The Ministry of Education and Science of The Russian Federation
Moscow Institute of Physics and Technology (State University)
Faculty of Control and Applied Mathematics
Department of «Intelligent Systems»
under Dorodnicyn Computing Center of RAS

Ignatov Andrey Dmitrievich

Deep Learning in information analysis of electrocardiogram signals for disease diagnostics

010900 — Applied Mathematics and Physics

Bachelor's thesis

Scientific advisor:
Senior Researcher at CC RAS, DSc
Konstantin V. Vorontsov

Moscow

year 2015

Contents

Abstract

With a rapid development of technology, new advanced methods for diseases detection become available. Thus, one possible way to diagnose the internal organs diseases is based on the information analysis of electrocardiogram signals. Since these signals contain information about various physiological processes taking place in the human body, it is possible to use them for health status evaluation. In this work, we propose a deep learning-based method for detecting various human diseases on the heart rate variability (HRV) data acquired from the electrocardiograms. We consider supervised and unsupervised approaches for training the models. In the first case, we use Convolutional Neural Networks and train them on the labeled HRV dataset. In the second case, we learn features from HRV data using Stacked Autoencoders and Restricted Boltzmann Machines, and then build a classifier on the obtained ones. The accuracy of the proposed models is evaluated on the dataset of labeled electrocardiograms with diagnosis in 14 different diseases. The final model demonstrates high performance with an average accuracy over 90% for most of the diseases. The source code of the models is written in Lua and available publicly.

Keywords: *Deep Learning, Convolutional Neural Network, Stacked Autoencoder, Boltzmann Machine, electrocardiogram, heart rate variability.*

1 Introduction

Problem relevance. Electrocardiograms are a valuable source of information about human health state, and an appropriate processing of them gives a vital information that can be applied for diagnosing different kinds of diseases. Nevertheless, relatively little research effort has been given to explore the connection between non-cardiovascular diseases and ECG signals. The only solution for this problem was proposed in works [1,2] and is based on hand-designed features extracted from the ECG signals. Though this method shows relatively high performance on a large number of diseases, it still has several issues, among which are incomplete use of information held in the source data and a complication in extending it on new diseases. To solve these problems, deep learning techniques both for feature generation and classification can be used.

Goals and Objectives. The goal of the project is to create a purely machine learning approach for the problem of electrocardiogram processing and diseases classification. For this purpose, we consider both supervised and unsupervised deep learning techniques and choose the most appropriate one.

Research methods. To build a model capable of detecting diseases on the ECG data we use Convolutional Neural Networks, Stacked Autoencoders and Restricted Boltzmann Machines. The implementation of our models is written in Lua using Torch machine learning library.

Scientific novelty.

- we propose a supervised approach based on Convolutional Neural Networks and available of diseases detection on the ECG data;
- we propose a modification of Convolutional Neural Networks in which convolutional layers are trained by unsupervised methods.

Practical value. We have developed a module that:

- preprocesses and visualizes the source electrocardiograms;
- implements Convolutional Neural Networks;
- implements Stacked Autoencoders and Restricted Boltzmann Machines;

- performs training of the proposed deep learning models;
- tests the models on the set of all available diseases and visualizes the results.

Statements to be defended.

- Convolutional Neural Network based model for diseases detection on the source HRV data;
- unsupervised modification of Convolutional Neural Network.

Publications.

- *Andrey D. Ignatov, Vadim V. Strijov.* Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer // Multimedia Tools and Applications, 2015, vol. 74, no. 11.

2 Literature review

2.1 Electrocardiogram processing

Electrocardiogram is a valuable indicator of the human health status. It contains a comprehensive information about physiological processes taking place in the human body and thus can be considered as a promising tool for health assessment. However, at the present time it is primarily used in medicine to diagnose only heart and vessel diseases [3, 4]. The majority of recent researches in this field are also focused on the detection of various cardiovascular illnesses. For instance, ECG signals are successfully applied for arrhythmia classification [5–7], myocardial ischemia detection [8–10], coronary artery disease detection [11–13], etc. A wide and exhaustive overview of the current state in ECG signals processing and interpretation is given in book [14].

Currently, only a few of works explored the connectivity between ECG data and non-cardiovascular diseases. In [15], authors have shown that the ECG signals change their form for a number of non-cardiac disorders, such as pulmonary embolism, central nervous system (CNS) diseases, myasthenia gravis, muscular tremors, hypothermia and hypothyroidism. Another study [16] has pointed out that besides CNS diseases the changes are presented for some esophageal disorders. There was also shown that drugs, poisons and electrical injuries have a considerable impact on the waveform of ECG signals. In another

paper [17], authors have presented a study where the connection between Friedreich's ataxia and electrocardiographic results was shown.

