DOI: https://doi.org/10.15276/hait.08.2025.7
UDC 004.8

Representation-based ECG signal prediction for neural networks pre-training

Serhii G. Stavychenko[1)]
ORCID: https://orcid.org/0009-0007-4329-292X; sergey.stavichenko@gmail.com
Anna E. Filatova[1)]
ORCID: https://orcid.org/0000-0003-1982-2322; Hanna. Filatova@khpi.edu.ua
[1)] National Technical University "Kharkiv Polytechnic Institute", 2, Kyrpychova Str. Kharkiv, 61002, Ukraine

ABSTRACT

A limited amount of training data is a well-known challenge in the application of deep learning methods. This issue is particularly relevant in biomedical signal processing, such as the analysis of electrocardiograms, due to the labor-intensive nature of data preparation, which requires the involvement of qualified specialists. Self-supervised learning methods, originally developed in such domains as natural language processing and computer vision, have emerged as a potential approach to addressing this challenge and are increasingly being explored in biomedical signal processing. However, direct adaptation of self-supervised learning techniques from other domains does not fully account for ECG-specific characteristics, such as quasi-periodicity, localized morphological features, and susceptibility to noise. This highlights the relevance of developing ECG-specific self-supervised learning methods. This study presents a novel self-supervised learning approach for pretraining neural networks on unlabeled ECG data. The proposed method is based on predicting the short consecutive signal segment using a preceding one and a learned representation vector. The representation extraction and prediction models are trained jointly on the MIMIC-ECG-IV dataset using backpropagation to minimize the mean squared error between the predicted and original signal segments. As an example of a downstream task, a linear binary classifier was trained on the PTB-XL dataset to diagnose pathological conditions using Lead I. The number of training examples for each diagnosis was limited to thirty-four samples. Firstly, the representation model was pre-trained on the unlabeled MIMIC-ECG-IV dataset, and then linear classifiers were trained on the learned representations for each selected diagnosis in PTB-XL. A comparison was also conducted with a randomly initialized representation model trained jointly with the classifier in a fully supervised manner. The proposed method was evaluated against adaptations of Contrastive Learning, Contrastive Predictive Coding, and Masked Autoencoders method. To ensure a controlled experimental setup, implementations of all considered methods were developed using a unified codebase and shared architectural components. Experimental results demonstrated a significant advantage of all self-supervised learning approaches over joint training of feature extraction and classification models. The proposed SSL method outperformed other tested approaches, particularly for diagnoses with subtle short-term morphological features, such as atrial fibrillation and flutter. These findings suggest the potential for further research in developing ECG-specific self-supervised learning methods as a promising approach to improving neural network performance in scenarios with limited labeled data.

Keywords: Biomedical signals; electrocardiogram; deep learning; self-supervised learning; representation vector; signal prediction

For citation: **Stavychenko S. G., Filatova A. E. "Representation-based ECG signal prediction for neural networks pre-training". *Herald of Advanced Information Technology*. 2025; Vol.8 No.1: 100–116. DOI: https://doi.org/10.15276/hait.08.2025.7**

1. INTRODUCTION

According to the last World Health Organization report published on 7 August 2024, cardiovascular diseases remain the leading cause of death globally. Automatic ECG analysis plays a crucial role in the early detection, diagnosis, and monitoring of cardiovascular diseases. With the growing amount of ECG data generated by wearable devices, clinical environments, and research studies, there is a pressing need for efficient, scalable, and reliable methods to analyze these signals automatically.

Automating ECG analysis methods began to be developed in the last century. One of the first attempts to use computers for ECG signal analysis started in 1957 in the laboratory of Dr. Hubert V. Pipberger [1]. In the 1960s Dr. Hubert V. Pipberger, with his colleagues, introduced a pilot facility for digitizing and storing ECG leads data on magnetic tape [2]. The method to find boundaries for PQRST waves, based on voltage change thresholding, was introduced later [3]. Automating ECG analysis became a popular topic for many organizations and researchers, and different algorithms and approaches were developed. Traditional algorithms are usually based on recognizing ECG segments (also known as ECG segments' annotation) with a consequent classification of their features. ECG segment annotation can be performed in various ways, in [4] authors performed a detailed comparison of 9 different segment annotation

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

algorithms, such as Adaptive Thresholding, Multi-scale Morphological Derivate, Discrete Wavelet Transform based methods, and others. Considering the non-stationarity and noisiness of biomedical signals, interpolation techniques, such as the interpolation of raw signal with cubic splines [5, 6] can improve the accuracy of analysis algorithms. Different approaches were studied for segment classification, such as Decision tree algorithms [7, 8], statistical classifiers [9], and others. Another important aspect of ECG analysis automatization is the transformation of the original signal into its representations suitable for consequent processing. Famous Ukrainian scientist Leonid Fainzilberg proposed to process ECG signal $z(t)$ on a phase plane $\{z(t), z'(t)\}$ composed of the signal value and its first derivative [10] and shows the advantages of such representation for human interpretation. The Ministry of Health of Ukraine later approved a Phasegraphy ECG analysis method as a recommended method for a medical screening practice. Another interesting approach is to encode the signal into a sequence of codegrams (or words) that encode the changes in selected heartbeat parameters making it possible to process such sequences with mathematical linguistic methods [11]. In [12] authors proposed an algorithm of 12 ECG leads aggregation into a single integral signal invariant to the electrical heart axis of a patient, providing standardized heart activity signal representation.

The integration of machine learning into ECG analysis has shown great promise, significantly improving accuracy and diagnostic capability compared to traditional methods. In 2017 Andrew Ng with his team showed that artificial neural networks could outperform cardiologists accuracy of cardiac rhythm classification [13] by comparing results obtained from 6 individual cardiologists and trained neural network with ground truth diagnoses obtained from board-certified cardiologists. An international scientific conference computing in Cardiology conducted several challenges for ECG analysis tasks. The latest challenges – “Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020” [14], and “Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021” [15] showed an unconditional advantage of deep learning methods according to the official results. However, the success of deep learning models is highly dependent on the availability of large, labeled datasets, which presents a major challenge in the medical field. Acquiring labeled ECG data is a time-consuming and resource-intensive process. Labeling medical data requires the expertise of certified specialists, to ensure accuracy and reliability. This is further complicated by the sensitive nature of medical data and the strict regulations that govern its accessibility and sharing.

Self-supervised learning (SSL) methods offer a promising solution to the challenge of limited labeled data [16]. Self-supervised learning enables models to learn useful representations from large amounts of unlabeled data by solving artificially stated problems with consequent transfer-learning on a target task. These problems often involve predicting certain transformations or parts of the data. This approach can significantly reduce the required amount of labeled data. Over the past few years, SSL has seen rapid evolution, with numerous methods emerging, particularly in computer vision and natural language processing (NLP) domains. A family of contrastive learning methods such as SWaV (Swapping Assignments Between View) [17], SimCLR (A Simple Framework for Contrastive Learning of Visual Representations) [18], CPC (Contrastive Predictive Coding) [19], and BYOL (Bootstrap your own latent) [20] have achieved best results in field of computer vision, while BERT (Bidirectional Encoder Representations from Transformers) [21] and GPT (Generative pre-trained transformers) [22] have revolutionized language models by solving masked tokens prediction SSL problems.

Self-supervised learning methods have also shown great promise in the field of biomedical signals and ECG data in particular. Most recent and successful SSL methods that proved their efficiency in other domains were successfully applied to biomedical signal processing and showed a great performance boost compared to the supervised-only approach. Such contrastive learning methods as SimCLR [23], CPC [24] BYoL [23], SWaV [23, 24], MoCo [25] as well as Masked Autoencoders (MAE) for both Transformers [26] and CNN-based architectures [27] were successfully applied to ECG signals.

Despite the growing interest from the scientific community in applying SSL methods to ECG analysis, most of the research focuses on adapting techniques initially developed for other domains, and relatively little attention has been given to the development of methods specifically designed for biomedical signals, which have some unique characteristics, such sensitivity to noise, local

features concentration. Some signals, like ECG, have an implicit periodic nature, which can be used to state SSL problems.

