

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

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Towards Automated Feature Extraction For Deep Learning Classification of Electrocardiogram Signals

FATIMA SAJID BUTT^{1,2}, MATTHIAS F WAGNER³(Member, IEEE), JÖRG SCHÄFER⁴(Member, IEEE), DAVID GOMEZ ULLATE⁵

¹Frankfurt University of Applied Sciences, Nibelungenpl. 1, 60318 Frankfurt am Main, Germany (e-mail: fbütt@fb2.fra-uas.de)

²Escuela Superior de Ingeniería, Universidad de Cádiz, 11001 Cádiz, Spain (e-mail: fatima.sajidbutt@alum.uca.es)

³Frankfurt University of Applied Sciences, Nibelungenpl. 1, 60318 Frankfurt am Main, Germany (e-mail: jschaefer@fb2.fra-uas.de)

⁴Frankfurt University of Applied Sciences, Nibelungenpl. 1, 60318 Frankfurt am Main, Germany (e-mail: mfwagner@fb2.fra-uas.de)

⁵Escuela Superior de Ingeniería, Universidad de Cádiz, 11001 Cádiz, Spain (e-mail: david.gomezullate@uca.es)

Corresponding author: Fatima Sajid Butt (e-mail: fatima.bütt@fb2.fra-uas.de)

This work was supported in part by the PhD-research programm of the Faculty of Computer Science and Engineering Fb2 of Frankfurt University of Applied Sciences.

ABSTRACT Many recent studies focus on the automatic classification of electrocardiogram (ECG) signals using deep learning (DL) methods. Most rely on existing complex DL methods such as transfer learning or providing the models with carefully designed extracted features based on domain knowledge. A common assumption is the deeper and more complex the DL model, the better it learns. In this study, we propose two different DL models for automatic feature extraction from ECG signals for classification tasks. A CNN-LSTM hybrid model and an attention / transformer based model with wavelet transform for the dimensional embedding. Both of the models extract the features from time series at the initial layers of the neural networks and can obtain a performance at least equal to, if not greater than, many contemporary deep neural networks. To validate our hypothesis, three publicly available data-sets are used to evaluate the proposed models. Our model achieves a benchmark accuracy of 99.92% for fall detection and 99.93% for PTB database for myocardial infarction versus normal heart beat classification.

INDEX TERMS Electrocardiograph; benchmark testing; fall detection; time series analysis; machine learning; deep learning; LSTM; CNN; attention; transformer; PTB XL

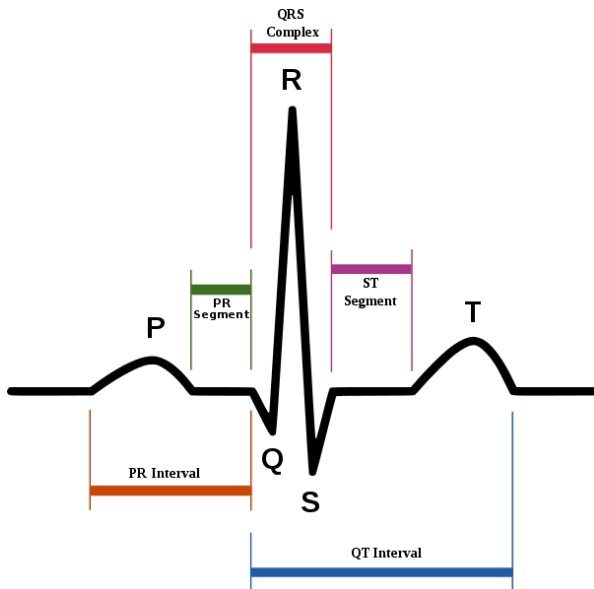
According to [2], in the United States of America alone, the leading cause of death for men and women irrespective of the racial and ethnic group is heart disease. Hence, a timely and accurate diagnosis of heart conditions is of vital importance. An Electrocardiogram (ECG) has been a well grounded method to measure and evaluate the performance of the cardiovascular system. Several techniques exist in both literature and practice to evaluate the ECG signals in different manners. It is one of the most important parameters indicating a person's physiological well being and is extensively used for evaluating the cardiac situation of the patients. It has found its usage widely from getting an overview health picture of a human heart to bio-metric purposes to fall detection and prevention as described in [3]. ECG is a non-invasive method to evaluate the health of the human cardiovascular system. It can detect many heart diseases like atrial

fibrillation, myocardial infarction, AV block, and ventricular tachycardia etc. It gives an insight into the central nervous system especially the autonomic nervous system. Many of the automatic classification techniques using deep learning for ECG use either very deep neural networks or a pre-trained neural network which require either the weights set up to a configuration after being trained on an immense amount of similar data sets. Another approach is to pre-process the data sets by applying some filtration or feature extraction which is based on data domain knowledge and then fed into a neural network to train this. All of the above-mentioned steps involve an explicit understanding of the domain and the pre-process itself. [7] overviews the many ECG feature extraction techniques present in the literature.

I. ELECTROCARDIOGRAM - A TIME SERIES

This section presents an introduction to the ECG signals and the importance of using automatic ECG classification techniques. An ECG is a physiological signal which is measured as the potential difference between electrodes placed on the body surface. The cardiac impulse passes through the heart causing the electrical current to spread from the heart to adjacent tissues. A small current extends to the surface of the body. The electrodes placed on the skin can effectively detect the current on opposite sides of the heart, and record the electrical potentials generated by the current. A normal ECG consists of 5 major deflections called P, Q, R, S, and T waves which make up a single cardiac rhythm as shown in Fig. 1. The P wave lasts about 0.08s and is the smallest, followed by the large QRS complex which lasts between 0.08 and 0.10 seconds. The end of the cardiac cycle is marked by the T wave which lasts about 0.16s. A single waveform varies depending on the size of the heart and the conductive properties of the body which in turn gives the waveform a unique pattern per person [6]. ECG has not only be used to monitor and evaluate cardiovascular system but has also been used as a biometric identifier [39], a predictor of gender and age as described in [1], and for detecting fall activities as in [8].

FIGURE 1. A Normal ECG



A disruption of blood flow to the muscle layer of the heart causes a cardiovascular condition called a myocardial infarction (MI). This disruption is mostly due to the build-up of the plaques in the arteries which result in reduced blood flow to that part of the heart muscle. MI is called a silent heart attack because the patient is not aware of the condition unless they suffer from a heart attack. An early diagnosis of MI is therefore of vital importance as it would help the patients to get timely treatment hence preventing the high percentage of mortality associated with it. Due to

the small amplitude (millivolts), the manual interpretation of ECG signals is time-consuming and prone to errors. This limitation can be mitigated by an automatic diagnosis of heart conditions based on the signals. Our study aims to work towards automation of the cardiovascular disease diagnosis from ECG signals.

In this study, we propose two methods to extract features from a time series automatically and then feed those features to another deep learning model for classification. First, a hybrid model for multiple ECG classification tasks is proposed as an alternative to many complex models which require many pre-processing steps before the actual training. We experimented with a robust hybrid deep learning model for the ECG classification tasks which proved to outperform many state-of-the-art complex models and achieve similar or even better accuracy with no pre-processing steps. The CNN placed in front of a LSTM also known as CNN-LSTM has lately been used for multiple classification tasks but its use for ECG classification has not been systematically explored. The CNN model first searches for the features in high dimensional input data and then after converting it into one-dimensional data, it is fed as an input to LSTM model. The role of the CNN in this context is as an automatic feature extractor. Secondly, a novel attention/transformer model using wavelets for dimensional embedding is introduced to improve the efficiency of the classification process. As it has less trainable parameters than CNN-LSTM it has advantages in terms of (training) performance as shown in Table 6. As a bonus, we evaluate both models also on data for fall detection.

II. RELATED WORK AND OUR CONTRIBUTION

Many recent works focus on automatic ECG classification. Among several different techniques present for ECG classification, deep learning has gained popularity in recent times. It is mainly due to its automatic feature learning and availability of large public data sets. Many deep learning techniques use feature extraction as an essential pre-processing step before feeding the data to the neural network. Most common feature extraction techniques for ECG classification are continuous wavelet transform (CWT), discrete cosine transform (DCT) [13], Pan-Tompkins algorithm [14] and discrete wavelet transform (DWT) [11]. One of the major disadvantages of using wavelet transform as a feature extractor is that the complexity of the process increases with the increase in decomposition level. All feature extraction processes require some domain knowledge in order to efficiently extract relevant features from the data. Therefore, we aim to explore the research question of whether a similar state-of-the-art results can be achieved with no pre-processing and with a simpler model architecture in an efficient manner in terms of resources and computation. Our contribution to this study is twofold. Firstly we use an ensemble of CNN-LSTM to achieve accuracy equivalent to many complex models, and secondly, to further optimize the automatic feature extraction, we introduce a new embedding technique in the attention/transformer encoder architecture which uses discrete

wavelet transform to extract features from the ECG time series and feeds them to the attention mechanism. In the following sub-sections, we present the state of the art in the related work and highlight our contributions.