These papers were very important for understanding the information function of the heart, but the proposed methods could not be expanded to detect other diseases while they rely on a specific form of the ECG abnormalities. The problem was largely solved by Uspenskiy in [1, 2]. In his work, he has created a set of 216 features that were further extracted from the electrocardiograms and used for disease classification. The proposed solution was tested on the set of 30 diseases and has shown high performance. However, this method still has weaknesses. First, is does not answer the question whether the proposed features will perform well on new diseases. Another issue is that features were generated manually, and thus some information held in the source data can be lost.

2.2 Deep learning

When dealing with such complex data as ECG signals, it seems reasonable to use machine learning both for feature generation and classification. This approach is widely used in many data analysis problems related to the considered one: speech recognition [18, 19], image classification [20, 21], EEG signals processing [22, 23]. Currently, considerable progress in solving these tasks has been made by models based on deep learning architecture that learn multiple levels of representation of increasing abstraction.

Training of these models is generally done using two target concepts. In the first case, supervised learning is used and the model is trained as a whole. Among different architectures that use this concept quite attractive performance have convolutional neural networks (CNNs) [23,24]. Compared to standard feed-forward neural networks, they have convolutional layers that consist of small sets of neurons which look at small sub-regions of the input data. These layers have shared weights, and thus the process of learning them is equivalent to the process of learning filters that are applied for each sub-region of the data. Due to the small number of connections, they are trained more effectively, and at the same time these filters automatically learn features from the data. While the previous experiments have shown that the major changes in the form of ECG signals are local, CNNs seems well suited for the purpose of their detection and classification.

The second training concept is based on unsupervised representation learning from the data. Three basic approaches [25] for representation learning are now primarily used: Principal Component Analysis, Autoencoders and Boltzmann Machines. The first one,

Principal Component Analysis [26], is one of the oldest feature extraction algorithms, that learns a linear transformation of the input data vectors. Though there is a number of simple and efficient implementations of this algorithm, it can generate only linear features, and therefore its power is quite limited.

The other two approaches have layerwise architecture represented by the "visible" layers that correspond to the data and the "hidden" layers corresponding to the inherent features of the data. The main difference between these approaches is that in Boltzmann Machines hidden units are considered as latent random variables and in Autoencoders they are considered as computational nodes. From the first perspective [27], the question of feature learning can be interpreted as an attempt to recover a set of these latent random variables that describe distribution over the observed data. Then learning is conceived in term of estimating a set of model parameters that maximizes the likelihood of the training data, and the posterior of the latent variables can be interpreted as a new data representation. In the Autoencoder framework, inversely, one starts by explicitly defining a feature-extracting function in a specific parameterized closed form [28,29]. This function is called the encoder and it allows the straightforward computation of a feature vector using the input data. Whereas the described approaches have developed a parallel line for the problem, when building deep models the connection between them becomes tenuous, and the obtained results are usually different in nature. Thus, in this work we try both of them and choose the most appropriate one.

The contributions of this work are as follows. First, we propose a supervised approach to build a model available of diseases detection on ECG data. Here, we use convolutional neural networks and train them directly on the labeled heart rate variability data acquired from the ECG signals. Second, we propose a modification of CNN in which the convolutional layers are trained by unsupervised methods. We test our models on the dataset from the work of Uspenskiy [1, 2], and compare the results with previously obtained on the same data. The implementation of our models is written in Lua using Torch machine learning library and is available publicly.

3 Preliminaries

In this section, we introduce some useful notations and give a brief survey of the considered deep learning models.

3.1 Convolutional Neural Network

CNN is a hierarchial feed-forward neural network which structure is inspired by the biological visual system. The principal difference from the standard networks is that instead of using fully connected layers, it proposes learning filters that are applied to a small sub-regions of the data. In terms of the network structure, this means that the output values of each layer are obtained by applying an identical weight matrix (kernel) and activation function to each block of the data. This results in a shift-invariant features generated for each sub-region which form a representation of the input data. More precisely, the structure of CNNs typically consists of the following three types of layers:

- *Convolutional layer.* In this layer, a dot product (or a convolution) of each sub-region of the input data with a kernel is computed, and the result values form the output of this layer. The layer is parameterized by the size and the number of kernels, steps of the convolution in the width and height dimensions, and the activation function h applied to introduce a non-linearity to this layer (usually *sigmoid* or *tanh*). If the kernel size is (k_m, k_n), a size of the two-dimensional input data $d_{x,y}$ is (m, n), and (s_x, s_y) are steps of the convolution, then the output values $d'_{x,y}$ are computed according to:

$$d'_{x,y} = h\big(\sum_{i=1}^{k_m}\sum_{j=1}^{k_n} d_{x'+i,y'+j} \times w_{i,j}\big), \tag{3.1}$$

  where $x' = x \cdot s_x - 2$, $y' = y \cdot s_y - 2$, and $w_{i,j}$ are the weights of the kernel. Using these notations, the size of the output layer then is:

$$\mathrm{m}' = \lfloor \frac{\mathrm{m} - k_x}{s_x} + 1 \rfloor, \quad \mathrm{n}' = \lfloor \frac{\mathrm{n} - k_y}{s_y} + 1 \rfloor. \tag{3.2}$$