2. ANALYSIS OF LITERARY DATA AND PROBLEM STATEMENT

The idea to construct a task to train neural networks on the data without manual labeling originated from early works in Natural Language Processing with the prediction of one or multiple words based on the given context words. For example, in [28] in 2013 authors proposed such tasks as a prediction of the masked middle word in a sequence of 9 consequent words and a prediction of surrounding words probabilities given by the input word. This idea remains dominant in NLP till today, such methods as BERT and GPT are trained to predict certain words by a given sequence of context words. In computer vision, early SSL approaches focused on restoring corrupted samples or predicting the characteristics of the corruption, like solving a Jigsaw puzzle [29], colorizing image converted to grayscale color space [30], predicting the relative positions of image patches within a larger image [31], or recognition of image rotation applied [32]. However, the prediction of missed parts of an input, which was widely used in NLP, became dominant later. One of the methods that showed great performance in different domains is Masked Autoencoders (MAE). This method has several applications in ECG data. In [27] authors used a novel CNN-based MAE architecture proposed by the Meta (Formerly Facebook) research group [33]. A traditional transformers-based MAE pipeline was also applied to ECG [26]. The authors reported promising results and showed that the MAE approach fits well for the ECG processing domain. However, learned representations may contain redundant information, such as time shifts relative to the start of the heartbeat cycle. For instance, different samples taken from the same ECG signal with a slight temporal shift should have distinct representations to be correctly reconstructed, even though they represent the same underlying signal. We hypothesize that this redundancy may negatively impact the quality of the learned representations, making them less effective for subsequent classification tasks. Another valuable group of SSL methods first developed in the computer vision domain is a family of contrastive methods. The essence of contrastive methods is to learn an image representation, also called latent space embedding, sufficient to distinguish augmented versions of the same image from different ones. The SSL task is usually formulated to have distinct embeddings for different input samples and as close as possible for the same. Different approaches to state the problem have been developed founding a family of contrastive learning methods. Contrastive methods are widely applied in the ECG processing domain. In [23] authors tested SimCLR [18], BYOL [20], and SWaV [17] contrastive learning frameworks and showed that contrastive learning pre-training yields better results on a target task compared to fully supervised learning with random weights initialization. However, according to the reported results, the detection of diagnoses that require analysis of short-term signal features, such as for atrial fibrillation has relatively bad performance. We hypothesize that it can be explained by the fact that contrastive methods don't tend to preserve all the information about the signal in its latent space, and local short-term features of a signal shape may be lost. In [24] another comparison of contrastive and the novel adaptation of Contrastive Predictive Coding (CPC) [19] was performed. The authors showed that CPC outperforms contrastive methods in terms of target task accuracy methods. The reason behind this can be explained by the fact that in CPC the training pipeline requires the prediction of tokens obtained from small local chunks of a signal sample, rather than the entire sample embeddings, which may preserve the short-term features information in latent space. However, the final signal representation can also vary on the sample's time shift from the heartbeat cycle start as in the MAE approach, which is not the case for the traditional contrastive methods. Early computer vision techniques like recognition of transformation applied to the input samples are also studied. In [34] authors utilize the fact that ECG signal has a certain shape and use prediction of the artificial spatial and temporal reverse of signal prediction and show that such an approach can outperform a popular SimCLR method in a certain setting. In [35] authors used the same idea but with additional transformations like segment permutations, time wrapping and others for pre-training with consequent emotion recognition.

A literature review showed that pre-training of deep learning models with SSL methods has great potential in improving model performance on target tasks. Most studies are focused on the adaptation of methods initially developed for different domains.

We also highlighted several potential issues with the application of common SSL methods developed for other domains to ECG data:

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

1) potential loss of small locally concentrated features in signal representation;

2) variance in representations of the same signal embeddings.

Therefore, developing the SSL methods considering the accumulated empirical knowledge in other domains and the unique features of ECG signal data seems to be an actual research problem.

3. GOAL AND RESEARCH OBJECTIVES

The purpose of the research. The purpose of the research is to improve the efficiency of machine learning methods for ECG analysis under limited training data conditions.

The objectives of the research are:

1. Analysis of the state of the art in self-supervised learning methods and their application for the ECG analysis domain.

2. Development of a domain-specific Self-Supervised Learning method for ECG analysis model pre-training.

3. Experimental evaluation of the method and analysis of the results.

The object of research is the automatic analysis of ECG using computers.

The subject of research is the application of SSL methods for ECG analysis.

4. MATERIALS AND METHODS

4.1. SSL task formulation

Let $\vec{X} \in R^m$ be a quasi-periodic discrete ECG signal (lead) with *m* readings, *T* be the average number of readings per ECG signal period, and $\vec{X}^s = (\vec{X}_s, \vec{X}_{s+1}, ..., \vec{X}_{s+k})$ be a certain sample of the signal $\vec{X}$ of length *k*, taken with an offset *s* from the beginning of the signal. Let the function $f: R^k \rightarrow R^h$ transform the sample $\vec{X}^s$ into a representation vector (embedding) $\vec{D}$ of size $h < k$.

$$D = f(\vec{X}^s) \quad (1)$$

Define a sequence of consecutive samples of a signal $\vec{X}$ of length l as:

$$\vec{Y}_t = (\vec{X}_{tl}, \vec{X}_{tl+1}, \ldots, \vec{X}_{tl+l-1}),\ t \in 0, \overline{\left\lfloor \frac{m}{l} \right\rfloor}.$$

Define a function $g: (\mathrm{R}^l, \mathrm{R}^h) \rightarrow \mathrm{R}^l$ such that:

$$\vec{Y}_{t+1} = g(\vec{Y}_t, \vec{D}) \quad (2)$$

where $k > T$, $l < \frac{t}{2}$. Thus, the sample $\vec{X}^s$ contains at least one full period, whereas the elements of the sequence $\vec{Y}_t$ contain less than half a period. Let's hypothesize that it is impossible to predict the element $\vec{Y}_{t+1}$ in the future based solely on the preceding sample $\vec{Y}_t$ of size *l* that does not contain a full period. In other words, there does not exist a function g' such that $\vec{Y}_{t+1} = g'(\vec{Y}_t)$. If functions *f* and *g* exist and satisfy the given conditions for any signal $\vec{X}$, then the embedding $\vec{D}$ possesses invariance with respect to the sample $\vec{X}^s$.

Substituting (1) into (2), we obtain:

$$\vec{Y}_{t+1} = g(\vec{Y}_t, f(\vec{X}^s)). \quad (3)$$

Let $\vec{X}^{s1}$ and $\vec{X}^{s2}$ be two samples of a signal $\vec{X}$ taken with different offsets. Substituting them into (3), we got: $\vec{Y}_{t+1} = g(\vec{Y}_t, f(\vec{X}^{s1}))$ on one side, and $\vec{Y}_{t+1} = g(\vec{Y}_t, f(\vec{X}^{s2}))$ on the other, so:

$$g(\vec{Y}_t, f(\vec{X}^{s1})) = g(\vec{Y}_t, f(\vec{X}^{s2})). \quad (4)$$

The function *g* is injective with respect to its second argument; otherwise, there would exist $\vec{X}^{s1}$ and $\vec{X}^{s2}$ such that:

$$f(\vec{X}^{s1}) \neq f(\vec{X}^{s2}),$$

therefore, to satisfy (4), the value of *g* does not depend on the second argument, reducing equation (2) to $\vec{Y}_{t+1} = g'(\vec{Y}_t)$, and contradicting our hypothesis. If *g* is injective, then from (4) it follows that:

$$f(\vec{X}^{s1}) = f(\vec{X}^{s2}),$$

which means that the signal embedding extraction function is invariant to the sample's offset from the beginning of the recording. The embedding $\vec{D}$ also contains information about the signal shape necessary for predicting the elements of the sequence $\vec{Y}_t$, which holds the signal values at all possible offsets relative to the period start when $m > T + l$.

The *f* and *g* functions approximation with machine learning methods will be considered as a self-supervised learning task for deep learning models pre-training. Let $\vec{X}_i \in R^m$ be a recording in the training dataset. For each recording, we take

three subsamples: $\vec{X}_i^s$ of length k and two consecutive subsamples $\vec{Y}_i^1$ and $\vec{Y}_i^2$ of length l, located in the future relative to $\vec{X}_i^s$ to avoid overlap. Let t_1 and t_2 be discrete uniformly distributed random numbers indicating the offsets from the beginning of the recording $\vec{X}_i$ to the first reading of subsamples $\vec{X}_i^s$ and $\vec{Y}_i^1$, respectively:

$$t_1 \in \overline{0, m-2k-1};$$
$$t_2 \in \overline{t_1+1, m-l}.$$

Define the operation of extracting a subsample of readings from a recording x starting at the a-th element with length b as $x[a:b]$, then:

$$\vec{X}_i^s = \vec{X}_i[t_1 : k];$$
$$\vec{Y}_i^1 = \vec{X}_i[t_2 : l];$$
$$\vec{Y}_i^2 = \vec{X}_i[t_2 + l : l].$$

An example of subsamples locations for certain t_1 and t_2 is illustrated in Fig. 1.

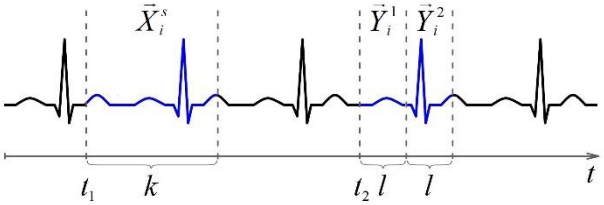

Fig. 1. **Schema of subsamples**
***Source*: compiled by the authors**

Let the models approximating functions f and g be denoted as $F(\vec{X}^s, \omega_f)$ and $G(\vec{Y}_t, \vec{D}, \omega_g)$, where ω_f and ω_g are model parameters.

