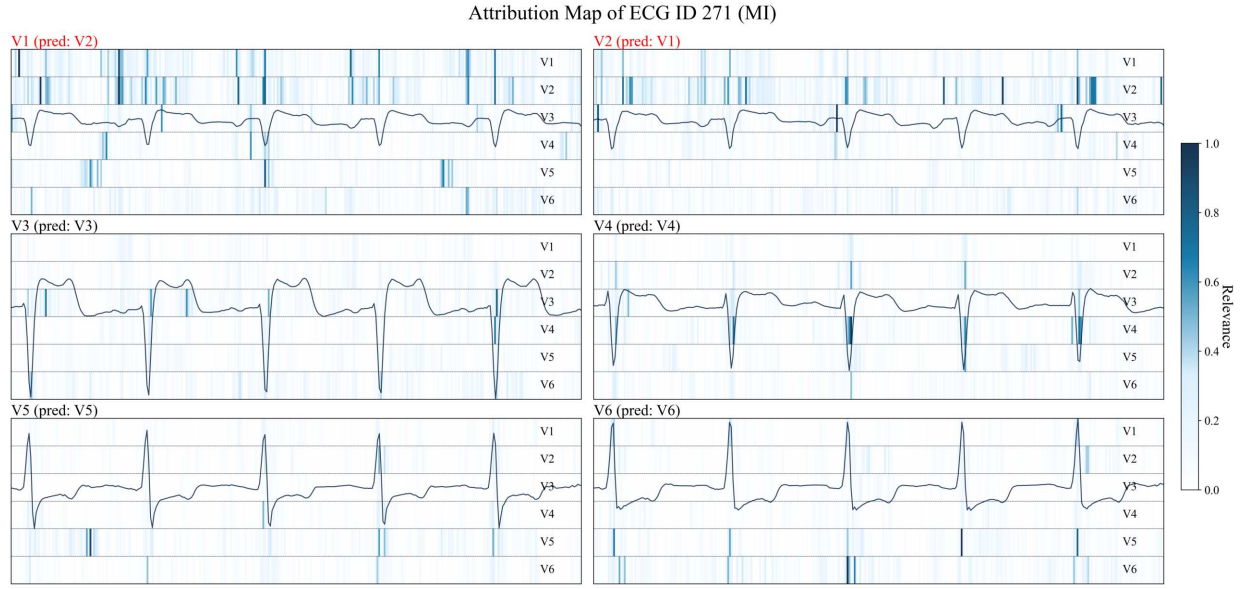


Fig. 13. Global attribution map for ECG ID 271.

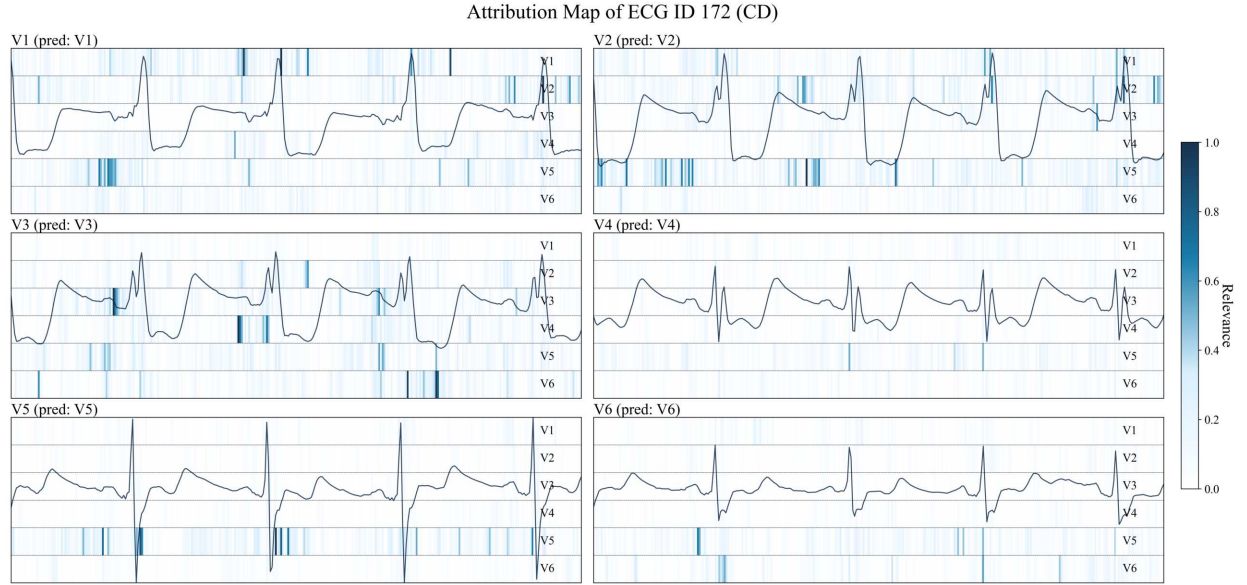


Fig. 14. Global attribution map for ECG ID 172.