- *Max-pooling layer.* This layer follows a convolutional layer and performs a down-sampling operation in order to reduce the feature size. It takes small rectangular blocks of the data and produces a singular output for each block. This can be done by several ways, but the common one is taking a maximum in the block. Thus, if the block size is 2 × 2, then the number of features will be reduced by 4 times.

- . Finally, after several convolutional and max-pooling layers, the obtained features are converted into a single one-dimensional vector that is used for the classification. The classification layers are fully connected and usually use one output unit per class label.

While the overall structure is similar for all convolutional networks, their architecture can be significantly varied depending on the problem type and the used approach. An example of LeNet architecture used for handwritten digits recognition in paper [30] is presented on the figure 1.

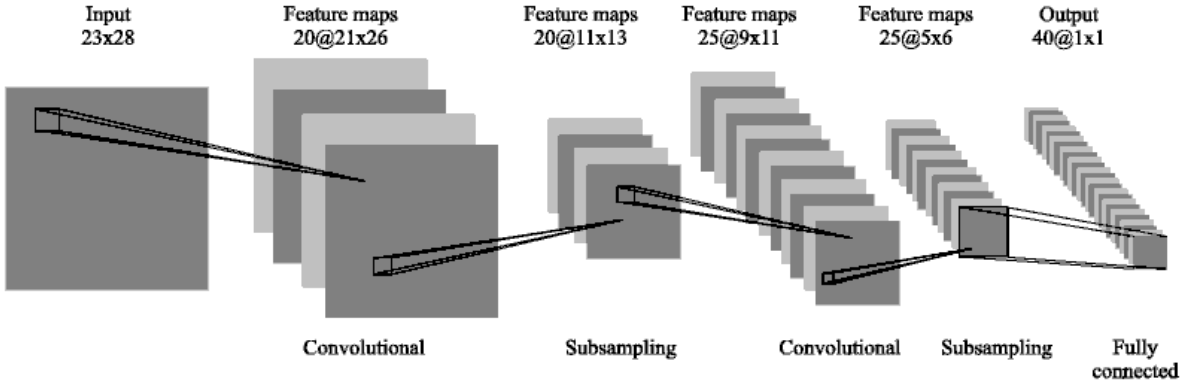

Figure 1: LeNet CNN architecture

3.2 Single-Layer Autoencoder

An Autoencoder is a symmetrical neural network that is used for unsupervised feature learning. Its training is done by minimizing the reconstruction error between the input data and its reconstruction at the output layer, and the activation values of the hidden layer are considered as a feature vector corresponding to the input data.

Encoding of an input vector $\mathbf{x} \in \mathbb{R}^N$ in the autoencoder is done by applying a linear transformation and a nonlinear activation function to $\mathbf{x}$:

$$\mathbf{h} = \sigma(\mathbf{W}_1\mathbf{x} + \mathbf{b_1}), \tag{3.3}$$

where $\mathbf{W}_1 \in \mathbb{R}^{H\times N}$ is a weight matrix, $\mathbf{b_1} \in \mathbb{R}^H$ is a bias, $\sigma(t) = (1 + \exp(-t))^{-1}$ is a logistic sigmoid function, and $\mathbf{h} \in \mathbb{R}^H$ represents activation values of the hidden layer. Decoding of the obtained vector $\mathbf{h}$ is done by another transformation with a separate weight matrix $\mathbf{W}_2$ and bias $\mathbf{b}_2$:

$$\mathbf{x}' = \sigma(\mathbf{W}_2^\intercal\mathbf{h} + \mathbf{b_2}), \tag{3.4}$$

where $\mathbf{x}'$ is a reconstruction of the vector $\mathbf{x}$. With these notations, in case of mean square error criterion the reconstruction error can be written as:

$$\mathcal{L}(\mathrm{X}, \mathrm{X}') = \frac{1}{2}\sum_{i=1}^{m} ||\mathbf{x}_i - \mathbf{x}'_i||_2^2, \tag{3.5}$$

where $\mathrm{X} = \{\mathbf{x}_i\}_{i=1}^m$ and $\mathrm{X}' = \{\mathbf{x}'_i\}_{i=1}^m$ are the training and reconstructed data respectively.