Then, the models' parameters optimization problem can be formulated as follows:

$$\min_{\omega_f, \omega_g} \sum_i^n E(\vec{Y}_i^2, G(\vec{Y}_i^1, F(\vec{X}_i^s, \omega_f), \omega_g)),$$

where n is the dataset size, and E is the mean absolute error between two signal samples:

$$E(\vec{A}, \vec{B}) = \frac{1}{l} \| \vec{A} - \vec{B} \|_1.$$

The described method uses the quasi-periodicity of the signal to formulate the pre-training task in the form of forecasting the subsequent subsample. It is also sensitive to the localization of diagnostic features due to the formulation of the task, since for successful signal forecasting, the vector representation must contain complete information about the signal shape. The proposed method reduces the dimensionality of the input signal without relying on noise suppression algorithms, which eliminates the possibility of suppressing local features. In this work, the pre-trained model F is used as a signal embedding extractor. Binary classification models a later trained on the extracted embeddings to detect the presence of various pathological conditions.

4.2. Compared methods

The proposed method of ECG signal model pre-training was compared with several broadly researched in recent times self-supervised learning approaches.

1. Contrastive Learning with Triplets Loss Function – Selected as a broad representation of contrastive methods such as SimCLR, SwaV, BYOL, and others which are based on embeddings comparison in latent space.

2. Contrastive Predictive Coding (CPC) – Selected for its ability to avoid direct signal comparison, focusing on token-based embeddings, potentially making it more robust to signal details than contrastive approaches.

3. Masked Autoencoders (MAE) – A leading method in large language models (LLMs) applied to different domains. Successfully applied for pre-training on ECG data in different studies [26, 27].

4.3. Implementation details

The two most common types of neural networks used today are convolutional networks (CNN) and transformer networks. Recently, the Mamba architecture, based on Selective State Space Model (SSM) architecture has been gaining popularity. Albeit SSM is different from transformer models, it operates on sequence rather than on spatial data as CNN does. Thus, it could be said that two principally different machine learning model architecture categories in how data modeling is performed are most promising. Contrastive and CPC SSL approaches were developed for CNN architecture, while MAE was formerly implemented for transformer models. However, a CNN-based SSL framework was introduced recently [33]. ECG adaptation of MAE was also researched [27] and showed quite promising results. Considering the above we decided to use CNN architecture in all implementations used in experiments, to eliminate the NN architecture influence on the results.

To reduce the implementation details impact, we developed custom implementation for every method, following the next principles:

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

- use of common neural network blocks, wherever it is possible, and not contradict the main concepts of a method;
- minimal data pre and post-processing;
- consistent training conditions, i.e. usage of the same weights optimization strategies, same hyperparameters, same batch generation strategies, same record lengths, and so on.

The two main building blocks used across all implementations are ConvBlock and ConvGroup (Fig. 2). ConvBlock is represented with 1D Convolution with kernel size 5, given dilation and filter count, followed by batch normalization and GELU activation function layers. ConvGroup is composed of two parallel branches of three ConvBlocks with different dilations (1 and 5). Input tensor is passed to the convolution layer of size 1 to up-sample channels and then passed to convolution block branches. The first branch is aimed at capturing local signal features, while the second branch can capture more global features due to a bigger dilation value. Then, channels of each branch are concatenated together and merged into a tensor with a target channel count via another convolution layer of size 1. Finally, the result is added to the up-sampled input via a residual connection to improve gradient flow and passed to the max pooling layer with a window of size 2 to reduce size. Every implementation uses encoder models which are composed of 6 sequential ConvGroup blocks with $C_i = (16, 32, 64, 96, 128, 128)$ filter count, each block reduces input size twice due to the pooling layer, so $S_i = S_{i-1} / 2$. The obtained tensor is then passed to the global max pooling layer to get a 1D embedding vector that is l2-normalized, i.e. element-wise divided with its magnitude. Another block that is used in the experimental setup is the decoder model. The decoder model is a mirrored version of the encoder and is composed of 6 sequential convolution groups with a number of number filters equal to reversed C_i. The convolution group in the decoder has the same architecture, except that the trailing pooling layer is removed and the starting channels up-sampling layer is replaced with nearest neighbors up-sampling along the time axis. At the beginning of the decoder model, a fully connected layer with a GELU activation function is used to convert embedding to bottleneck representation with the shape as an input tensor for the global max pooling layer used in the encoder.

Contrastive Triplets. Contrastive triplets implementation uses a single encoder model to convert the three input triplet samples into latent space embeddings. Triplet contains an anchor sample (*a*), a positive sample (*p*) which is taken from the same signal but with a different offset and a negative sample (*n*) taken from a different patient signal recording. Obtained representations are then compared with Triplet Loss (Fig. 3). The embedding obtained from the encoder is later used in a classification task.

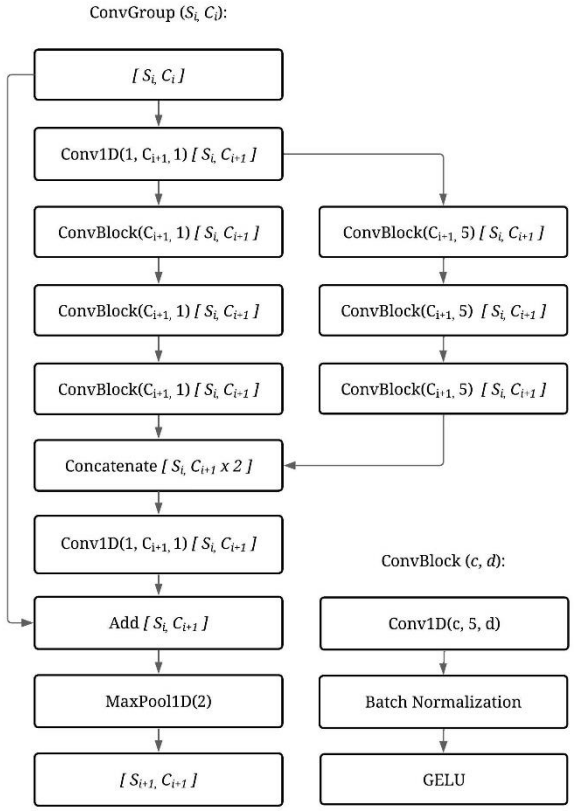

Fig. 2. **Convolution blocks architecture. Conv1D (c, k, d) – represents 1D convolution layers with *c* filters; *k* kernel size and *d* dilation**
***Source*: compiled by the authors**

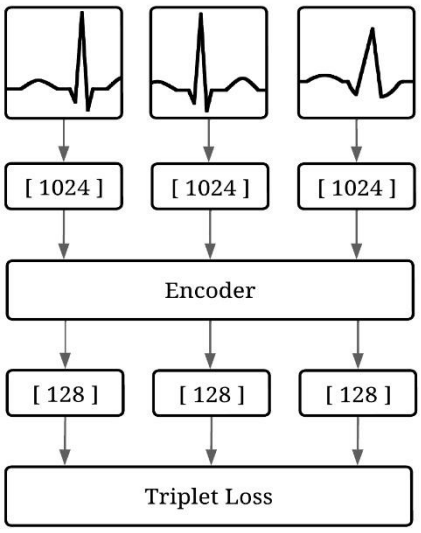

Fig. 3. **Contrastive Triplets architecture**
***Source*: compiled by the authors**

Contrastive Predictive Coding (CPC). CPC implementation uses two consequent samples for the signal as two inputs.

The first input is passed to the Tokenizer network (Fig. 4), which splits it into chunks of size 256 and encodes to the tokens of size 128 via a default encoder model with $S_0 = 128$ and default C_i. Then, encoded tokens are passed to the Gated Recurrent Unit (GRU) layer with 128 hidden units and tanh activation function. The hidden state of the GRU layer is later passed to the Predictor (Fig. 4) – a prediction model implemented as 4 linear fully connected layers each of which predicts one of the next 4 tokens $\hat{z}_i$ where $i \in \{1, 2, 3, 4\}$ an index in the predicted future signal part tokens sequence is.

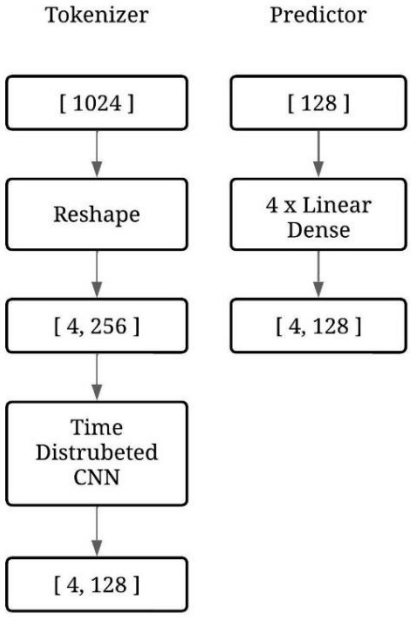

Fig. 4. **Contrastive Predictive coding architecture. Tokenizer and predictor networks**
***Source*: compiled by the authors**

The second input, which represents a future continuation of the signal, is also passed to the same Tokenizer network to obtain a set of z_i which represents a sequence of actual tokens of input sample chunks. Obtained sets of tokens $\hat{z}_i$ and z_i for every signal sample in the SGD mini-batch are finally compared among others in the batch via InfoNCE loss function (Fig. 5).