A. CNN-LSTM ARCHITECTURES

Jambukia et al. (2015) [6] present an overview of ECG classification into different types of arrhythmia. Another current review on deep learning methods for ECG arrhythmia classification [18] deduced that among many deep learning models, CNNs and LSTMs were among the most effective for learning arrhythmia in ECG classification tasks. The use of CNN-LSTM architecture for classification is not entirely novel. Socher et al. [28] in 2012 proposed a model for 3d object classification which combined a CNN with an RNN. They concluded that the CNN provides the translation variance for lower level features whereas RNNs can learn the interactions and compositional features in the data. Zheng et al. [24], transformed the data acquired by three-axis accelerometer into an image format and then used CNN with three convolution layers to classify human activities. Xia et al. (2020), [25] used CNN after a LSTM layer to classify human activity recognition (HAR) with an accuracy of 95.85%. Ordóñez et al. (2016), [26] proposed an activity recognition classifier, which combined deep CNN and dense layers. However, not many studies focus on hybrid CNN-LSTM models for ECG classification. Studies like [20], [23], [54] and [21] have implemented CNNs and its variants for ECG classifications. [27] used RNNs to classify ECG signals. In this study, we have not only performed multiple classifications with CNN-LSTM model for ECG but also worked with three different ECG data sets including data for fall detection to present proof of concept that CNN placed in front of LSTM surpasses many complex and pre-trained models.

B. ATTENTION AND TRANSFORMER ARCHITECTURES

The seminal paper by Vaswani et al. "Attention is All you Need" [43] has triggered an enormous amount of successful applications of attention mechanisms and transformer architectures in deep learning. However, the vast majority of research is focusing on the natural language processing (NLP) domain. Little research has been carried out in applying attention-based architectures in other domains such as time series analysis. One of the first papers in this regard is LSTNet by Lai et al. [45], where the authors introduce long- and short-term time-series network (LSTNet) using the convolution neural networks and recurrent neural networks to extract short-term local dependency patterns and to discover long-term patterns for time series trends. In [41] Shun-Yao Shih et al. apply an attention mechanism for multivariate time series data in three medical domains. Song et al. [42] have applied attention models to clinical time series analysis. A systematic and comprehensive analysis and study of utilising attention mechanisms, however, in the time-series domain is still outstanding.

One of the shortcomings of the self-attention mechanism preventing its application for e.g. time-series is the $\mathcal{O}(n^2)$ complexity with regards to the length of the input vector, i.e. the length of the time series in our case. To address this problem LinFormer has been introduced by Wang et al. in [44]. Linformer is the first theoretically proven linear-time transformer architecture and henceforth might be suitable also for long time series. The linear scaling is achieved by discovering that self-attention is low rank and henceforth projecting information on a low rank constant sub-dimensions achieves to decouple from the $\mathcal{O}(n^2)$ scaling. Recently Rabe and Staats [46] have proposed an algorithmic solution to at least reduce the memory (but not the time) complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$.

In this paper, we propose a novel attention architecture using projection on discrete wavelet components as a means to address the $\mathcal{O}(n^2)$ problem and for dimensional embedding. Moreover, the results show that using this technique, an *attention only* architecture is on par with or even outperforms more complex models and has several additional advantages such as e.g. better run-time performance.

III. ALGORITHMS

This section provides a brief overview of the algorithms and the technologies that were used during the course of this study and also presents the state of the art in the respective technologies.

A. CNN-LSTM MODEL

We define some basic terms related to the convolutional neural network and LSTM for clarity in the following section.

1) CNNs and LSTMs

Convolutional neural networks, introduced as LeNet in 1989 by LeCun, have revolutionized the field of image recognition and are among the most prominently used deep neural networks. They were named after the linear matrix operation called convolution. Since convolution is a linear operation, the convolution layer is often followed by a non-linear layer. Though introduced earlier, it gained popularity after its application as the first deep neural network applied for object recognition in ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012. AlexNet proved to excel on the largest computer vision data set as compared to contemporary methods. In recent times, [15] presents a state of the art review of the recent deep CNNs architectures. The individual CNN components are explained in [16] in a structured way. The most common architectures of CNNs include an input layer, a convolution layer followed by a pooling layer, a drop-out layer, and a fully connected layer followed by an output layer. The number of layers and their layout can change depending on different problem sets. The

convolution operation¹ itself is given by:

$$V_{i,j} = X * W_{i,j} + b = \sum_L X^L * W_{i,j}^L + b$$

$$= \sum_L \sum_{k,l} X_{kl}^L * W_{i+k,j+l}^L + b$$

where X or X^L resp. denote the L -th input matrix. W is the convolution kernel matrix, b is the bias, and $V_{i,j}$ is the output matrix after convolution.

CNNs have been known for their excellent feature extraction capability. One of the most salient features of CNN is its translation invariance. Therefore, it can extract features irrespective of the spatial context. Though it has proven to be beneficial in image recognition, its application and usefulness in time series is still to be exploited fully. Cases, where historical context is relevant for classification, would not work well with CNN alone as it does not carry any information about the history of the time series. The CNNs initially extract the local features in the sub-regions of the time series and then the information is merged in later stages to detect the higher order features. We applied 1D convolution to the time series both with univariate and multivariate data sets. ECG Human activity recognition (HAR) data set and PTB diagnostic data set contained one feature each, so the 2D convolutional operation would not be suitable as it will incorrectly convolve across multiple time series. Long short term memory (LSTM) networks – a variation of recurrent neural networks (RNNs) – were introduced by Hochreiter [51] in 1997. They tend to present a solution to the common problem associated with RNNs called vanishing and exploding gradients. Classical RNNs can in principle keep track of long term dependencies in the sequence. However in practice, during the backpropagation phase of training, these long term gradients either vanish or explode due to the successive multiplicative operations. A LSTM consists of a chained loop structure. Each LSTM unit is made up of an input gate, an output gate and a forget gate. The LSTMs keep the long term memory by maintaining a cell state that sustains a part of the information from earlier states by forgetting and/or applying increment operations on the previous states. Adding a CNN in front of a LSTM helps to feed the LSTM the features from CNN which were extracted from the time series.

2) CNN-LSTM Architecture and Algorithm

Fig. 2 and Fig. 3 provide a more graphical overview of our model. Initially (1*N) time series with N time stamps are convolved with k filters each of size M*1. Later the k feature maps each of size (N-M)+1 time stamps are generated which are passed through a dropout layer followed by the max pooling layer and later fed into the LSTM layer where the encoded extracted features are fed into it from CNN. LSTM unit is followed by a fully connected or Dense layer which

applies Softmax as an output function to classify the input time series into one of the output classes.

Layer (type:depth-idx)	Param #
=====	
CNN_LSTM	--
--CNN: 1-1	--
--ReLU: 2-1	--
--Sequential: 2-2	--
--Conv1d: 3-1	561
--BatchNorm1d: 3-2	374
--ReLU: 3-3	--
--Sequential: 2-3	--
--Conv1d: 3-4	24,000
--BatchNorm1d: 3-5	128
--ReLU: 3-6	--
--MaxPool1d: 3-7	--
--Sequential: 2-4	--
--Conv1d: 3-8	8,256
--BatchNorm1d: 3-9	128
--ReLU: 3-10	--
--MaxPool1d: 3-11	--
--Sequential: 2-5	--
--Conv1d: 3-12	8,256
--BatchNorm1d: 3-13	128
--ReLU: 3-14	--
--MaxPool1d: 3-15	--
--Sequential: 2-6	--
--Conv1d: 3-16	8,256
--BatchNorm1d: 3-17	128
--ReLU: 3-18	--
--MaxPool1d: 3-19	--
--Sequential: 2-7	--
--Conv1d: 3-20	8,256
--BatchNorm1d: 3-21	128
--ReLU: 3-22	--
--MaxPool1d: 3-23	--
--Sequential: 2-8	--
--Conv1d: 3-24	16,512
--BatchNorm1d: 3-25	256
--ReLU: 3-26	--
--Dropout: 3-27	--
--MaxPool1d: 3-28	--
--LSTM: 1-2	264,192
--Linear: 1-3	2,562
=====	
Total params: 342,121	
Trainable params: 342,121	
Non-trainable params: 0	
=====	

FIGURE 2. CNN-LSTM Model for PTB DB

Algorithm 1: Classification of ECG signals with Raw Signals using CNN-LSTM

Input: A time series ECG raw data ts
Output: The classified label l

- 1: $ts \leftarrow \text{RAW_VALUE_EXTRACTION}(ts)$
- 2: $features \leftarrow \text{CNN}(ts)$
- 3: $l \leftarrow \text{LSTM_CLASSIFICATION}(features)$
- 4: **return** l

Algorithm 1 layouts the algorithms for extracting features from the ECG signal and classifying them using a CNN-LSTM model. The number of CNNs and LSTMs can be varied but we used a maximum of 5 1-d convolution layers in front of 3 LSTM layers.