3.3 Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is a two-layer bipartite graphical model with a set of visible units $\mathbf{v}$, a set of hidden units $\mathbf{h}$, and symmetrical connections between these two layers represented by a weight matrix $\mathbf{W}$. The joint distribution between the hidden and visible variables is given by:

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \exp\big(-E(\mathbf{v}, \mathbf{h})\big), \tag{3.6}$$

$$E(\mathbf{v}, \mathbf{h}) = -\mathbf{h}^\intercal \mathbf{W}\mathbf{v} - \mathbf{b}^\intercal \mathbf{v} - \mathbf{c}^\intercal \mathbf{h}, \tag{3.7}$$

where $E(\mathbf{v}, \mathbf{h})$ is an energy function, $\mathbf{b}$ and $\mathbf{c}$ are visible and hidden units biases respectively, Z is the partition function, and $v_i, b_i \in \{0, 1\}$. The hidden units are conditionally independent of one another given the visible layer, and vice versa. In principle, the model can be trained by performing stochastic gradient ascent on the log-likelihood of the data, however, in this case it is computationally intractable. Instead, one uses the contrastive divergence approximation (Hinton, 1999, 2002), which performs well in practice.

3.4 Stacked Autoencoders and Deep Belief Networks

While the representation power of a single Autoencoder and RBM is limited, they are usually stacked to form a deep architecture — a Stacked Autoencoder and a Deep Belief Network (DBN) respectively. These models consist of many layers, and each two adjacent layers have a full set of connections between them. Training these models as a whole is computationally costly, and thus it is usually done by greedily training each layer using the activations from the previous level as an input. Once a stack of Autoencoders or a DBN has been built, its highest level output representation can be used as an input to a stand-alone supervised learning algorithm. Alternatively, a logistic regression layer can be added to the top layer, yielding a deep neural network that can be used for supervised learning.

4 Data description

The considered dataset consists of electrocardiograms obtained from 192 healthy persons and 9396 patients with various internal diseases. These ECG were acquired using electrocardiography complex "KAD-3" , which comparing to the common cardiographs has a broader frequency band of input signals (from 0.5 to 500 Hz). To make the diagnoses

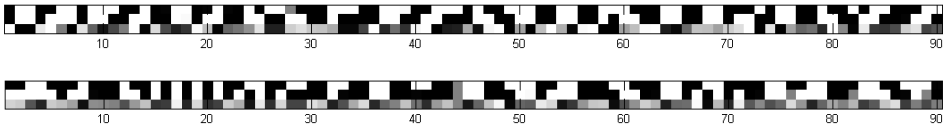

Figure 2: ECG visual representation

of patients more reliable, in the dataset were included those of them that were confirmed both by clinical evaluation and laboratory-instrumentation methods.

While electrocardiograms have an excessively complex structure, their simplified representation is usually used instead. In cardiology, the intervals between QRS complexes [4] are commonly measured, and the configuration of these signals is then visually observed. In this work, besides these intervals we also consider an amplitude or "swing" of the QRS complexes [31], and a ratio between each amplitude-interval pair, which form heart rate variability data. Thus, in this case the cardiogram is represented by three one-dimensional time series of amplitudes, intervals and their ratio respectively. To obtain valid information, there were registered 600 modes of electrocardiogram that leads to the same length of the mentioned time series.

Finally, we proposed a visual representation of this data. Instead of using raw time series, we considered increments between each two adjacent points and applied a hyperbolic arctangent function to normalize it. Thus, for each electrocardiogram we obtained a grayscale image of size 599×3, well suitable for its further processing. Besides a good visualization, this allowed the problem to be treated a visual recognition problem, which has opened up new ways of dealing with it. An example of the obtained visual representation is presented on the figure 2.

5 Model architecture

When developing a model structure, the principal task is to take into account the specifics of the data. One important property of the measured signal is that its waveform is quite the same both for healthy and ill people, particularly, this difference can not be observed by visual analysis. However, when considering local behavior of the signal, it has tangibly different variations for people with illnesses. To catch this distinction, the model should learn local data structure at the first levels of abstraction. For this purpose, two different approaches are considered.