Masked Autoencoders. According to the result reported in the original ConvNeXt V2 paper [33] regular U-net-like symmetric decoder provides almost similar results to the officially recommended asymmetric one. So, a common encoder with a symmetric decoder is used to keep consistency with other model implementations (Fig. 6).

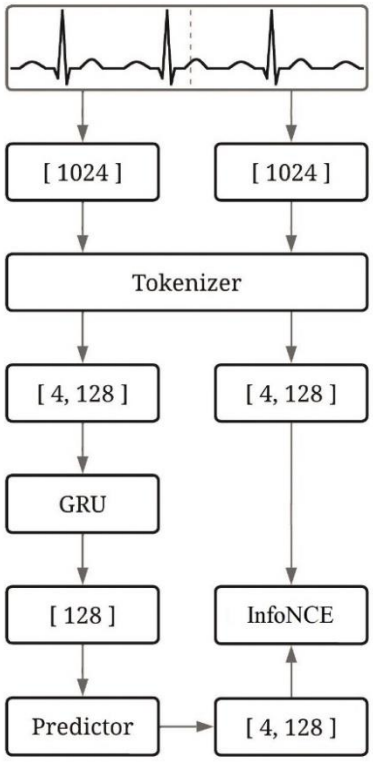

Fig. 5. **Contrastive Predictive Coding architecture. Overall pipeline**
***Source*: compiled by the authors**

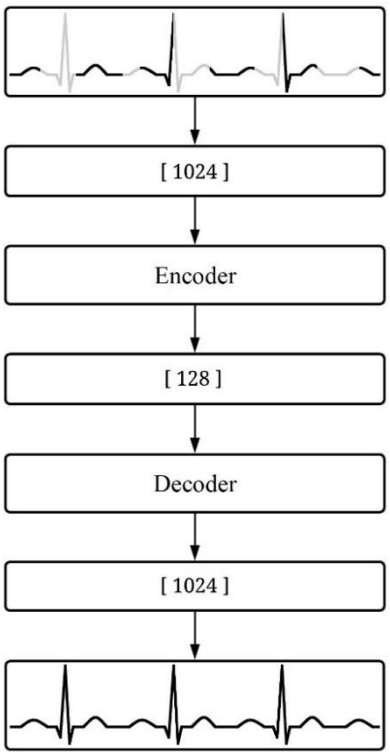

Fig. 6. **Masked Autoencoders pipeline architecture**
***Source*: compiled by the authors**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

The Global Response Normalization (GRN) layer is also omitted since the feature collapse phenomenon reported in the original study [33] wasn't observed in our signal processing domain with 1D convolution.

Different masking rates were evaluated: 60% as suggested in the original ConvNeXt paper; 20% since such a rate corresponds to the seen/predicted ratio used for the proposed method. The next patch sizes were tested:

1) 4 used because it constitutes the same percentage of the total sample data as in the original study;

2) 16 as an empirically selected minimum recognizable chunk of the signal;

3) 32 as an experiment with a larger patch size.

Signal masking is implemented via suppression of signal from masked regions by multiplication of input tensor with a binary mask before and after regular dense convolution, as suggested in [33] as an alternative approach to sparse tensors. Such an approach allowed the usage of the same codebase for all implementations.

The training process consists of the following steps:

1) masked input is passed to the encoder to obtain a signal embedding;

2) embedding is passed to the decoder model to reconstruct the whole signal sample;

3) only sample regions masked in the input are compared with the original sample with mean squared error loss function, with consequent error backpropagation.

Proposed method. The next hyperparameters of the proposed model were used in the reported results:

- total sample length $m = 5000$;
- representation source ($\vec{X}_i^s$) subsample length $k = 1024$;
- short-time subsamples ($\vec{Y}_i^1$, $\vec{Y}_i^2$) length $l = 256$.

The proposed method uses the Descriptor Extractor network (F) to obtain an embedding from a 1024-reading-long sample. The Descriptor Extractor is implemented as a common encoder model. The future sample of length 256 is then passed to the Predictor (G) network (Fig. 7) to predict the next 256 readings long sample. The Predictor uses another encoder model to convert the signal sample into a latent space token, a dense layer to predict the transformation matrix to be applied to a latent token, and a decoder model to convert the transformed token into a future signal subsample.

Let $\omega_g = (\omega_{enc}, \omega_{lin}, \omega_{dec})$ be a tuple of G parameters, then Predictor network can be denoted as:

$$G(Y, D, \omega_g) = Dec(Enc(Y, \omega_{enc})Linear(D, \omega_{lin}), \omega_{dec})$$

where $Enc: R^l \rightarrow R^z$ and $Dec: R^z \rightarrow R^l$ are encoder and decoder networks, and $z = 128$ is token size.

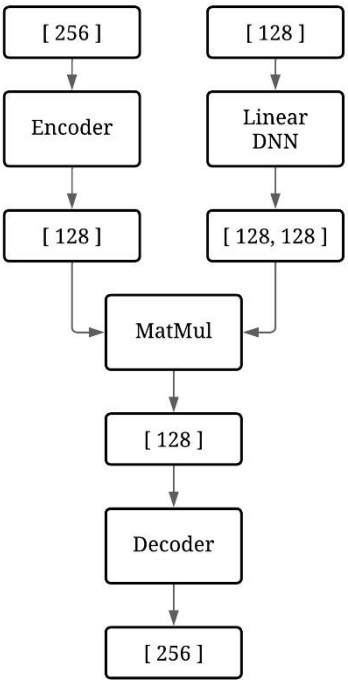

Fig. 7. **Proposed method. Predictor model**

***Source*: compiled by the authors**

Finally predicted real future samples are compared with mean absolute error loss (Fig. 8).

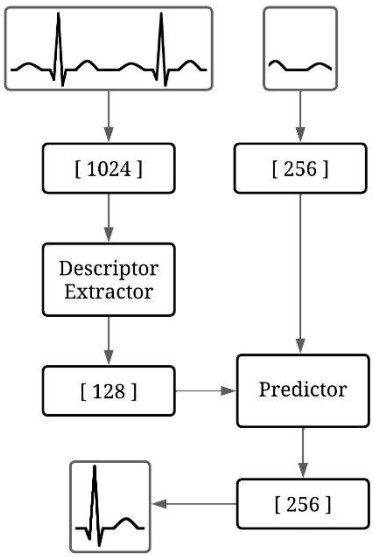

Fig. 8. **Proposed pipeline**

***Source*: compiled by the authors**

4.4. Datasets

Experiments have two phases. During the first phase, an encoder model is trained to learn I lead ECG signal representation in a self-supervised way on a large unlabeled dataset. Later binary linear classification model is trained to classify signal embeddings obtained from encoders on a small-size target-task dataset. The pre-training dataset is obtained from the MIMIC-IV-ECG dataset which consists of approximately 800k ECG records across nearly 160k unique patients. Every record is 10 seconds long and has 5000 discrete readings with a 500 Hz sampling rate. Sub-sample of 1024 readings is randomly taken on each training epoch from every record as the main input for the encoder model. The CPC model also takes the consequent 1024-length sample to compute the loss function. For the MAE model, the length of the input sample is computed to keep the length of the unmasked part of the data equal to 1024 as for other models (2560 values for a 60% masking rate and 1280 for 20%). Input samples are split into chunks (different chunk sizes are evaluated), and a random number (corresponding to the masking rate) of chunks are masked out. For the proposed method implementation two additional consequent sub-samples of length 256 are taken from every signal.

Only the first standard ECG lead is taken for every sample for the pre-training step.

The target dataset is derived from the PTB-XL dataset with 21799 clinical 12-lead ECGs from 18869 patients. PTB-XL dataset has the same discretization parameters as MIMIC-IV-ECG (5000 readings with 500 Hz sampling rate), so no resampling is required. The dataset is labeled with 71 ECG statements conforming to the SCP-ECG standard. Since the I-lead diagnosis case is used in experiments, statements that represent diagnoses recognizable by the shape of the I-lead ECG signal were selected.

According to [36] the next pathologies have distinct signal shape features and can be diagnosed with I ECG lead:

1. Atrial fibrillation (AFIB). P waves are absent and there is an irregular irregularity in R-R interval (Fig. 9).

2. Atrial flutter (AFLT). Low and interchanging ventricular response rate (Fig. 10).

3. Incomplete (ILBBB) and complete (CLBBB) left branch bundle blocks. QRS complex is broad (138 ms as measured on 12 L-ECG). Repolarization matches LBBB with ST depression and discordant T waves. (Fig. 11)

4. Incomplete (IRBBB) and complete (CRBBB) right branch bundle blocks. The total QRS complex is broad (161 ms as measured on 12 L-ECG). The R wave is small, the S wave is broad and the T wave is positive (Fig. 12).