B. ATTENTION MODEL

For the reader's convenience, we recall the basic definitions of the attention mechanism following [43].

¹in the two dimensional case – the one-dimensional case is analogous

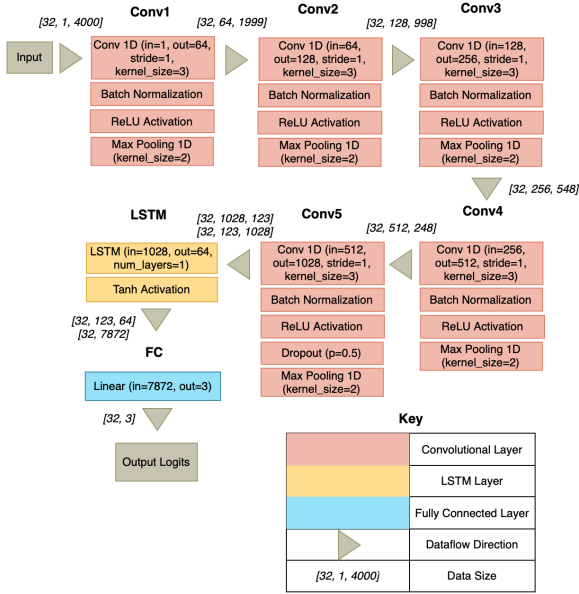


FIGURE 3. Final CNN-LSTM architecture for fall and HAR using ECG signals

1) Attention

Attention is defined as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (1)$$

The transformer uses Multi-Head Self-Attention (MHA) allowing the model to jointly attend to information at different positions of the time-series or different semantics of the domain. MHA is defined as

$$\text{MultiHead}(Q, K, V) = \quad (2)$$

$$\text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) W^O, \quad (3)$$

where Q, K and $V \in \mathbb{R}^{n \times d_m}$ are input embedding matrices, n is the length of the (time) series, and d_m is the embedding dimensions, and h is the number of heads. Each head is defined as

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (4)$$

$$= \text{softmax}\left[\frac{QW_i^Q(KW_i^K)^T}{\sqrt{d_k}}\right]VW_i^V, \quad (5)$$

where $W_i^Q, W_i^K \in \mathbb{R}^{d_m \times d_k}$, $W_i^V \in \mathbb{R}^{d_m \times d_v}$, and $W^O \in \mathbb{R}^{hd_v \times d_m}$ (projection onto the output) are learned matrices and d_k, d_v are hidden dimensions of projection subspaces. For simplicity in the sequel, we drop the differentiation between d_m, d_k and d_v and refer to them by d .

The matrices Q, K and V are usually referred to as query, key and value matrices to remind of the associative memory architecture of a transform, compare e.g. also the analysis in [40].

2) Attention and Dimensional Embedding

For applying the attention mechanism to time-series one has to decide on the proper dimensional embedding, i.e. on the dimension of the embedding subspace and on the embedding transformation. We recall that in the domain of ECG the “natural” dimension is small. For instance, the signals are one-dimensional if a one-dimensional channel is used (as is the case in this paper for the attention/transformer model, i.e. Algorithm 2). Even if multi-channel ECGs are used usually the number of channels is limited to a small number of 3 to maximally 12 channels. Henceforth, if we used the channel as the embedding, the dimension would be 1 in our case, i.e. $d = 1$. This is way too small to capture interesting patterns and, indeed, a test showed that the gradient descent does not converge, but stays constant after one or two initial updates. Furthermore, as depicted in the previous section, the self-attention suffers from an $\mathcal{O}(n^2)$ problem. We propose the following architecture to solve both problems simultaneously:

- 1) Assuming $m \ll n$. For simplicity of the notation, we assume without loss of generality that n is divisible by m , i.e. $n = mw$. This effectively segments the time-series n into n_m “windowed” segments of length w , where $m \in 1, \dots, k$ with $k := n/w$. If n is not divisible without remainder we could fill the time series with zeros (padding).
- 2) For each “windowed” sub-time-series t_{n_k} we calculate the decomposition to a chosen (fixed) wavelet by performing a discrete wavelet transformation (DWT), see below. Assuming that the result of applying the DWT is in dimension p , we have transformed the input from \mathbb{R}^n into $\mathbb{R}^{m \times p}$ depicted in Fig. 4.

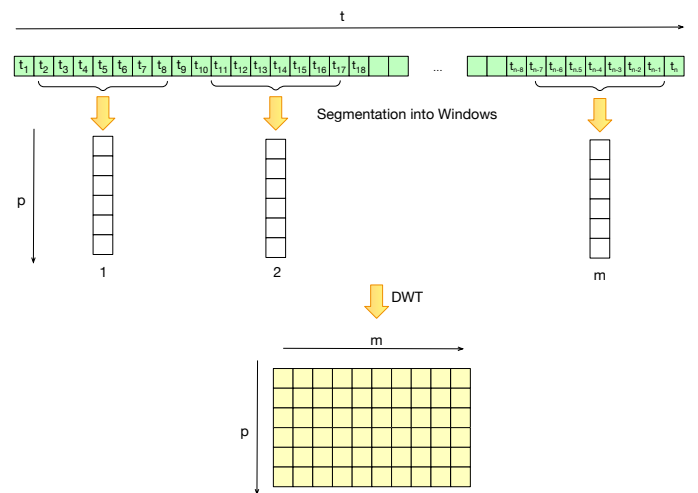


FIGURE 4. Dimensional Embedding

Remark 1: In this paper, we propose a deterministic embedding using DWT rather than a learned, randomly initialised embedding, which is an alternative approach that has been used in other attention architectures in the past. This should

– in theory – require less training data – a conjecture that we want to validate in future work using synthesized data.

3) Discrete Wavelet Transformation (DWT) for Dimensional Embedding

For the reader's convenience, we recall a few well-known definitions and theorems wavelet theory following [47] and [48]. We consider signals as real-valued functions. We call a function $\psi \in L^2(\mathbb{R})$ an orthonormal wavelet if its dyadic translations and dilations of $\psi_{jk}(x) = 2^{\frac{j}{2}} \psi(2^j x - k)$ constitute a Hilbert space and if in addition, it satisfies a regularity (admissibility) condition, namely $\int_0^\infty |\psi(t)|^2 \frac{dt}{t} < \infty$, ensuring convergence and $\int \psi(t) dt = 0$. The function ψ is usually called the *mother wavelet* and its child wavelets are defined as $\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k2^j}{2^j}\right)$. A projection of a function $x(t)$ onto $\psi_{j,k}$ is then given by

$$\gamma_{jk} = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k2^j}{2^j}\right) dt. \quad (6)$$

The crucial idea is to decompose the space $L^2(\mathbb{R})$ into resolution spaces of different resolutions. First, the *resolution space* V_0 is defined as the space of piecewise constant functions on subintervals of $[n, n+1]$ with $n = 0, \dots, N$. If we define the corresponding step function

$$\phi(t) = \begin{cases} 1, & \text{if } 0 \leq t < 1 \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

then $V =$ has dimension N and the N functions $\phi_0 := \{\phi(t - j)\}_{j=0, \dots, N-1}$ constitute an orthogonal basis. Analogously the *refined resolution spaces* V_k are defined as the spaces of functions constant on each subinterval $[n/2^k, (n+1)/2^k]$. This yields a nested sequence of embedded spaces

$$V_0 \subset V_1 \subset V_2 \subset \dots \subset V_k. \quad (8)$$

We denote the orthogonal complements of V_{k-1} in V_k as W_{k-1} , i.e. $V_k = V_{k-1} \oplus W_{k-1}$ and call it the *detail space*. This yields an orthogonal decomposition at level k as follows:

$$V_k = V_0 \oplus W_0 \oplus W_1 \oplus \dots \oplus W_{k-1} \quad (9)$$

Then the k -level *discrete wavelet transformation* (DWT) is defined as the change of coordinates from ϕ_k to $(\phi_0, \psi_0, \psi_1, \dots, \psi_{k-1})$, where $\phi_k := \{\phi_{jk}\}_{j=0, \dots, N-1}$ and $\psi_k := \{\psi_{jk}\}_{j=0, \dots, N-1}$, resp. denote the family of functions obtained from the mother wavelet.