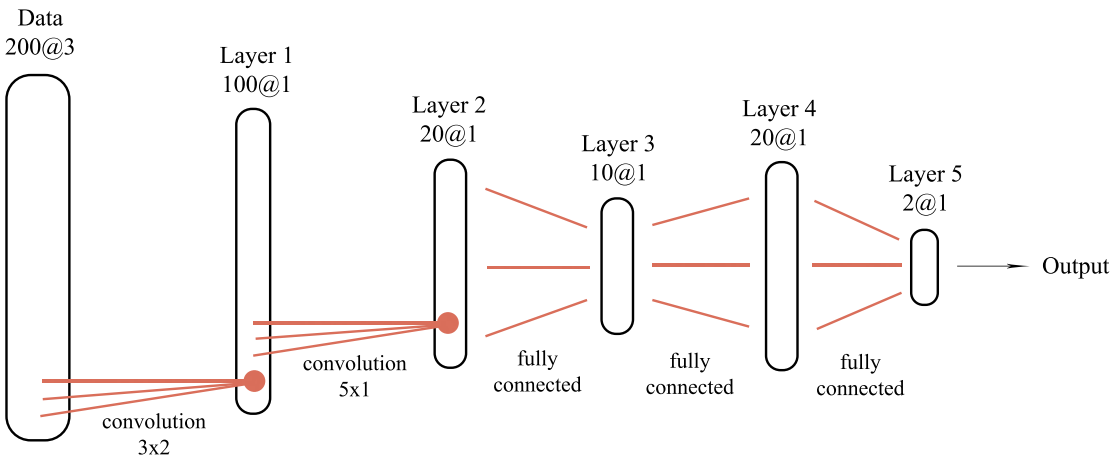

Figure 3: CNN architecture

5.1 Convolutional Neural Network

The first approach is based on the supervised learning. Our initial experiments have shown that a standard neural network with fully connected layers is not applicable to this problem. Because of the complex data structure and a small number of the learning samples (about 800 per class), it is prone to overfitting the training set and is not able to catch the data specifics mentioned above. Thus, a model with a more appropriate structure was required. Generally, this model should also meet the following two criterions: it should have enough levels of representation to learn the described data variations, and at the same time have a relatively small number of training parameters not to overfit the small dataset. To achieve this, a model based on the Convolutional Neural Network was developed.

The architecture of our neural network is summarized in figure 3. It contains five learned layers: two convolutional and three fully-connected. The first layer learns invariant features that detect local abnormalities in the ECG signals, and consists only of one convolutional kernel of size 3×2. The computational experiment described below has shown that we can use non-overlapping data sub-regions while performing a convolution. This has reduced a number of output values by six times and made the use of max-pooling layer unnecessary.

The second convolutional layer uses a kernel of size 5×1, and therefore each output from this layer holds the information about 30 subsequent points from the input data. While this is enough to learn the data structure, the proposed architecture prevents overfitting and is easy to train because it has only a few different weights to learn in the

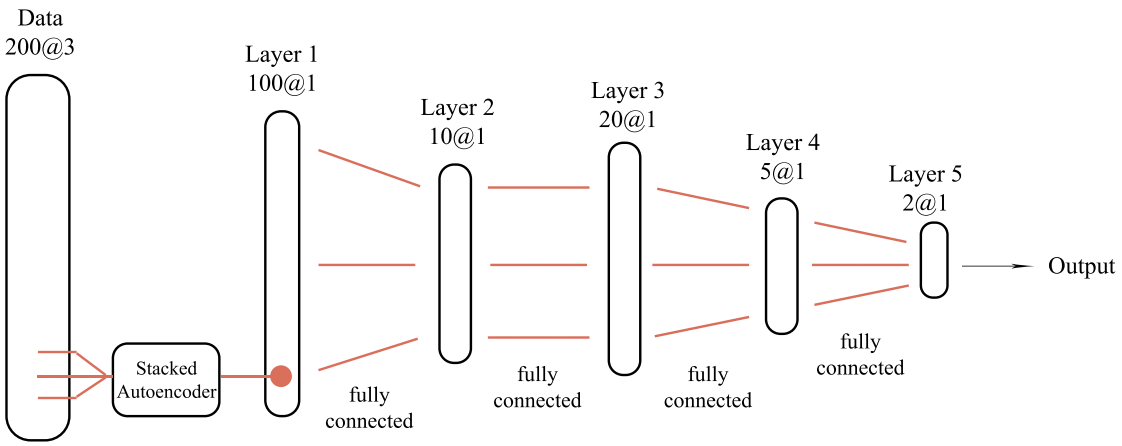

Figure 4: The architecture of the unsupervised-based model

first two layers. The last three layers are fully-connected and they perform a classification of the features obtained after the convolutional layers. The output of the last layer is fed to a softmax which produces a distribution over the class labels. The details about training and regularization of the proposed model are given in the computational experiment.

5.2 CNN with unsupervised convolutional layer

The second approach was developed using unsupervised learning techniques. We were facing the same challenges as described in the section above. A standard Autoencoder and Restricted Boltzmann Machine are good at learning data trends, but they are not able to encode the small local variation, which in this case can be considered as a noise. Therefore, here we propose using a similar model structure to the Convolutional Neural Network, but the role of a convolutional kernel here plays a stacked encoder that learns small pieces of the input data. Thus, the principal difference is that this "unsupervised convolutional layer" is training separately from the whole model and its outputs represent purely encoded data instead of the learned feature maps.