5. Atrioventricular block of first degree (1AVB). The PR interval surpasses 200 ms or 230 ms on 12 L-ECG (Fig. 13).

6. Third-degree AV blocks (3AVB). No correlation between the P waves and the broad QRS complexes (Fig. 14).

7. Premature atrial contraction (PAC). P waves are present before all QRS complexes indicating a sinus base rhythm. The R-R interval, besides the PACs, is regular (Fig. 15).

8. Premature ventricular contraction (PVC). The QRS complex of the PVC is broader and has a different axis, resulting from an electric signal not conducted through the normal conducting system (Fig. 16).

The number of records labeled with each of the listed statements is displayed in Table 1.

Fig. 9. **Atrial fibrillation (AFIB)**
***Source*: compiled by the [36]**

Fig. 10. **Atrial flutter (AFLT)**
***Source*: compiled by the [36]**

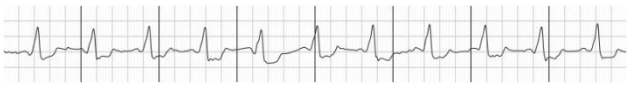

Fig. 11. **Left branch bundle block (ILBBB, CLBBB)**
***Source*: compiled by the [36]**

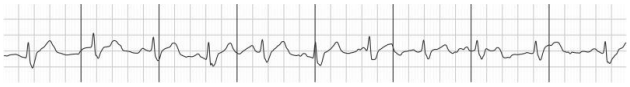

Fig. 12. **Right branch bundle block (IRBBB, CRBBB)**
***Source*: compiled by the [36]**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Fig. 13. **First-degree atrioventricular block (1AVB)**
***Source*: compiled by the [36]**

Fig. 14. **Third-degree atrioventricular block (3AVB)**
***Source*: compiled by the [36]**

Fig. 15. **Premature atrial contraction (PAC)**
***Source*: compiled by the [36]**

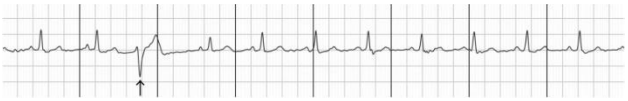

Fig. 16. **Premature ventricular contraction (PAC)**
***Source*: compiled by the [36]**

Table 1. **Number of records labeled with selected statements in PTB-XL dataset**

Statement	Number of records
AFIB	48
AFLT	42
IRBBB	1003
CRBBB	444
ILBBB	61
CLBBB	464
1AVB	767
2AVB	12
3AVB	13
PAC	33
PVC	791

***Source*: compiled by the authors**

A separate dataset is composed for each of the selected statements consisting of records labeled with a given statement and the same amount of randomly selected records without the corresponding label. To keep a class balance, the number of records of each statement is randomly trimmed to 42. Then every record is split into 4 chunks of 1250 discrete readings. 1024 readings of each record with random offset are finally taken for the training set on every epoch to increase data variability. The test set is composed of the same 1024 long records taken with zero offset.

4.5. Training and evaluation algorithm

For the SSL stage, every model is trained in a feed-forward manner with the AdamGrad weights optimization algorithm. 20% of dataset items are taken for validation. The learning rate was initially set to 1e-3 with an iterative decrease by 1e-1 twice on the validation loss plateau. The encoder pre-trained with every tested SSL method is taken with frozen weights and used as a feature extractor to train a linear binary classifier for every target diagnosis. Since the amount of training data for the downstream task is quite small the k-folds cross-validation approach was used with 5 folds, which means that 34 samples are used for training on every iteration. The training process is performed with AdamGrad optimizer with a 1e-3 learning rate. The binary cross-entropy loss function is used as an objective. The training was performed until there was no advance in validation loss value. The best value of the area under the curve (AUC) metric on the validation part is taken from every fold and then averaged as final performance results. The standard deviation between the best AUC values across different folds is also calculated as a measure of training stability. Finally, the means of best AUC and AUC deviation are calculated as integral metrics of method quality.

The invariance of embeddings to the selected source subsample of the same signal is also analyzed with the relative variance measure *V*, which can be calculated as follows. Let $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ be *n* randomly selected signals of size *m* from a certain dataset, and $\vec{X}_{i1}, \vec{X}_{i2}, \ldots, \vec{X}_{iq}$ be *q* samples of size *k* taken with a step *s* from each signal:

$$s = \left\lfloor \frac{m}{q} \right\rfloor;\ \vec{X}_{ij} = \vec{X}_i[js:k];\ j \in \overline{0, q-1},$$

where *X*[*a*:*b*] denotes the operation of extracting a subsample starting from the *a*-th reading with a length of *b*. Let the function for extracting an embedding from a signal sample be denoted as $\hat{f}: R^k \to R^h$, and the embedding for the *j*-th sample of the *i*-th signal as $\vec{D}_{ij} = \hat{f}(\vec{X}_{ij})$.

The standard deviation (STD) of the embeddings for all subsamples of the *i*-th signal and the STD of the embeddings for the *j*-th subsample across all signals are denoted as σ_i^1 and σ_j^2:

$$\bar{D}_i^1 = \frac{1}{q}\sum_j^q \vec{D}_{ij};\ \bar{D}_j^2 = \frac{1}{n}\sum_i^n \vec{D}_{ij};$$

$$\sigma_i^1 = \frac{1}{q}\sum_j^q \| \vec{D}_{ij} - \bar{D}_i^1 \|_2 \; ; \; \sigma_j^2 = \frac{1}{n}\sum_i^n \| \vec{D}_{ij} - \bar{D}_j^2 \|_2 .$$

Then, the measure V of the embeddings invariance to the sample selection for the function $\hat{f}$ is defined as the ratio of mean values for σ_i^1 and σ_j^2 overall signals and samples, respectively:

$$V = \frac{\frac{1}{n}\sum_i^n \sigma_i^1}{\frac{1}{q}\sum_j^q \sigma_j^2},$$

V takes values in the range of $[0,\infty)$ and provides an estimate of the variation in the embeddings within a single signal relative to the variation in the test dataset: $V = 0$ indicates minimal variation of embeddings within a single signal; $V = 1$ indicates that different subsamples of a single signal exhibit a variation comparable to the variation across the entire dataset; $V > 1$ indicates that the variation within a single signal exceeds that of the dataset.

5. RESULTS

All models pre-trained with tested SSL approaches significantly outperform fully supervised training with random weight initialization. The obtained average AUC values and their deviations are represented in Tables 2 and 3. The proposed method achieved the best average AUC (0.9). It also shows a quite low embedding variance *V*, second only to the contrastive approach. CPC showed worse overall performance (0.85), but it slightly outperformed the proposed method in incomplete right branch bundle block (IRBBB) diagnosis (0.66 vs 0.64). The contrastive approach has relatively poor performance at atrial fibrillation and flutter diagnoses. However contrastive approach showed the best result in incomplete and complete right branch bundle blocks (IRBBB, CRBBB). It also showed the lowest *V* value. MAE with optimal hyperparameters (16 readings patch size and 60% masking ratio) showed the best results at the right brunch bundle blocks diagnosis, however, it struggles with long-range dependent diagnoses like premature atrial and ventricular contraction (PAC, PVC).

As was expected, the embedding variance *V* is relatively high for both MAE and CPC compared to proposed and contrastive methods. The *V* values for every method are reported in Table 4. The principal components analysis (PCA) algorithm was applied to 10 samples per 10 random signals' embeddings to visualize it on a 2D plot (Fig. 17). Despite the success on the target task, the proposed model achieved ambiguous results on the SSL task. The best Mean Absolute Error of 0.0375 was achieved, which is less than for another reconstruction-based MAE method (0.032).