This yields a filter bank interpretation of DWT, wherein each step the signal is decomposed into an averaged and a detailed signal using low-pass and a high-pass filter depicted graphically in Fig. 5.

Thus, any signal can be decomposed into an averaged and a detailed signal, namely V_0 and W_0 . Due to the recursive nature, this can be extended to any desired level k . Please note, that due to the dyadic nature the data size is reduced by a factor of 2 in each step. For further details, we refer to [47] and [48] as well as e.g. [49].

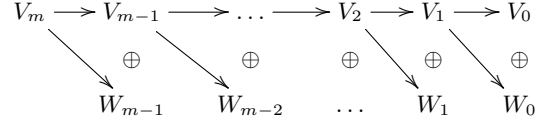


FIGURE 5. Wavelet Transform Pyramid

We used Haar and Daubechies wavelets as well as symlets (symmetrized version of Daubechies wavelets) as a mother wavelet.

4) Transformer Architecture and Algorithm

The DWT can be used not only for dimensional embedding but also for noise reduction as one can ignore some or all coefficients for the detailed spaces. To explore the impact we tried several configurations, see Table 1².

TABLE 1. Attention Models

	Wavelet	Spaces	Dim Embedding	Hid Dim	Frequency	Epochs	Accuracy
ATT2	db8	V_0	13	150	0.1k	1k	98.14
ATT3	db8	V_0	12	150	0.1k	1k	98.83
ATT4	db8	$V_0 \oplus W_0$	24	150	0.1k	1k	99.18
ATT5	db8	$V_0 \oplus W_0$	24	150	0.1k	1k	99.11
ATT6	db8	$V_0 \oplus W_0$	26	250	0.1k	1k	99.45
ATT7	sym6	$V_0 \oplus W_0$	22	250	0.1k	1k	99.38
ATT8	db8	$V_0 \oplus W_0$	26	250	10k	1k	99.59
ATT9	db8	$V_0 \oplus W_0$	26	250	10k	2k	99.59
ATT10	db5	$V_0 \oplus W_0$	20	250	10k	1k	99.24
ATT11	db5	$V_0 \oplus W_0 \oplus W_1$	28	250	10k	1k	99.73

We deployed a transformer architecture with one head, four transformer encoder layers, and a dimension of 150 or 250 hidden units of the feed-forward network. The network's architecture of the best model, att11, is depicted in Fig. 6.

Layer (type:depth-idx)	Param #
PositionalEncoding: 1-1	---
TransformerEncoder: 1-2	---
ModuleList: 2-1	---
L-TransformerEncoderLayer: 3-1	17,638
L-TransformerEncoderLayer: 3-2	17,638
L-TransformerEncoderLayer: 3-3	17,638
L-TransformerEncoderLayer: 3-4	17,638
Dropout: 1-3	---
Linear: 1-4	954
Total params: 61,762	
Trainable params: 61,762	
Non-trainable params: 0	

FIGURE 6. Transformer Architecture of Model ATT11

The original input tensor has a dimension of $[14552, 1, 187]$ corresponding to 14522 data rows, 1 feature, and a sequence

²Frequency refers to the frequency of the positional encoding. It should be remarked, that model ATT5 differs from ATT4 by an additional residual connection.

length of 187. The sequence was split into 17 chunks of window length 11. Each subsequence of length 11 was converted using DWT. For instance, for db8 for att7, this leads to a transformed tensor of [14552, 22, 17]. (Note, that due to boundary effects, the embedding dimension is not always a multiple of 11.).

As a positional encoding, we tried the usual Fourier encoding and used frequencies f of $f = 100$ and $f = 10,000$, resp. While adding positional encoding is questionable after embedding using DWT, we experimentally found the results to be improved by a small amount.

All models were trained with a batch size of 256 and a learning rate of 0.001 using the Adam optimizer [50]. Algorithm 2³ layouts the algorithms for extracting features from ECG data and classifying it using a transformer/attention model.

Algorithm 2: Classification of ECG signals with Raw Signals using Attention/Transformer

Input: A time series ECG raw data ts

Output: The classified label l

- 1: $X \leftarrow \text{RAW_VALUE_EXTRACTION}(ts)$
 - 2: $(V_0 \oplus W_0, \dots, V_n \oplus W_m) \leftarrow \text{DWT}(X)$
 - 3: $\mathbf{V} \oplus \mathbf{W} \leftarrow (V_0 \oplus W_0, \dots, V_n \oplus W_m)$
 - 4: $\mathbf{V} \oplus \mathbf{W} \leftarrow \mathbf{V} \oplus \mathbf{W} + \text{POS_ENC}(\mathbf{V} \oplus \mathbf{W})$
 - 5: **for** $i \in \text{Layers}$ **do**
 - 6: $X \leftarrow \text{TRANSFORMER}_i(\mathbf{V} \oplus \mathbf{W})$
 - 7: **end for**
 - 8: $l \leftarrow \text{LINEAR_FEED_FORWARD}(X)$
 - 9: **return** l
-

IV. DATA PREPARATION AND EXPERIMENTAL SETUP

Since our study includes multiple data sets, therefore this section explains the preparation steps taken for each data set and the overall experimental setup. The experiments were carried out on a GPU server. All the experiments were implemented using PyTorch library because of its supportive architecture with GPUs. The main aim of the experiments was Fall detection and MI detection using ECG signals in an automated and efficient manner.

A. ECG DATA SET FOR FALL DETECTION

The ECG HAR data set is the only one to our best knowledge for the detection of different human activities including falls using ECG signals. It was originally collected by [8], as an experiment that was part of the study from [17]. It originally consisted of two classes: one for the ECG of a person falling from the bed and another one for the ECG of a resting person. It was later augmented with two more data sets, [10] and [9], by up-sampling the original data set. In addition to that, another augmentation method called slicing was applied to

the data set. Slicing has been explained in detail in [19]. After the addition of new data sets, the final version has three classes namely: fall, rest, and daily activities.

The overview of the final class distribution in data set is depicted in Table 2.

TABLE 2. Total number of samples in the ECG HAR data set [8]

Label	Total Count
Fall	500
Rest	474
DA	296
Total	1270

In the previous experiments, the data set was filtered, converted to wavelet transform, and later to 3-D images called scalograms. These scalograms were used to first fine-tune and then train, a pre-trained AlexNet and GoogLeNet. The state-of-the-art validation accuracy obtained for classification with this data set is 98.44%. This accuracy was obtained after applying extensive pre-processing to the data set. Our current model beats the state-of-the-art validation accuracy and we achieve a 99.21% with no pre-processing and only fine tuning the ensemble model.

B. PTB DIAGNOSTICS DATA SET

After working with ECG for falls and daily activities, the model had to be tested on a data set that is standardized and is available publicly. In the second set of experiments, a publicly available data set called, PTB diagnostic, was used which is freely available but is used as a standard for ECG classification tasks.

The original PTB data set consists of 549 records from 290 subjects which were aged 17 to 87 with a mean age of 57.2. 209 subjects were males with a mean age of 55.5 and 81 females with a mean age of 61.6 (for 1 female and 14 male subjects, age was not recorded). Each record has 15 measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) together with the 3 Frank lead ECGs (vx, vy, vz). The data from lead II was used to train the model which outperforms the databases which even use all 12 lead data [60]. ECG beats have been extracted using the method described in [20]. The data set used was curated into two classes named normal and abnormal (myocardial infarction). In the previous prominent work [12], all the 12 leads were separately evaluated to determine which leads contributes most to the classification. We used lead II of the data set to differentiate between healthy controls and that with myocardial infarction. Since in the previous two experimental phases only one lead of ECG was used, we used another publicly available data set and used all 12 of its leads to reaffirm the usefulness of the ensemble model for both univariate and multivariate data sets.

³Note, that for simplicity of notation we use the variable $\mathbf{V} \oplus \mathbf{W}$ from the decomposition in equation 8 generically, i.e. some of the components of $\mathbf{V} \oplus \mathbf{W}$ might be empty.