This layer brings several advantages to the training process. First, an encoder can be trained on the small data pieces, and this greatly increases a training sample while each data sub-region can be used for training. This also prevents overfitting and makes it possible for encoder to consist of many layers. Second, an encoder learns the main patterns that occur in HRV data for healthy and ill people, which is crucial for our problem. Finally, the proposed layer reduces the size of the data by generating only a few features for each sub-region almost without losing an information held in it.

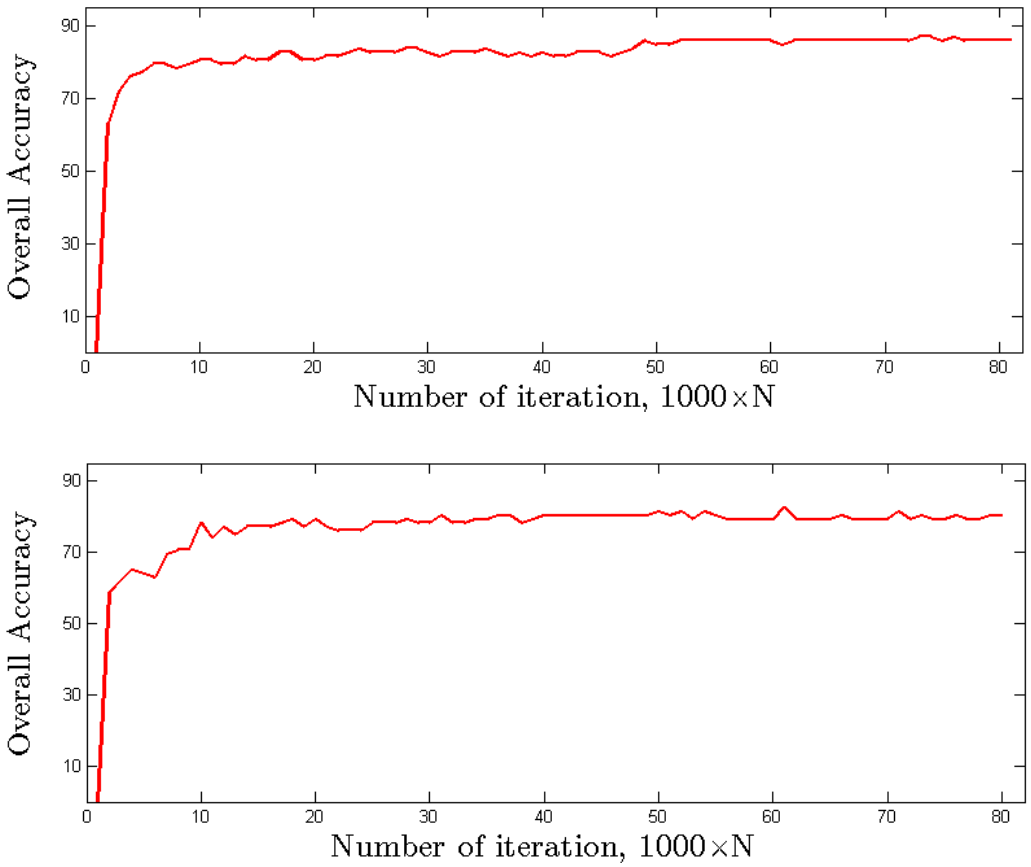

Figure 5: The accuracy of the models on the validation set during the training process

In this work, we use an encoder that consists of six hidden layers, and train it as a whole on the HRV data. The features obtained after this unsupervised layer are then classified with a Deep Neural Network, which together with this layer forms an architecture presented in the figure 4.

6 Computational experiment

In this section, we make a quantitative evaluation of the proposed approaches on the dataset of ECG signals. The signals were acquired from 192 healthy people and 9396 patients ill with 14 different diseases. The dataset was initially split into two non-overlapping sets: 80% of ECGs were used for training the models, and 20% for testing.

The models were trained using stochastic gradient descent with momentum of 0.9 and weight decay of 0.02. A weight decay was important for the models to learn because of an excessively small amount of the training data. It prevented them from overfitting, and together with a developed model structure allowed the models not to lose gener-

alization ability during the whole training process. The accuracy of the supervised and unsupervised models on the validation set during their training is presented in the figure 5.

We report the performance of our models in tables 1 and 2. Here, for each of 14 diseases we specify the sample size, sensitivity, specificity, and the overall diagnostic accuracy obtained on the test set. Sensitivity here implies the percentage of people for whom the disease was successfully detected, and specificity means the percentage of healthy people who were not misdiagnosed this particular disease.

The first approach, based on the Convolutional Neural Network, shows an average performance of 88.45% among all diseases. For the majority of them it has an overall diagnostic accuracy over 88%. The most complicated task for this model was to distinguish between healthy people and people ill with diseases number 1, 11 and 13, and in this case the overall accuracy was only about 80%.