Table 2. **The per-fold mean of AUC values for different approaches and diagnoses**

Method	AFIB	AFLT	IRBB	CRBB	ILBB	CLBB	1AVB	2AVB	3AVB	PAC	PVC	Avg
Rnd	0.65	0.81	0.6	0.73	0.67	0.66	0.6	0.53	0.67	0.56	0.54	0.64
Contrasrtive	0.88	0.91	0.68	0.99	0.89	1	0.81	0.72	0.98	0.61	0.7	0.83
CPC	0.96	0.95	0.66	0.94	0.84	0.99	0.85	0.8	0.92	0.72	0.71	0.85
Proposed	0.99	0.99	0.64	0.97	0.91	1	0.97	0.89	1	0.75	0.75	0.90
MAE	0.96	0.95	0.71	0.98	0.84	1	0.9	0.76	0.98	0.64	0.7	0.86

***Source*: compiled by the authors**

Table 3. **The per-fold standard deviation of AUC values for different approaches and diagnoses**

Method	AFIB	AFLT	IRBB	CRBB	ILBB	CLBB	1AVB	2AVB	3AVB	PAC	PVC	Avg
Rnd	0.05	0.1	0.09	0.17	0.09	0.17	0.08	0.03	0.17	0.06	0.09	0.10
Contrasrtive	0.06	0.06	0.13	0.02	0.09	0.01	0.05	0.21	0.05	0.08	0.04	0.07
CPC	0.03	0.04	0.13	0.04	0.1	0	0.04	0.18	0.05	0.12	0.04	0.07
Proposed	0.02	0.01	0.13	0.06	0.06	0.01	0.03	0.18	0	0.05	0.05	0.05
MAE	0.02	0.04	0.11	0.03	0.11	0	0.03	0.2	0.02	0.07	0.05	0.06

***Source*: compiled by the authors**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Table 4. **The embeddings variance (*V*) for different methods**

Method	Contrastive	CPC	MAE	Proposed
V	0.16	0.5	0.93	0.32

***Source*: compiled by the authors**

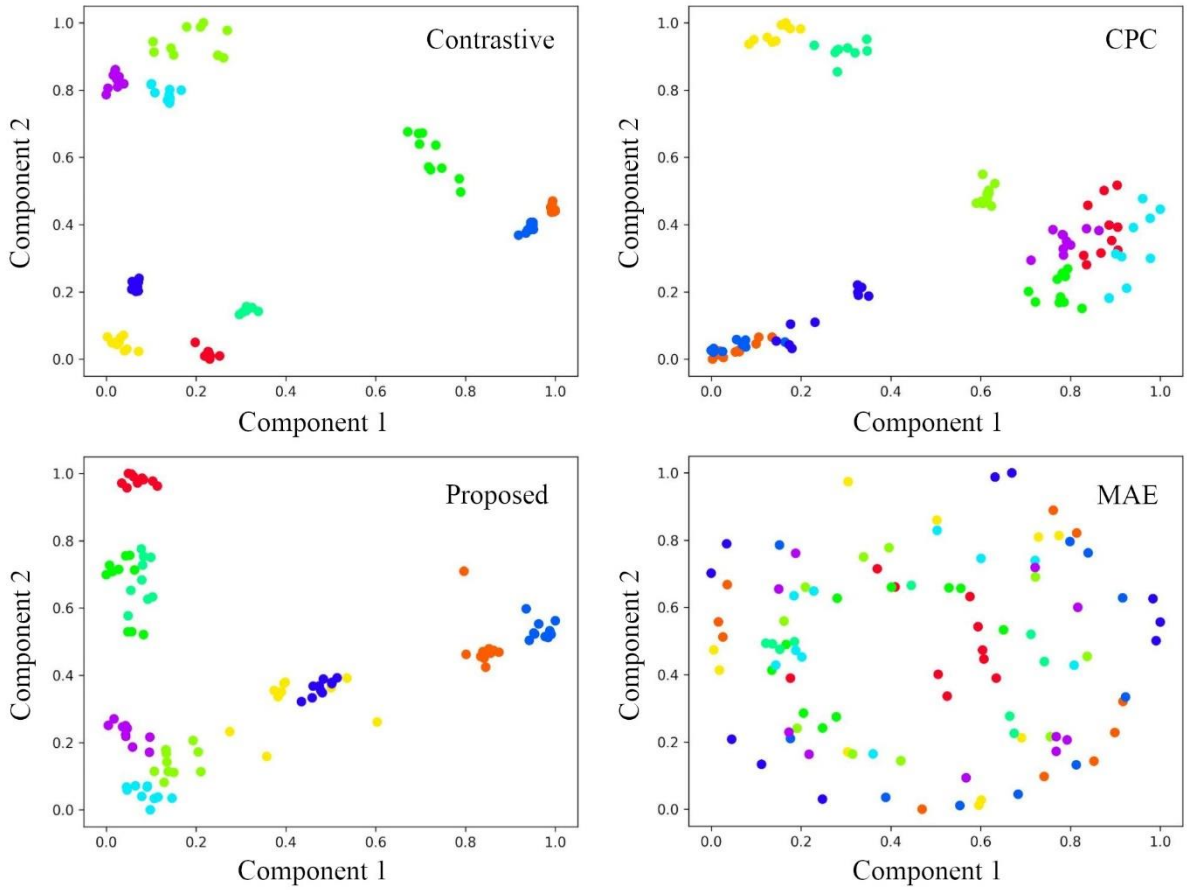

Fig. 17. **Two-components PCA-transformed embeddings of ten samples per ten different signals. Embeddings for different signals are displayed with different colors**

***Source*: compiled by the authors**

6. DISCUSSION OF RESULTS

The experiments showed that the proposed NN pre-training method via SSL pre-task showed promising results on transfer-learning to single-lead ECG diagnosis tasks with a small amount of training samples. It significantly outperformed the fully supervised training approach. It also showed better performance than our implementation of other SSL methods proved their effectiveness in other domains and successfully applied to the ECG analysis. We can also note that the experimental results are consistent with our hypotheses about potential issues with existing methods. The contrastive learning approach fails on short-term feature detection for atrial and flutter fibrillation statements. CPC is
focused on signal chunks comparison and produces better results, but still relies on mutual information maximization in a latent space, rather than direct signal comparison. MAE showed comparable results in local feature recognition but has a relatively bad overall performance, especially in diagnoses that require recognition of long-term signal features
such as premature atrial and ventricular contraction. The small amount of training data with a high level of intra-sample signal variance can be a potential problem, but further research is required. The proposed SSL method produces very contrasting results on pre-text tasks during self-supervised pre-training. Predictions visualization showed that the model either produces good predictions or completely fails to predict a signal peak or at least significantly smoothes it down (Fig. 18). Since the ECG signal has a very sharp peak, a slight temporal shift in peak prediction can lead to a significant increase in mean absolute error value which is

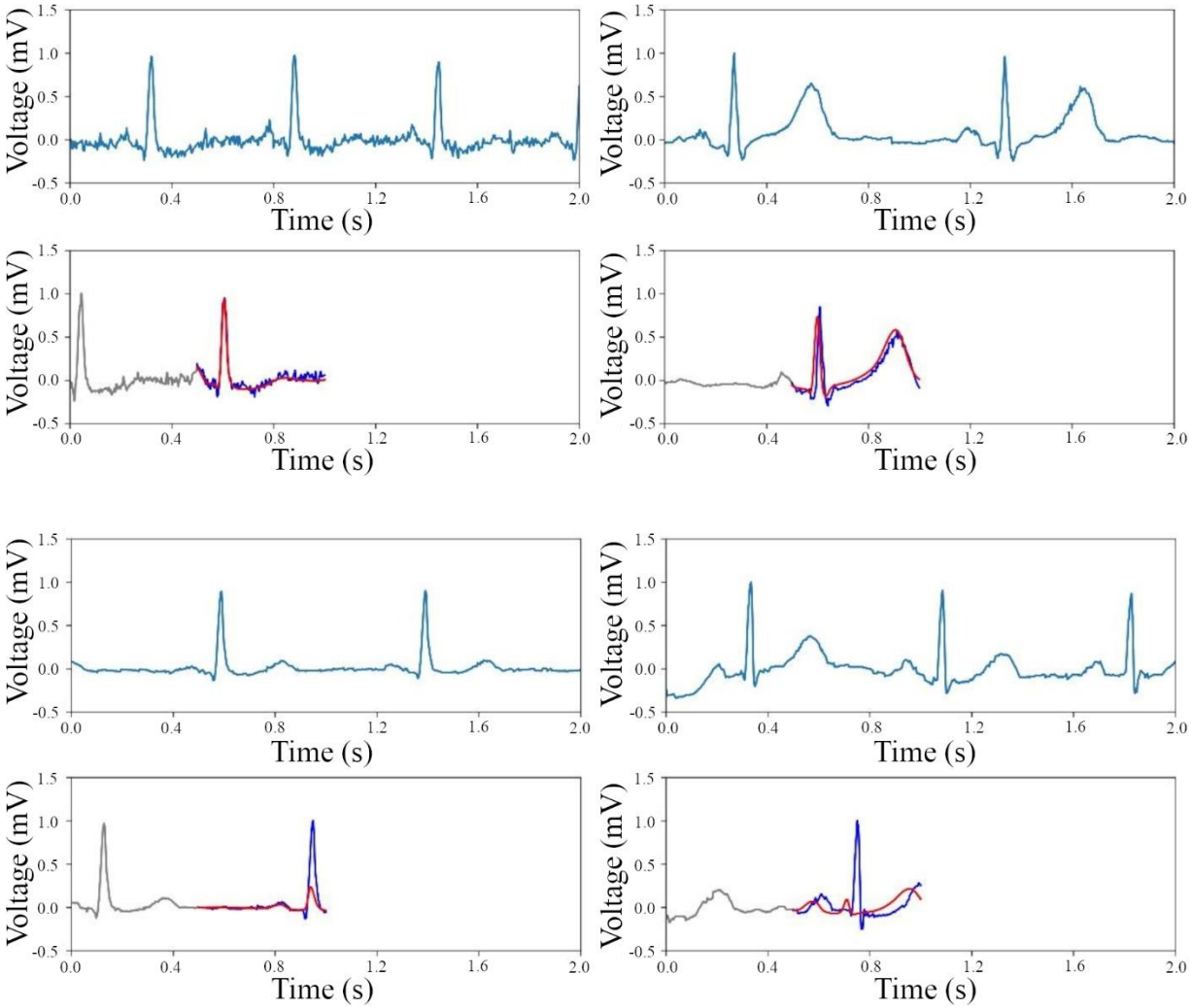

Fig. 18. **Example of SSL-task predictions**
***Source*: compiled by the authors**

used as a loss function. This can negatively affect the smoothness of the objective function of weight optimization and lead to the finding of local minima. It also could be said that the performance of the SSL task is not perfect, but the neural network can still generalize to the target task. Considering this, we believe that research on the ways to improve the convergence of the weights optimization problem can significantly improve the quality of learned representations.