C. PTB XL DIAGNOSTICS

PTB-XL is one of the largest freely accessible ECG data sets available. It was collected over a span of seven years between 1989-1996. It was made publicly available in 2020 in a structured database by Physikalisch-Technische Bundesanstalt (PTB). The data set consists of a total of 21837 records of 12-lead ECG each comprising of 10 seconds. It is a gender-balanced data set with 52% male and 48% female records with an age range of 0-95 years. The data set consists of various diagnoses and a large number of healthy controls as well [5]. PTB XL has a standardized set of pre-processing instructions for the data set. Since different labels are heavily imbalanced and imbalanced classes can introduce bias in the trained model, it is important to divide the data set in a way that each of the classes is represented equally in each subset. Stratified sampling was used to divide it into training-validation-testing data sets. The data set has multiple classification categories as shown in Fig. 7. The goal is to classify MI from other heart conditions, models were trained for diagnostic superclass and myocardial infarction detection using Algorithm 1.

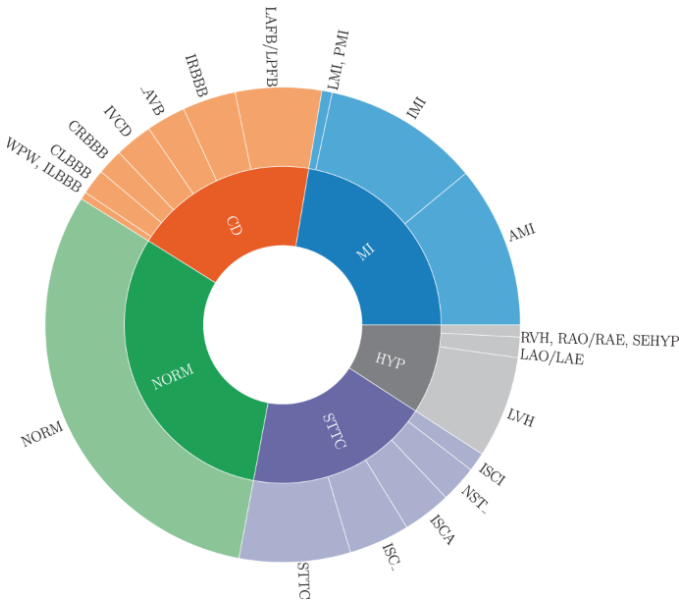


FIGURE 7. Class Distribution for PTB-XL data set

V. RESULTS

Several methods exist in the literature to evaluate a DL classifier. We evaluate our classifiers using accuracy, area under curve (AUC), confusion matrix, sensitivity, and specificity. A detail of the results obtained by applying each of the algorithms and the respective data set is explained in this section. In the sequel, we present the results acquired for each data set and also compare our results to the other state-of-the-art work. An overview of the section is presented in Table 10.

A. ECG HAR DATA SET

The data set was trained initially using a plain LSTM to compare the model performance with the previous experiments. The data was fed into the model without any pre-processing. The LSTM initially gave an accuracy of 49.80% which was increased to 54% by fine tuning the hyperparameters. In the previous experiments, extensive pre-processing was carried out to extract the related features and then those features were fed into the model. That approach though gave excellent accuracy but was not automated. LSTMs have been shown to have a sense of previous timestamps or history in the time series, but CNNs have superior feature extraction capability. To test our hypothesis, a CNN was placed on top of the LSTM layer. The accuracy immediately improved to 93%. After some fine tuning the hyperparameters and adjusting the number of CNN layers, the validation accuracy got better than the state-of-the-art results. A testing accuracy of 99.21%-100% was achieved and a validation accuracy of 99.21% was achieved. For the first data set, the results were almost perfect with a validation accuracy of 99.21% and a testing accuracy of 99.21%-100%. The previous work achieved similar accuracy but with transfer learning and pre-processing the signals by converting them into wavelet transforms and then into scalograms. This model achieves similar accuracy even by avoiding all those steps. The following Table 3 shows the confusion matrix for the testing data set showing an almost perfect accuracy of 99.22%.

TABLE 3. Confusion Matrix for Fall Detection ECG data set using CNN-LSTM (Algorithm 1)

		Actual		
		DA	Fall	Rest
Predicted	DA	30	0	0
	Fall	1	52	0
	Rest	0	0	45

Fall detection using ECG signals was also done by applying Algorithm 2 to the HAR data set. Each sequence in the data set consists of 4000 time stamps. The initial tensor size was [1273, 1, 4000], which is of the format [Total Sequences, Number of features, Sequence length]. Each of the ECG sequences was divided into 100 chunks of 40 time stamps each, and then wavelet transform was calculated for each chunk resulting in a final dimension of [1273, 108, 40]. The model was trained in 403.39 seconds. The result is again achieved without any manual feature extraction or transfer learning model. The following Table 4 shows the confusion matrix for the testing data set showing also accuracy of 95.31%

B. PTB DIAGNOSTICS

Algorithm 1, i.e. CNN-LSTM was used to model PTB diagnostic to differentiate normal from abnormal heartbeats. Previous studies have emphasised feature extraction before feeding into the neural network, or transfer learning where the model is initially trained with an existing data set and

TABLE 4. Confusion Matrix for Fall Detection ECG data set using Attention (Algorithm 2)

		Actual		
		DA	Fall	Rest
Predicted	DA	29	1	0
	Fall	2	49	2
	Rest	0	0	44

later on trained with the same learned weights on the desired data set such as in [4]. In the current benchmark for MI classification using PTB diagnostic, ConvNetQuake neural network model was adapted to achieve an accuracy of 99.44%. Similarly, heavy pre-processing like wavelet transformation [22], data balancing [53], or transfer learning [20] is used in the literature to achieve higher accuracy for ECG signal classification. In our work, no pre-processing to the individual readings were applied and the model achieving 99.66% accuracy, exceeds the state-of-the-art accuracy for normal versus abnormal classification which was previously 99.43%.

Algorithm 2, i.e. the attention /transformer model was also used to model PTB diagnostic to differentiate normal from abnormal heart beats. This yields an accuracy of 99.73%, a precision of 99.73%, a sensitivity of 99.2%, and a specificity of 99.91%.

The confusion matrices of both algorithms are depicted in Table 5 below.

TABLE 5. Confusion Matrices for PTB data set

		Predicted			
		Algorithm 1		Algorithm 2	
		NORMAL	ABNORMAL	NORMAL	ABNORMAL
Actual	NORMAL	371	1	372	1
	ABNORMAL	2	1081	3	1080

In comparison to CNN-LSTM, i.e. Algorithm 1, the attention model with wavelet embedding has shown to be more efficient as it uses fewer parameters and less training time as compared to CNN-LSTM model as shown in Table 6. However, the time and number of parameters for Algorithm 2 might increase eventually with the increase in the number of attention heads and encoder layers.

Multiple metrics were used to evaluate the performance of the model. The terms tp , fp , tn , and fn refer to true positive, false positive, true negative, and false negative respectively. Since in medical terminology, true positive would refer to the medical condition being diagnosed, tp in our context refers to the diagnosis of MI. The performance metrics were

TABLE 6. Comparison between Attention model with wavelet embedding and CNN-LSTM hyper parameters and performance

		Total Parameters	Training Time	Learning Rate	Epochs
HAR	Attention Model	461,583	403.39 s	0.00002	500
	CNN-LSTM	2,471,503	80.56 s	0.001	100
PTB	Attention Model	61762	823.54s	0.001	1000
	CNN-LSTM	210,025	1923.87 s	0.001	1000

calculated using following formulas:

$$Accuracy = (tp + tn) / (tp + fp + fn + tn)$$

$$Precision = tp / (tp + fp)$$

$$Sensitivity = tp / (tp + fn)$$

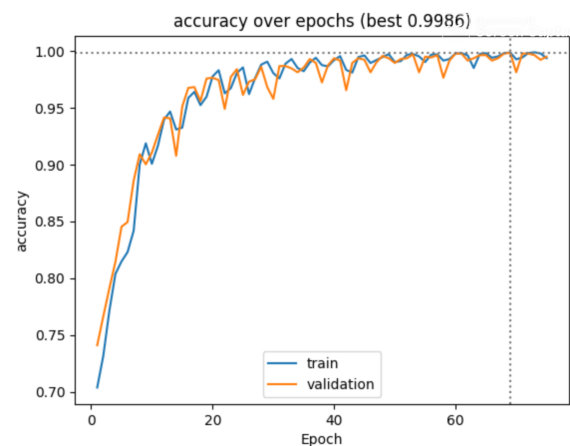
$$Specificity = tn / (tn + fp)$$

The results for our leading models are summarized in Table 7.