The second, unsupervised-based approach, shows a noticeable performance increase comparing to the CNN. The average accuracy of this model is 2.19% higher, and reaches 90,64%. The lowest accuracy, which was obtained for disease number 1, was also increased by about 4% to 82.35%. Overall, this method shows an improvement for all diseases except 7, 9 and 10, for which the results remain the same.

To compare these results with previously obtained on the same set of diseases, we have implemented a feature generation algorithm proposed by Uspenskiy as it was described in his work [32]. To classify the features, we used k-nearest neighbor method with a number of neighbors set to 1 for all diseases as a baseline. The obtained results are summarized in table 3. The average accuracy of this method is 87.71%, and generally the results are very similar to those ones obtained by the CNN. However, when considering the unsupervised-based model, this yields a significantly better performance for all diseases form the set, and in some cases (for diseases number 1, 5, 10 and 11) the increase can be over 5%.

7 Conclusion

In this work, we present a deep learning-based method for human diseases detection on the electrocardiogram signals. Taking into account the data specifics, we propose two different approaches to solve the problem that are based on the supervised and unsupervised learning. For the first approach, we use the Convolutional Neural Networks, and train them on a raw HRV data. For the second one, we propose an unsupervised modification of the convolutional layer, where instead of the convolutional kernel a Stacked

Table 1: The overall accuracy of the supervised model

Disease number	Sample size	Sensitivity	Specificity	Overall accuracy, %
1	660	88.28	47.62	78.24
2	1837	100	58.62	97.04
3	266	94.12	85.37	90.22
4	1232	98.82	64.52	95.09
5	625	94.87	73.91	88.96
6	745	92.11	71.43	88.23
7	836	97.01	76.92	93.20
8	700	94.20	85.00	92.13
9	662	95.56	61.11	88.30
10	329	94.66	76.31	90.38
11	244	73.81	86.67	80.46
12	238	84.62	91.18	87.21
13	266	93.22	57.78	80.43
14	756	94.63	65.85	88.42

Table 2: The overall accuracy of the unsupervised model

Disease number	Sample size	Sensitivity	Specificity	Overall accuracy, %
1	660	94.53	45.24	82.35
2	1837	99.20	75.86	97.54
3	266	94.19	90.24	92.39
4	1232	99.21	80.65	97.19
5	625	94.87	76.09	89.57
6	745	97.37	65.71	91.44
7	836	98.80	69.23	93.20
8	700	97.83	75.00	92.70
9	662	94.82	63.89	88.30
10	329	92.00	86.21	90.38
11	244	83.33	84.44	83.90
12	238	96.15	88.24	93.02
13	266	91.53	78.79	86.96
14	756	94.63	73.17	90.00

Table 3: The overall accuracy of the method proposed by Uspenskiy

Disease number	Sample size	Sensitivity	Specificity	Overall accuracy, %
1	660	85.93	35.71	73.53
2	1837	97.34	62.07	94.83
3	266	94.11	85.37	90.22
4	1232	99.61	74.19	96.84
5	625	93.16	60.87	84.05
6	745	95.40	68.57	90.37
7	836	97.61	69.23	92.23
8	700	97.10	75.00	92.13
9	662	93.33	61.11	86.55
10	329	89.33	72.41	84.62
11	244	85.71	71.11	78.16
12	238	88.46	91.18	89.53
13	266	93.22	72.72	85.87
14	756	96.64	60.98	88.95

Autoencoder is used. We test the models on the set of labeled ECG signals obtained from healthy people and people ill with 14 different illnesses. To compare the results with previously proposed method [32] that uses hand-designed features, we implemented it and tested on the same dataset. The approach based on CNN and the method [32] show a similar performance with 88.45% and 87.71% of average diseases detection accuracy respectively. However, the use of generated by an Autoencoder features together with a deep architecture allows the unsupervised model to outperform previous results by more that 2% and achieve an average accuracy of 90,64%.

References

[1] V.M. Uspenskiy. Information function of the heart. a measurement model. *MEASUREMENT 2011*, 1:383–386, 2011.

[2] V.M. Uspenskiy. Diagnostic system based on the information analysis of electrocardiograph. In *Embedded Computing (MECO), 2012 Mediterranean Conference on*, pages 74–76, June 2012.

[3] Wagner Galen and Strauss David. *Marriott's Practical Electrocardiography*. LWW, 2013.

[4] K. Perloff Joseph and J. Marelli Ariane. *Clinical Recognition of Congenital Heart Disease*. Elsevier, 2012.

[5] Babak Mohammadzadeh Asl, Seyed Kamaledin Setarehdan, and Maryam Mohebbi. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial Intelligence in Medicine*, 44(1):51 – 64, 2008.