Encoder Type	Params (#)	Set Accuracy (%)		Lead Accuracy (%)					
		All	All	NORM	MI	STTC	CD	HT	PACE
Standard 1D CNN	278 592	93.82	97.65	99.37	95.34	97.47	95.25	98.20	40.74
Branched CNN	68 844	93.31	97.45	99.37	95.46	97.38	94.92	98.01	37.04
GRU	182 400	93.17	97.34	99.34	95.00	97.79	94.95	97.89	40.74
Standard 1D CNN + GRU	353 152	94.41	97.79	99.42	<b>96.42</b>	97.73	95.32	98.51	<b>51.85</b>
Branched CNN + GRU	307 017	<b>94.92</b>	<b>97.89</b>	<b>99.60</b>	95.86	<b>98.04</b>	<b>95.72</b>	<b>98.69</b>	48.15

Table VI. Accuracy of the model on the validation set using different encoders (best results are displayed in bold), as well as the number of learnable parameters in each. Accuracy is displayed for each PTB-XL diagnostic superclass: NORM (normal, sinus rhythm), MI (myocardial infarction), STTC (ST-T change), CD (conduction disturbance), HT (hypertrophy). A separate column for pacemaker samples (PACE) is also included.

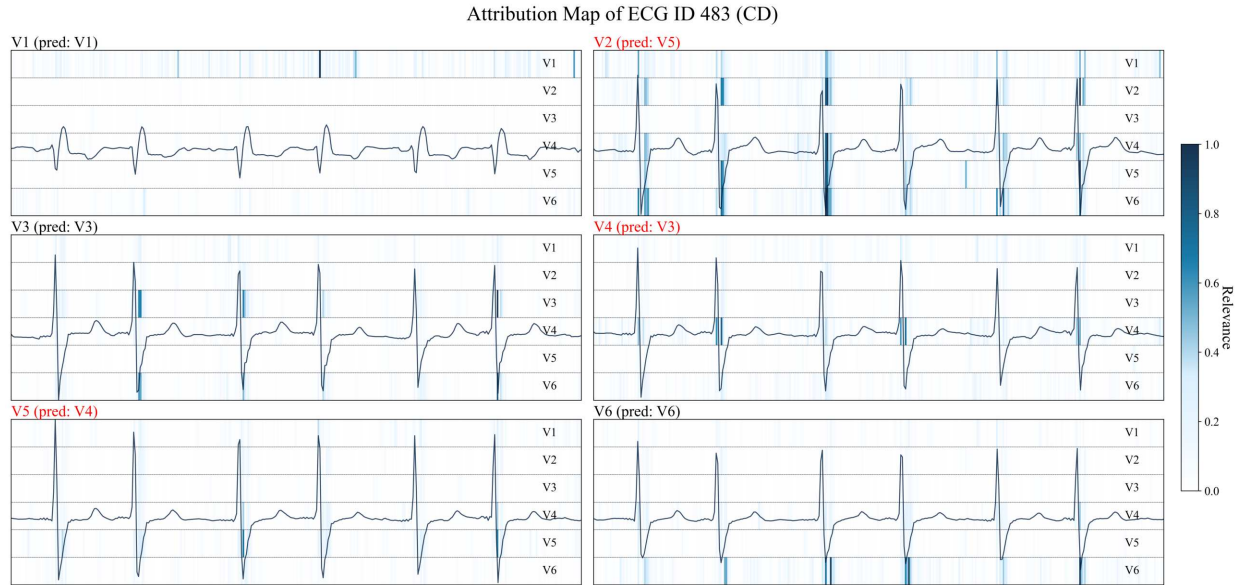


Fig. 15. Global attribution map for ECG ID 483.

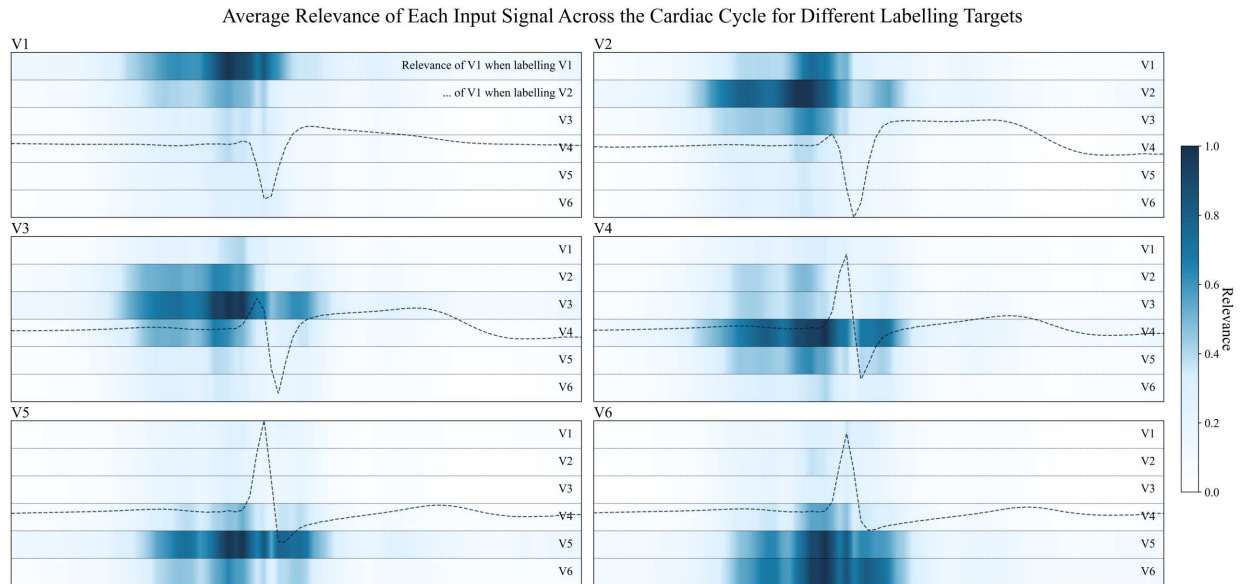


Fig. 16. The *local attributions* of each signal when labelling different leads. Calculated using GradientShap without multiplying the gradients by the input.