TABLE 7. Metrics for leading CNN-LSTM and Attention

		Predicted	
		CNN-LSTM	ATTENTION
Actual	Accuracy	99.79	99.73
	Precision	99.91	99.91
	Sensitivity	99.81	99.72
	Specificity	99.73	99.73

Fig. 8 and Fig. 9 show examples of training accuracy and losses for PTB dataset, respectively.

FIGURE 8. Training and validation graph over epochs for the PTB data set for Algorithm 1

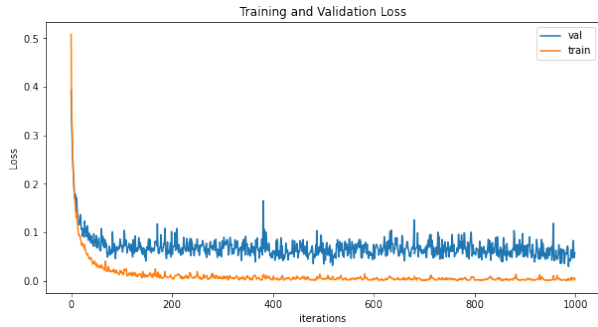


FIGURE 9. Training and Validation Loss for the PTB data set for Algorithm 2

C. PTB XL DIAGNOSTICS

Since PTB XL is a relatively new data set, many recent studies using this data set adapt it for different classification tasks such as super diagnostic, sub diagnostic, form, etc. For our study, 5 super diagnostic (SD) classes were classified. A validation accuracy of 75.70% and a testing accuracy of 74.33% were achieved. AUC score of 0.8395 was achieved, see Table 8.

TABLE 8. Performance indicators

Classification classes	Val. Accuracy	Testing Accuracy	AUC score
Super-diagnostic Classes	75.70%	74.33%	0.8395
MI vs. Normal Class	90.94%	89.1%	0.87
MI vs. Super-diagnostic classes	91.07%	90.2%	0.762

A direct comparison to the state-of-the-art is not very straightforward for PTB-XL mainly because it is a newer data set and many new studies, which use it, focus on different classification tasks. Though our results for this data set do not beat the state-of-the-art accuracy, the results are still comparable. Virginia et al. [55] and Martin et.al [57] achieved an accuracy of 90.8% and 77.12% for MI classification respectively. Similarly, Śmigielet al. [56] achieved an accuracy of 78.0–75.2% for 5 class classification.

Our Algorithm 1 applied for myocardial infarction detection using PTB XL data set yields a validation accuracy of 90.94% when it is MI versus normal class in super diagnostic, and 91.27% when it is MI vs other 4 superclasses. Please note the AUC as a metric was calculated only for PTB-XL data set as it is used widely for comparison of this particular data set in the literature. The experiment with Algorithm 2 on multi-lead data sets such as PTB XL is still in a preliminary stage and requires further investigation on how to merge the natural domain dimension of multi-lead with the dimensional embedding technique. However, similar results for another research project of ours yield promising results for the Algorithm 2 with a multidimensional or multi-featured data

set, see [58]. Henceforth these experiments have not been mentioned and as work in progress will be published in a future contribution.

D. SUMMARY

As seen in Table 9, our model leads in almost all of the evaluation criteria for the classification of PTB data set.

VI. DISCUSSION

As mentioned earlier, state-of-the-art accuracy's are achieved using CNN-LSTM model for three data sets and the attention model for PTB data set. In similar previous works, mostly one set of experiments is performed with a single database to prove the usability of the models. However, we worked on three data sets separately. The first data set is used for the classification of human activities including falls. The second one consists of extracted heartbeats for the classification of MI vs normal heartbeats. The third one consists of a 12-lead ECG data set for multiple cardiovascular conditions. The success of our proposed algorithms on all three data sets generalizes its usefulness for ECG classifications over multiple tasks.

Hybrid models help to combine the features of the base models. This is often more powerful than very deep models with hundreds of layers because for medium-sized data sets, deeper models tend to over-fit. A LSTM model keeps track of the past trends in the time series and can also help in the prediction of the next time stamps. In our study, the results of the CNN-LSTM model have shown to be always better than both of the models implemented individually. This was verified for the HAR data set by [52] and we compare the results from Table 9 for PTB data set where multiple variations of CNN and LSTMs have been applied separately in the previous works. The performance of the model on HAR data set is observed to increase up to a certain level with the increase in a) the number of filters in conv1d layer for CNN-LSTM and b) the number of dimensions in the dimensional embedding with the attention model. Since the data set is not very large, a final conclusion cannot be drawn at this stage but it merits further investigation. The attention algorithm clearly has a computational advantage over CNN-LSTM algorithm as seen in Table 6. It takes less time to converge and even has fewer parameters to train than the CNN-LSTM algorithm. Our work presents the following advantages:

- A hybrid CNN-LSTM model and attention with discrete wavelet transformation as an embedding are proposed.
- No or very little manual feature extraction is required for training the model.
- Three publicly available data sets are used separately for the training using the proposed models.
- State-of-the-art accuracy of 99.86% and 99% is achieved for the PTB dataset and ECG for HAR classification respectively without any feature extraction or pre-processing.

TABLE 9. Our result compared with other similar studies in literature which used PTB database(Built upon [12])

Work	Accuracy(%)	Sensitivity(%)	Specificity(%)	Precision(%)
Acharya et al. [22]	93.5	93.7	-	92.8
Safdarian et al. [31]	94.7	-	-	-
Kojuri et al. [32]	95.6	93.3	-	97.9
Sun et al. [33]	-	92.6	-	82.4
Liu et al. [30]	94.4	-	-	-
Sharma et al. [34]	96	93	-	99
Kachuee et al. [20]	95.9	95.1	-	95.2
Remya et al. [35]	93.61	93.22	94.28	-
Reasat et al. [36]	84.54	85.33	84.09	-
Zewdie et al. [37]	98.3	98.7	96.4	-
Feng et al. [38]	95.4	98.2	86.5	-
Strodthoff et al.	-	93.3	89.7	-
Huang et al.	96.96	99.89	92.51	95.35
Liu et al. [29]	98.59	99.53	94.50	-
Gupta et al. [12]	99.43	99.40	99.45	99.46
Ours (CNN-LSTM)	99.93	99.81	99.73	99.91
Ours (Attention)	99.73	99.72	99.73	99.91

TABLE 10. Overview of the experiments with different data sets and the acquired performances

Classification task	Data set	Achieved Accuracy	Algorithm	State-of-the-art accuracy
Fall detection	ECG HAR	99.21%	Attention	98.44% [8]
Fall detection	ECG HAR	99.21%	CNN-LSTM	98.44% [8]
MI detection	PTB	99.73%	Attention	99.44% [12]
MI detection	PTB	99.93%	CNN-LSTM	99.44% [12]
SD Class	PTB XL	75.70%	CNN-LSTM	-
MI detection	PTB XL	91.07%	CNN-LSTM	-

Hence, we did address our research question and achieved results equivalent to many recent studies without any pre-processing or feature extraction. We have also shown to train the models in an efficient manner computationally.

VII. CONCLUSION

The models proposed and explained in this paper aim to better classify ECG time series for different conditions using minimum pre-processing steps. The availability of publicly available data sets have made it possible to verify the robustness and usefulness of the proposed models via achieving state-of-the-art accuracy using multiple data sets. This would eventually help medical practitioners to identify multiple heart conditions in an automatic manner with minimum feature extraction. Specifically for the MI classification, since the results are close to 100% the model is ready to be deployed for medical evaluation.

VIII. DATA AND CODE AVAILABILITY

The data sets used are publicly available and present in the corresponding repositories:

- ECG HAR dataset : [59]
- PTB DB dataset : [60]
- PTB XL : [61]

The code is available on GitHub at <https://github.com/buttatimasajid/Towards-Automated-Feature-Extraction-For-Deep-Learning-Classification-of-Electrocardiogram-Signals>.

IX. ACKNOWLEDGMENT

The authors thank Kylie Pusch who contributed some ideas in her bachelor thesis [52]. This work was supported in part by the PhD-research program of the Faculty of Computer Science and Engineering Fb2 of Frankfurt University of Applied Sciences.