7. CONCLUSION

The next contributions were made according to the initially set objectives:

1. Various approaches to Self-Supervised Learning methods and their applications for pretraining neural network models for ECG analysis were examined during the literature review part. The key characteristics of these methods, their specific applications to ECG data, and potential shortcomings were analyzed. Such potential issues as the tendency to lose information about short-term signal features and dependency on the shift between the recorded signal's start reading and the beginning of a heartbeat period are highlighted.

2. A theoretical SSL task without highlighted shortcomings, neural network model architecture, and a training algorithm were proposed.

3. A series of experiments were conducted. The obtained results demonstrated the advantage of the proposed method over the considered alternatives (0.90 average AUC for classification problems compared to the nearest 0.86 value obtained with the MAE method). The proposed method also showed better signal feature vector invariance compared to non-contrastive methods (0.32 vs 0.5 for CPC) which aligns with the initial hypotheses regarding the limitations of existing approaches. Results show the

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

potential of the proposed approach and the broader development of ECG-specific SSL methods.

The main direction for further research is an improvement in the SSL task performance. Different objective functions should be considered for the SSL task, to overcome the problem of quick objective raise with a slight temporal shift in predicted signal, and improve solution convergence.

REFERENCES

1. Macfarlane, P. W. & Kennedy, J. "Automated ECG interpretation – a brief history from high expectations to deepest networks". *Hearts*. 2021; 2: 433–448. DOI: https://doi.org/10.3390/hearts2040034.

2. Pipberger, H. V., Freis, E. D., Taback, L. & Mason, H. L. "Preparation of electrocardiographic data for analysis by digital electronic computer". *Circulation.* 1960; 21: 413–418.

3. Stallman, F. W. & Pipberger, H. V. "Automatic recognition of electrocardiographic waves by digital computer". *Circ. Res.* 1961; 9: 1138–1143.

4. Beraza, I. & Romero, I. "Comparative study of algorithms for ECG segmentation*". Biomedical Signal Processing and Control.* 2017; 34: 166–173. DOI: https://doi.org/10.1016/j.bspc.2017.01.013.

5. Rodrigues, L. & Marengoni, M. "Detecting QRS complex in ECG using wavelets and cubic spline interpolation". *Biomedical Engineering (BioMed)*. 2012. DOI: https://doi.org/10.2316/P.2012.764-135

6. Krak, I. V., Pashko, A., Stelia, O., Barmak, O. V. & Pavlov, S. "Selection parameters in the ECG signal for analysis of QRS complexes". *International Workshop on Intelligent Information Technologies & Systems of Information Security*. 2020.

7. Kaur, B. & Singla, S. "ECG analysis with signal classification using decision tree induction (DTI)". *Proceedings of the International Conference on Advances in Information Communication Technology & Computing (AICTC '16). Association for Computing Machinery*. 2016; 95: 1–6. DOI: https://doi.org/10.1145/2979779.2979874.

8. Mohebbanaaz, Kumari, L. V. R. & Sai, Y. P. "Classification of ECG beats using optimized decision tree and adaptive boosted optimized decision tree". *Signal, Image and Video Processing*. 2022. p. 695–703. DOI: https://doi.org/10.1007/s11760-021-02009-x.

9. Engin, M., Fedakar, M., Engin, E. Z. & Korürek, M. "Feature measurements of ECG beats based on statistical classifiers", *Measurement*. 2007; 40 (9-10): 904–912. DOI: https://doi.org/10.1016/j.measurement.2006.10.012.

10. Fainzilberg L. S. "Heart functional state diagnostic using pattern recognition of phase space ECG-images". *Proc. of the 6th European Congress on intelligent techniques and soft computing (EUFIT'98)*. 1998; 3 (B-27): 1878–1882.

11. Fainzilberg L. S. "Cyclic signals classification by codegrams characterizing the dynamics of cycles shape changing". *International Scientific Technical Journal Problems of Control and Informatics*. 2022; 67: 112–123. DOI: https://doi.org/10.34229/2786-6505-2022-3-9.

12. Filatova, A., Povoroznyuk, A., Nosachenko, B. & Fahs, M. "Synthesis of an integral signal for solving the problem of morphological analysis of electrocardiograms". *Herald of Advanced Information Technology*. 2022; 5 (4): 263–274. https://doi.org/10.15276/hait.05.2022.19.

13. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P. & Ng, A. Y. "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network". *Nat Med*. 2019. p. 65–69. DOI: http://doi.org/10.1038/s41591-018-0268-3.

14. Perez Alday, E. A., Gu, A., Shah, A. J., et al. "Classification of 12-lead ECGs: the PhysioNet/Computing in cardiology challenge 2020". *Physiol Meas*. 2020. DOI: http://doi.org/10.1088/1361-6579/abc960.

15. Reyna, M. A., Sadr, N., Perez Alday, E. A., et al. "Will two do? Varying dimensions in electrocardiography: The PhysioNet/Computing in cardiology challenge 2021". *Computing in Cardiology*. 2021; 48: 1-4. DOI: http://doi.org/10.23919/CinC53138.2021.9662687.

16. Ericsson, L., Gouk, H., Loy, C. C. & Hospedales, T. M. "Self-supervised representation learning: Introduction, advances, and challenges". *IEEE Signal Processing Magazine*. 2022; 39 (3): 42–62. DOI: http://doi.org/10.1109/MSP.2021.3134634.

17. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P. & Joulin, A. “Unsupervised learning of visual features by contrasting cluster assignments”. *Neural Information Processing Systems*. 2020. DOI: http://doi.org/10.48550/arXiv.2006.09882.

18. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. “A simple framework for contrastive learning of visual representations”. *In Proceedings of the 37th International Conference on Machine Learning (ICML’20)*. 2020. p. 1597–1607. DOI: https://dl.acm.org/doi/abs/10.5555/3524938.3525087.

19. Oord, Aäron van den, et al. “Representation learning with contrastive predictive coding.” *ArXiv*. 2018; abs/1807.03748. DOI: https://doi.org/10.48550/arXiv.1807.03748.

20. Grill, J.-B., et al. “Bootstrap your own latent – a new approach to self-supervised learning”. *Advances in Neural Information Processing Systems*. 2020; 33: 21271–21284. DOI: https://doi.org/10.48550/arXiv.2006.07733.

21. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. “2019. BERT: Pre-training of deep bidirectional transformers for language understanding”. *North American Chapter of the Association for Computational Linguistics*. 2019; 1: 4171–4186. DOI: http://doi.org/10.48550/arXiv.1810.04805.

22. Yenduri, G., Murugan, R., Govardanan, C. et al. “GPT (Generative Pre-Trained Transformer) – A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions”. *IEEE Access*. 2024. p. 1–1. DOI: https://doi.org/10.1109/ACCESS.2024.3389497.

23. Soltanieh, S., Hashemi,, J. & Etemad A. “In-distribution and out-of-distribution self-supervised ECG representation learning for arrhythmia detection”. *IEEE Journal of Biomedical and Health Informatics*. 2024; 28 (2): 789–800, https://www.scopus.com/authid/detail.uri?authorId=58693319100. DOI: https://doi.org/10.1109/JBHI.2023.3331626.

24. Mehari, T. & Strodthoff, N. “Self-supervised representation learning from 12-lead ECG data”. *Computers in Biology and Medicine*. 2022; 141, https://www.scopus.com/authid/detail.uri?authorId=57218712908. DOI: https://doi.org/10.1016/j.compbiomed.2021.105114.

25. Wang, Z., Ma, C., Zhang, S., Zhang, Y., Li, J. & Liu, C. “Lightweight arrhythmia detection based on momentum contrast learning”. *Computing in Cardiology***.** 2023, https://www.scopus.com/authid/detail.uri?authorId=58188792800. DOI: https://doi.org/10.22489/CinC.2023.164.