[6] Mi Hye Song, Jeon Lee, Sung Pil Cho, Kyoung Joung Lee, and Sun Kook Yoo. Support vector machine based arrhythmia classification using reduced features. *International Journal of Control, Automation, and Systems*, 3:571–579, 2005.

[7] A.J. Joshi, S. Chandran, V.K. Jayaraman, and B.D. Kulkarni. Hybrid svm for multiclass arrhythmia classification. In *Bioinformatics and Biomedicine, 2009. BIBM '09. IEEE International Conference on*, pages 287–290, Nov 2009.

[8] V. Ahanathapillai and J.J. Soraghan. Myocardial ischemia detection algorithm (mida): Automated echocardiography sequence analysis for diagnosis of heart muscle damage. In *Computing in Cardiology, 2010*, pages 405–408, Sept 2010.

[9] T. Konttila, M. Stenroos, H. Vaananen, H. Hanninen, M. Lindholm, I. Tierala, and T. Katila. Support vector classification of acute myocardial ischemia with reduced sets of body surface potential map electrodes. In *Computers in Cardiology, 2005*, pages 869–872, Sept 2005.

[10] P Ranjith, P.C Baby, and P Joseph. {ECG} analysis using wavelet transform: application to myocardial ischemia detection. *ITBM-RBM*, 24(1):44 – 47, 2003.

[11] Ismail Babaoglu, Mustafa Servet Kiran, Erkan Ulker, and Mesut Gunduz. Diagnosis of coronary artery disease using artificial bee colony and k-nearest neighbor algorithms. *International Journal of Computer and Communication Engineering*, 2:56–59, 2013.

[12] R. Alizadehsani, M.J. Hosseini, Z.A. Sani, A. Ghandeharioun, and R. Boghrati. Diagnosis of coronary artery disease using cost-sensitive algorithms. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 9–16, Dec 2012.

[13] A. Kampouraki, G. Manis, and C. Nikou. Heartbeat time series classification with support vector machines. *Information Technology in Biomedicine, IEEE Transactions on*, 13(4):512–518, July 2009.

[14] Adam Gacek and Witold Pedrycz. *ECG Signal Processing, Classification and Interpretation*. Springer-Verlag London, 2012.

[15] L. Yash, L. Pallavi, and A. Sameer. Ecg in non cardiac disorders.

[16] Mieghem Carlos, Sabbe Marc, and Knockaert Daniel. The clinical value of the ecg in noncardiac conditions. *Chest*, 125:1561–1576, 2004.

[17] John S. Child, Joseph K. Perloff, Philip M. Bach, Allan D. Wolfe, Susan Perlman, and R.A. Pieter Kark. Cardiac involvement in friedreich's ataxia: A clinical study of 75 patients. *Journal of the American College of Cardiology*, 7(6):1370–1378, 1986.

[18] Mikael Henaff, Kevin Jarrett, Koray Kavukcuoglu, and Yann Lecun. Unsupervised learning of sparse features for scalable audio classification.

[19] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations.

[20] Hoo-Chang Shin, M.R. Orton, D.J. Collins, S.J. Doran, and M.O. Leach. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1930–1943, Aug 2013.

[21] Lee Honglak, Largman Yan, Pham Peter, and Ng Andrew. Unsupervised feature learning for audio classification using convolutional deep belief networks.

[22] Wei-Long Zheng, Jia-Yi Zhu, Yong Peng, and Bao-Liang Lu. Eeg-based emotion classification using deep belief networks. In *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, pages 1–6, July 2014.

[23] H. Cecotti and A. Graeser. Convolutional neural network with embedded fourier transform for eeg classification. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, Dec 2008.

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou,

and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[25] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[26] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[27] Geoffrey E. Hinton. A practical guide to training restricted boltzmann machines. In *Neural Networks: Tricks of the Trade - Second Edition*, pages 599–619. 2012.

[28] Geoffrey E. Hinton and Richard S. Zemel. Autoencoders, minimum description length and helmholtz free energy, 1994.

[29] H. Bourlard and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59(4-5):291–294, 1988.

[30] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, P. Simard, and V. Vapnik. Comparison of learning algorithms for handwritten digit recognition. In *INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS*, pages 53–60, 1995.

[31] A.G. Ramakrishnan, A.P. Prathosh, and T.V. Ananthapadmanabha. Threshold-independent qrs detection using the dynamic plosion index. *Signal Processing Letters, IEEE*, 21(5):554–558, May 2014.

[32] V.M. Uspenskiy, K.V. Vorontsov, V.R. Tselykh, and V.A. Bunakov. Information function of the heart: Discrete and fuzzy encoding of the ecg-signal for multidisease diagnostic system. In *Advanced Mathematical and Computational Tools in Metrology*, 2014.