REFERENCES

- [1] Wagner, P., Strodthoff, N., Bousseljot, R., Kreiseler, D., Lunze, F., Samek, W. & Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7, 154 (2020), <https://doi.org/10.1038/s41597-020-0495-6>
- [2] Centers for Disease Control and Prevention, Underlying Cause of Death, 1999-2020 Request, <https://wonder.cdc.gov/ucd-icd10.html>
- [3] Tan, M. & Kenny, R. Cardiovascular Assessment of Falls in Older People. *Clinical Interventions In Aging*, 1 pp. 57-66 (2006,2)
- [4] Yang, F., Wang, G., Luo, C. & Ding, Z. Improving Automatic Detection of ECG Abnormality with Less Manual Annotations using Siamese Network. 2021 43rd Annual International Conference Of The IEEE Engineering In Medicine Biology Society (EMBC), pp. 1120-1123 (2021)
- [5] Strodthoff, N., Wagner, P., Schaeffter, T. & Samek, W. Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL. *IEEE Journal Of Biomedical And Health Informatics*, 25, 1519-1528 (2021)
- [6] Jambukia, S., Dabhi, V. & Prajapati, H. Classification of ECG signals using machine learning techniques: A survey. 2015 International Conference On Advances In Computer Engineering And Applications. pp. 714-721 (2015)

- [7] Karpagachelvi, S., Arthanari, M. & Sivakumar, M. ECG Feature Extraction Techniques - A Survey Approach. (arXiv,2010), <https://arxiv.org/abs/1005.0957>
- [8] Butt, F., La Blunda, L., Wagner, M., Schäfer, J., Medina-Bulo, I. & Gómez-Ullate, D. Fall Detection from Electrocardiogram (ECG) Signals and Classification by Deep Transfer Learning. *Information*. **12**, 63 (2021,2), <http://dx.doi.org/10.3390/info12020063>
- [9] Kim, Y., Shin, D., Park, M., Lee, S., Jeon, M., Yoon, D. & Park, R. ECG-VIEW II, a freely accessible electrocardiogram database. *PloS One*. **12**, e0176222 (2017), <https://europepmc.org/articles/PMC5402933>
- [10] Kher, R. Wearable Ambulatory Electrocardiogram (ECG) and EEG dataset. (IEEE Dataport,2020), <https://dx.doi.org/10.21227/ysnc-gc65>
- [11] Dallali, A., Kachouri, A. & Samet, M. A Classification of Cardiac Arrhythmia Using WT, HRV, and Fuzzy C-Means Clustering. *Signal Processing: An International Journal (SPIJ)*. **Volume (5)** pp. 101-108 (2011,1)
- [12] Gupta, A., Huerta, E., Zhao, Z. & Moussa, I. Deep Learning for Cardiologist-Level Myocardial Infarction Detection in Electrocardiograms. 8th European Medical And Biological Engineering Conference. pp. 341-355 (2021)
- [13] Khorrami, H. & Moavenian, M. A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. *Expert Syst. Appl.* **37** pp. 5751-5757 (2010,8)
- [14] Sathyapriya, L., Murali, L. & Manigandan, T. Analysis and detection R-peak detection using Modified Pan-Tompkins algorithm. 2014 IEEE International Conference On Advanced Communications, Control And Computing Technologies. pp. 483-487 (2014)
- [15] Khan, A., Sohail, A., Zahoor, U. & Qureshi, A. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*. **53**, 5455-5516 (2020), <https://doi.org/10.1007/s10462-020-09825-6>
- [16] Albawi, S., Mohammed, T. & Al-Zawi, S. Understanding of a convolutional neural network. 2017 International Conference On Engineering And Technology (ICET). pp. 1-6 (2017)
- [17] Blunda, L., Gutiérrez-Madroñal, L., Wagner, M. & Medina-Bulo, I. A Wearable Fall Detection System Based on Body Area Networks. *IEEE Access*. **8** pp. 193060-193074 (2020)
- [18] Ebrahimi, Z., Loni, M., Daneshmand, M. & Gharehbaghi, A. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems With Applications: X*. **7** pp. 100033 (2020), <https://www.sciencedirect.com/science/article/pii/S2590188520300123>
- [19] Cui, Z., Chen, W. & Chen, Y. Multi-Scale Convolutional Neural Networks for Time Series Classification. *ArXiv*. **abs/1603.06995**(2016)
- [20] Kachuee, M., Fazeli, S. & Sarrafzadeh, M. ECG Heartbeat Classification: A Deep Transferable Representation. 2018 IEEE International Conference On Healthcare Informatics (ICHI). pp. 443-444 (2018)
- [21] Wang, T., Lu, C., Sun, Y., Yang, M., Liu, C. & Ou, C. Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network. *Entropy*. **23** (2021), <https://www.mdpi.com/1099-4300/23/1/119>
- [22] Acharya, U., Fujita, H., Oh, S., Hagiwara, Y., Tan, J. & Adam, M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*. **415-416** pp. 190-198 (2017), <https://www.sciencedirect.com/science/article/pii/S020025517308009>
- [23] Strodthoff, N. & Strodthoff, C. Detecting and interpreting myocardial infarction using fully convolutional neural networks. *Physiological Measurement*. **40**, 015001 (2019,1), <http://dx.doi.org/10.1088/1361-6579/aaf34d>
- [24] Zheng Y., Liu Q., Chen E., Ge Y., Zhao J.L. (2014) Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks. In: Li F., Li G., Hwang S., Yao B., Zhang Z. (eds) *Web-Age Information Management. WAIM 2014. Lecture Notes in Computer Science*, vol 8485. Springer, Cham. https://doi.org/10.1007/978-3-319-08010-9_33
- [25] K. Xia, J. Huang and H. Wang. "LSTM-CNN Architecture for Human Activity Recognition," in *IEEE Access*, vol. 8, pp. 56855-56866, 2020, doi: 10.1109/ACCESS.2020.2982225.
- [26] Ordóñez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 2016, **16**, 115. <https://doi.org/10.3390/s16010115>
- [27] S. Saadatnejad, M. Oveis and M. Hashemi, "LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 515-523, Feb. 2020, doi: 10.1109/JBHI.2019.2911367.
- [28] Socher, R., Huval, B., Bath, B., Manning, C. & Ng, A. Convolutional-Recursive Deep Learning for 3D Object Classification. *Advances In Neural Information Processing Systems*. **25** (2012), <https://proceedings.neurips.cc/paper/2012/file/3eae62bba9ddf64f69d49dc48e2dd214-Paper.pdf>
- [29] Liu, N., Wang, L., Chang, Q., Xing, Y. & Zhou, X. A Simple and Effective Method for Detecting Myocardial Infarction Based on Deep Convolutional Neural Network. *Journal Of Medical Imaging And Health Informatics*. **8** pp. 1508-1512 (2018,9)
- [30] Liu, B., Liu, J., Wang, G., Huang, K., Li, F., Zheng, Y., Luo, Y. & Zhou, F. A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection. *Computers In Biology And Medicine*. **61** (2014,8)
- [31] Safdarian, N., Dabanloo, N. and Attarodi, G., A New Pattern Recognition Method for Detection and Localization of Myocardial Infarction Using T-Wave Integral and Total Integral as Extracted Features from One Cycle of ECG Signal. (2014) *Journal of Biomedical Science and Engineering*, **7**, 818-824. doi: 10.4236/jbise.2014.710081.
- [32] Kojuri, J., Boostani, R., Dehghani, P., Nowroozpour, F. & Saki, N. Prediction of acute myocardial infarction with artificial neural networks in patients with nondiagnostic electrocardiogram. *Journal Of Cardiovascular Disease Research*. **6** pp. 51-59 (2015,5)
- [33] Sun, L., Lu, Y., Yang, K. & Li, S. ECG Analysis Using Multiple Instance Learning for Myocardial Infarction Detection. *IEEE Transactions On Biomedical Engineering*. **59**, 3348-3356 (2012)
- [34] Sharma, L. & Sunkaria, R. Inferior myocardial infarction detection using stationary wavelet transform and machine learning approach. *Signal, Image And Video Processing*. **12** pp. 199-206 (2018)
- [35] Remya, R., Indiradevi, K. & Babu, K. Classification of Myocardial Infarction Using Multi Resolution Wavelet Analysis of ECG. *Procedia Technology*. **24** pp. 949-956 (2016,12)
- [36] Tahsin, R. & Shahnaz, C. Detection of Inferior Myocardial Infarction using Shallow Convolutional Neural Networks. (2017,12)
- [37] Zewdie, G. & Xiong, M. Fully Automated Myocardial Infarction Classification using Ordinary Differential Equations. <https://arxiv.org/abs/1410.6984> (2014)
- [38] Feng, K., Pi, X., Liu, H. & Sun, K. Myocardial Infarction Classification Based on Convolutional Neural Network and Recurrent Neural Network. *Applied Sciences*. **9** pp. 1879 (2019,5)
- [39] Ingale, M., Cordeiro, R., Thentu, S., Park, Y. & Karimian, N. ECG Biometric Authentication: A Comparative Analysis. *IEEE Access*. **8** pp. 117853-117866 (2020)
- [40] Ramsauer, H., Schäfl, B., Lehner, J., Seidl, P., Widrich, M., Adler, T., Gruber, L., Holzleitner, M., Pavlović, M., Sandve, G., Greiff, V., Kreil, D., Kopp, M., Klambauer, G., Brandstetter, J. & Hochreiter, S. Hopfield Networks is All You Need. (2021)
- [41] Shih, S., Sun, F. & Lee, H. Temporal pattern attention for multivariate time series forecasting. *Mach. Learn.* **108**, 1421-1441 (2019), <http://dblp.uni-trier.de/db/journals/ml/ml108.html#ShihSL19>
- [42] Song, H., Rajan, D., Thiagarajan, J. & Spanias, A. Attend and diagnose: Time series analysis using attention models. 32nd AAAI Conference On Artificial Intelligence, AAAI 2018. pp. 4091-4098 (2018)
- [43] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł. & Polosukhin, I. Attention is All you Need. *Advances In Neural Information Processing Systems*. **30** (2017), <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fd053c1c4a845aa-Paper.pdf>
- [44] Wang, S., Li, B., Khabsa, M., Fang, H. & Ma, H. Linformer: Self-Attention with Linear Complexity. (arXiv,2020), <https://arxiv.org/abs/2006.04768>
- [45] Lai, G., Chang, W., Yang, Y. & Liu, H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. *CoRR*. **abs/1703.07015** (2017), <http://arxiv.org/abs/1703.07015>
- [46] Rabe, M. & Staats, C. Self-attention Does Not Need $O(n^2)$ Memory. (2021)
- [47] Daubechies, I. *Ten Lectures on Wavelets*. (Society for Industrial,1992)
- [48] Mallat, S. & Zhong, S. Characterization of signals from multiscale edges. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. **14**, 710-732 (1992)
- [49] Ryan, Ø. *Linear Algebra, Signal Processing, and Wavelets - A Unified Approach: Python Version*. (2019,1)
- [50] Kingma, D. & Ba, J. Adam: A Method for Stochastic Optimization. (2014), <http://arxiv.org/abs/1412.6980>, cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015
- [51] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* **9**, 8 (November 15, 1997), 1735–1780. DOI:<https://doi.org/10.1162/neco.1997.9.8.1735>