26. Zhang, H., Liu, W., Shi, J., Chang, S., Wang, H., He, J. & Huang, Q. “MaeFE: Masked autoencoders family of electrocardiogram for self-supervised pretraining and transfer learning”. *IEEE Transactions on Instrumentation and Measurement*. 2023; 72: 1–15. DOI: https://doi.org/10.1109/TIM.2022.3228267.

27. Wang, G., Wang, Q., Iyer, G. N., Nag, A. & John, D. “Unsupervised pre-training using masked autoencoders for ECG analysis”. *BioCAS 2023-2023 IEEE Biomedical Circuits and Systems Conference*. 2023, https://www.scopus.com/authid/detail.uri?authorId=57222019892.
DOI: https://doi.org/10.1109/BioCAS58349.2023.10388636.

28. Mikolov, T., Chen, K., Corrado, G. & Dean, J. “Efficient estimation of word representations in vector space”. *1st International Conference on Learning Representations, ICLR 2013-Workshop Track Proceedings*. 2013, https://www.scopus.com/authid/detail.uri?authorId=34969425500.
DOI: https://doi.org/10.48550/arXiv.1301.3781.

29. Noroozi, M., & Favaro, P. “Unsupervised learning of visual representations by solving jigsaw puzzles”. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2016; 9910: 69–84, https://www.scopus.com/authid/detail.uri?authorId=57191433285. DOI: https://doi.org/10.1007/978-3-319-46466-4_5.

30. Zhang, R., Isola, P. & Efros, A. A. “Colorful image colorization”. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2016; p. 649–666, https://www.scopus.com/authid/detail.uri?authorId=57223101349. DOI: https://doi.org/10.1007/978-3-319-46487-9_40.

31. Doersch, C., Gupta, A. & Efros, A. A. “Unsupervised visual representation learning by context prediction”. *Proceedings of the IEEE International Conference on Computer Vision, 2015 International Conference on Computer Vision (ICCV)*. 2015; art. No. 7410524: 1422–1430, https://www.scopus.com/authid/detail.uri?authorId=36445038000.
DOI: https://doi.org/10.1109/ICCV.2015.167.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

32. Yamaguchi, S., Kanai, S., Shioda, T. & Takeda, S. “Image enhanced rotation prediction for self-supervised learning”. *Proceedings-International Conference on Image Processing, (ICIP)*. 2019. p. 489–493, https://www.scopus.com/authid/detail.uri?authorId=57219583496.
DOI: https://doi.org/10.1109/ICIP42928.2021.9506132.

33. Woo, S. et al. “ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders”. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR*). 2023. p. 16133–16142. DOI: https://doi.org/10.1109/CVPR52729.2023.01548.

34. Zhang, W., Geng, S. & Hong S. “A simple self-supervised ECG representation learning method via manipulated temporal–spatial reverse detection”. *Biomedical Signal Processing and Control*. 2023; 79 (104194), https://www.scopus.com/authid/detail.uri?authorId=57432103700. DOI: https://10.1016/j.bspc.2022.104194.

35. Sarkar, P. & Etemad, A. “Self-supervised ECG representation learning for emotion recognition”. *IEEE Transactions on Affective Computing*. 2020; 13 (3): 1541–1554. https://www.scopus.com/authid/detail.uri?authorId=57211192178.
DOI: https://doi.org/10.1109/TAFFC.2020.3014842.

36. Witvliet, M. P., Karregat, E., Himmelreich, J., de Jong, J., Lucassen, W. & Harskamp, R. “Usefulness, pitfalls and interpretation of handheld single-lead electrocardiograms”. *Journal of Electrocardiology*. 2021; 66: 33–37. DOI: https://doi.org/10.1016/j.jelectrocard.2021.02.011.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 2312.2024
Received after revision 14.03.2025
Accepted 20.03.2025

DOI: https://doi.org/10.15276/hait.08.2025.7
УДК 004.8

Прогнозування сигналу ЕКГ на основі представлень для переднавчання нейронних мереж

Ставиченко Сергій Григорович[1)]
ORCID: https://orcid.org/0009-0007-4329-292X; sergey.stavichenko@gmail.com
Філатова Ганна Євгенівна[1)]
ORCID: https://orcid.org/0000-0003-1982-2322; Hanna.Filatova@khpi.edu.ua
[1)] Національний Технічний Університет “ХПІ”, вул. Кирпичова 2. Харків 61002, Україна

АНОТАЦІЯ

Обмежений набір даних для навчання є відомою проблемою при застосуванні методів глибокого навчання нейронних мереж. Проблема є особливо актуальною в галузі обробки біомедичних сигналів, таких як сигнал електрокардіограми, оскільки підготовка навчальних даних є трудомісткою і вимагає залучення кваліфікованих спеціалістів. Методи самокерованного навчання які спочатку зародилися в таких галузях машинного навчання, як обробка природної мови та комп’ютерний зір, є одним із шляхів розв’язання цієї проблеми і набувають все більшого поширення в сфері обробки біомедичних сигналів. Однак, пряма адаптація методів самокерованного навчання розроблених для інших доменів, не враховує таких особливостей даних електрокардіограм, як квазіперіодичність, локалізація морфологічних ознак, чутливість до шумів. Це робить актуальним розвиток специфічних методів самокерованного навчання для даних електрокардіограм. У цій роботі представлено новий метод самокерованного навчання для переднавчання нейронних мереж на нерозмічених даних електрокардіограм. Запропонований метод використовує задачу прогнозування короткої наступної підвибірки сигналу на основі попередньої підвибірки та вектора представлення. Моделі прогнозування та трансформації сигналу у вектори представлення навчаються спільно на датасеті MIMIC-ECG-IV із використанням методу зворотного поширення помилки при мінімізації середньоквадратичної помилки між спрогнозованою та оригінальною підвибірками сигналу. Як приклад цільової

задачі було обрано навчання лінійних бінарних класифікаторів на датасеті PTB-XL для діагностики патологічних станів пацієнта за І-м відведенням. Розмір навчальної вибірки для кожного діагнозу обмежено тридцятьма чотирма прикладами. Спочатку модель отримання представлень була навчена на нерозміченому датасеті MIMIC-ECG-IV, а потім навчалися лінійні моделі класифікації отриманих представлень для кожного обраного діагнозу в PTB-XL. Також проводилося порівняння з навчанням моделі представлень із випадково ініціалізованими вагами разом із навчанням класифікатора. Ефективність запропонованого методу порівнювалася з адаптаціями таких методів, як Contrastive Learning, Contrastive Predictive Coding та Masked Autoencoders. Для забезпечення принципів контрольованості експериментів були розроблені власні реалізації розглянутих методів із використанням спільної кодової бази та архітектурних блоків. Результати експериментів показали значну перевагу всіх розглянутих методів порівняно із сумісним навчанням моделі виділення ознак, а також перевагу запропонованого методу самокерованного навчання над іншими розглянутими методами, особливо на діагнозах із неявно вираженими короткочасними ознаками форми сигналу, таких як фібриляція та тріпотіння передсердь. Проведені експерименти демонструють перспективність подальших досліджень у галузі розробки специфічних методів самокерованного навчання для даних електрокардіограм як ефективного підходу до підвищення ефективності нейромережевих методів в умовах обмеження навчальних даних.

Ключові слова: біомедичні сигнали; електрокардіограма; глибоке навчання; самокероване навчання; вектор представлення; прогнозування сигналу

ABOUT THE AUTHORS

Serhii G. Stavychenko - PhD Student of Computer Engineering and Programming Department. National Technical University “Kharkiv Polytechnic Institute", 2, Kyrpychova Str. Kharkiv, 61002, Ukraine.
ORCID: https://orcid.org/0009-0007-4329-292X; sergey.stavichenko@gmail.com
Research field: Architecture of operating systems, development and design of databases, signal and image processing, design of computer diagnostic systems, optimization of processes in multiservice systems and networks

Ставиченко Сергій Григорович - аспірант кафедри Комп'ютерної інженерії та програмування. Національний Технічний Університет “ХПІ”, вул. Кирпичова 2. Харків, 61002, Україна

Anna E. Filatova - Doctor of Engineering Sciences, Professor, Professor of Computer Engineering and Programming Department. National Technical University “Kharkiv Polytechnic Institute", 2, Kyrpychova Str. Kharkiv, 61002, Ukraine
ORCID: https://orcid.org/0000-0003-1982-2322; Hanna.Filatova@khpi.edu.ua
Research field: Architecture of operating systems, development and design of databases, signal and image processing, design of computer diagnostic systems, optimization of processes in multiservice systems and networks

Філатова Ганна Євгенівна - доктор технічних наук, професор кафедри Комп'ютерної інженерії та програмування, Національний Технічний Університет “ХПІ”, вул. Кирпичова 2. Харків 61002, Україна

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)