- [52] Pusch, K. ECG Classification Using Different Machine Learning Models for Human Activity Recognition, Bachelor Thesis, Frankfurt University of Applied Sciences, 2021
- [53] Rai, H. & Chatterjee, K. Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data. *Applied Intelligence*. **52**, 5366-5384 (2022), <https://doi.org/10.1007/s10489-021-02696-6>
- [54] Wang, J., Qiao, X., Liu, C., Wang, X., Liu, Y., Yao, L. & Zhang, H. Automated ECG classification using a non-local convolutional block attention module. *Computer Methods And Programs In Biomedicine*. **203** pp. 106006 (2021), <https://www.sciencedirect.com/science/article/pii/S01692607210081X>
- [55] S., C. & E., R. A Novel Deep Learning based Gated Recurrent Unit with Extreme Learning Machine for Electrocardiogram (ECG) Signal Recognition. *Biomedical Signal Processing And Control*. **68** pp. 102779 (2021), <https://www.sciencedirect.com/science/article/pii/S1746809421003761>
- [56] Smigiel, S., Pałczyński, K. & Ledziński, D. Deep Learning Techniques in the Classification of ECG Signals Using R-Peak Detection Based on the PTB-XL Dataset. *Sensors*. **21** (2021), <https://www.mdpi.com/1424-8220/21/24/8174>
- [57] Martin, H., Morar, U., Izquierdo, W., Cabrerizo, M., Cabrera, A. & Adjouadi, M. Real-time frequency-independent single-Lead and single-beat myocardial infarction detection. *Artificial Intelligence In Medicine*. **121** pp. 102179 (2021), <https://www.sciencedirect.com/science/article/pii/S093336572100172X>
- [58] Jörg Schäfer, Human Activity Recognition With CSI Data – Attention Is All You Need, pre-print, Frankfurt University of Applied Sciences, 2022
- [59] Butt, F., La Blunda, L., Wagner, M., Schäfer, J., Medina-Bulo, I. & Oteiza, D. ECG data for deep transfer learning. (IEEE Dataport, 2020), <https://dx.doi.org/10.3390/info12020063>
- [60] Bousseljot R, Kreiseler D, Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik, Band 40, Ergänzungsband 1* (1995) S 317 <https://doi.org/10.13026/C28C71>
- [61] Wagner, P., Strodthoff, N., Bousseljot, R., Samek, W., and Schaeffter, T., PTB-XL, a large publicly available electrocardiography dataset (version 1.0.0). *PhysioNet*, (2020), <https://doi.org/10.13026/qmgm-0d46>



FATIMA SAJID BUTT is a doctoral candidate for Engineering Informatics at the University of Cádiz, Spain and a research assistant at Frankfurt University of Applied Sciences, Frankfurt, Germany. She received her Bachelors in Information Technology from University of the Punjab, Lahore, Pakistan and her Masters in High Integrity Systems (HIS) from Frankfurt University of Applied Sciences, Germany in 2010 and 2019 respectively. She is a part of the research group Industrial

Data Sciences (INDAS) along with Jörg Schäfer, Matthias Wagner and Dirk Stegelmeyer at Frankfurt University of Applied Sciences, Germany. Her research interests includes time series analysis for classification and application of machine learning algorithms for industrial problems such as predictive maintenance .



MATTHIAS WAGNER received a Diploma and a Dr. rer.nat. in Physics from the Johannes Gutenberg - Universität Mainz (Germany). He was head of Measuring Technology Software Development at Hottinger Baldwin Messtechnik (HBM) in Darmstadt (Germany) from 1990 until 2002. In 2002, he was appointed as Professor of Computer Science at the Frankfurt University of Applied Sciences in Frankfurt am Main (Germany). Since 2005, he has been the Program Director of the international M.Sc. program "High Integrity Systems". From 2017 to March 2020, he served as Vice-Dean for Research and International Relations of the FB2, Department of Computer Science and Engineering. Since 2010, he has been head of the Research Group Wireless Sensor Networks and Internet of Things (WSN & IoT). His research interests cover safety critical computer systems, smart sensors and actuator networks. Software and Systems Engineering and Computational Science are supported by research stays at the UCASE Software Engineering Research Group of the Universidad de Cádiz (Spain) and the Dipartimento di Fisica e Astronomia of the Università degli Studi di Firenze (Italy).



JÖRG SCHÄFER (M'12-22) is a professor of Computer Science at Frankfurt University of Applied Sciences. He received his PhD degree from Bochum University (Germany) in Mathematical Physics in 1992. After spending more than 10 years working as a principal architect in IT consulting for large international companies he was appointed as full professor of Computer Science at Frankfurt University of Applied Sciences in 2009. Since 2012 he served as the chairman of the

computer science B.Sc. program. His main research interest is in theoretical understanding of deep learning architectures and applying machine learning and probabilistic models in ubiquitous computing applications. He runs a research group on human activity recognition (HAR) and channel state information (CSI) and jointly with Matthias Wagner and Dirk Stegelmeyer the research group industrial data science (INDAS). He is collaborating with UCASE Software Engineering Research Group of the Universidad de Cádiz (Spain).



DAVID GÓMEZ-ULLATE received his PhD in Physics from Universidad Complutense de Madrid in 2001. He is a Distinguished Researcher at University of Cádiz, where he founded and serves as Director of UCA Datalab. He is Professor of Applied Mathematics on leave from Complutense University of Madrid, Adjunct Professor at IE Business School and Visiting Professor of Mathematics and Data Science at the University of Loughborough (United Kingdom). His research interests span a wide range of topics, ranging from mathematical physics and approximation theory to applied machine learning and data science. His major contribution is the theory of exceptional orthogonal polynomials, which has earned him international recognition expressed as plenary talks in the main conferences of the field, and invited seminars at Cambridge, Harvard, Edinburgh, Stockholm, Copenhagen and Rome, among others. For the past 10 years he has specialized in knowledge transfer of mathematics to industry. He is currently President of the Knowledge Transfer Commission of the Royal Spanish Mathematical Society, and a member of its Governing Board. In 2016 he received a Leonardo Scholarship from the BBVA Foundation for a project on credit card fraud detection. He has coordinated 8 knowledge transfer contracts with industry in the financial, fisheries, biomedical, legaltech and aeronautical sectors. He is co-founder and scientific advisor of the technological startup Komorebi AI.

• • •