

# Committee Machine Networks to Diagnose Cardiovascular Diseases

Mehmet Can  
[mcan@ius.edu.ba](mailto:mcan@ius.edu.ba)

International University of Sarajevo  
Faculty of Engineering and Natural Sciences  
Hrasnicka Cesta 15, 71000 Sarajevo  
Bosnia and Herzegovina

## Abstract

**A parallel committee machines technique for neural network systems with back propagation together with a majority voting scheme is presented in this paper. Previous research with regards to predict the presence of cardiovascular diseases has shown accuracy rates up to 72.9% but it comes with a cost of reduced prediction accuracy of the minority class. The designed neural network system in this article presents a significant increase of robustness and it is shown that by majority voting of the parallel networks, recognition rates reach to > 90 in the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation data set.**

**Keywords**— Machine learning, parallel neural networks, boosting by filtering, cardiovascular diseases

## 1. INTRODUCTION

Cardiovascular disease, also called heart disease, is a class of diseases that involve the heart or blood vessels (arteries, capillaries and veins).[1]

Cardiovascular disease refers to any disease that affects the cardiovascular system, principally cardiac disease, vascular diseases of the brain and kidney, and peripheral arterial disease.[2] The causes of cardiovascular disease are diverse but atherosclerosis and/or hypertension are the most common. Additionally, with aging come a number of physiological and morphological changes that alter cardiovascular function and lead to subsequently increased risk of cardiovascular disease, even in healthy asymptomatic individuals.[3]

Cardiovascular disease is the leading cause of deaths worldwide, though since the 1970s, cardiovascular mortality rates have declined in many high-income countries. [4] At the same time, cardiovascular deaths and disease have increased at a fast rate in low- and middle-income countries.[5] Although cardiovascular disease usually affects older adults, the antecedents of cardiovascular disease,

notably atherosclerosis, begin in early life, making primary prevention efforts necessary from childhood.[6] There is therefore increased emphasis on preventing atherosclerosis by modifying risk factors, such as healthy eating, exercise, and avoidance of smoking.

### *1.1 Types of cardiovascular diseases*

- Coronary heart disease (also ischaemic heart disease or coronary artery disease)
- Cardiomyopathy - diseases of cardiac muscle
- Hypertensive heart disease - diseases of the heart secondary to high blood pressure
- Heart failure
- Cor pulmonale - a failure of the right side of the heart
- Cardiac dysrhythmias - abnormalities of heart rhythm
- Inflammatory heart disease
- Endocarditis – inflammation of the inner layer of the heart, the endocardium. The structures most commonly involved are the heart valves.
- Inflammatory cardiomegaly

- Myocarditis – inflammation of the myocardium, the muscular part of the heart.
- Valvular heart disease
- Cerebrovascular disease - disease of blood vessels that supplies to the brain such as stroke
- Peripheral arterial disease - disease of blood vessels that supplies to the arms and legs
- Congenital heart disease - heart structure malformations existing at birth
- Rheumatic heart disease - heart muscles and valves damage due to rheumatic fever caused by streptococcal bacteria infections

### 1.2 Risk factors

Epidemiology suggests a number of risk factors for heart disease: age, gender, high blood pressure, high serum cholesterol levels, tobacco smoking, excessive alcohol consumption, family history, obesity, lack of physical activity, psychosocial factors, diabetes mellitus, and air pollution. While the individual contribution of each risk factor varies between different communities or ethnic groups the consistency of the overall contribution of these risk factors to epidemiological studies is remarkably strong. Some of these risk factors, such as age, gender or family history, are immutable; however, many important cardiovascular risk factors are modifiable by lifestyle change, drug treatment or social change.

### 1.3 Research

The first studies on cardiovascular health were performed in 1949 by Jerry Morris using occupational health data and were published in 1958.[7] The causes, prevention, and/or treatment of all forms of cardiovascular disease remain active fields of biomedical research, with hundreds of scientific studies being published on a weekly basis. A trend has emerged, particularly in the early 2000s, in which numerous studies have revealed a link between fast food and an increase in heart disease. These studies include those conducted by the Ryan Mackey Memorial Research Institute, Harvard University and the Sydney Center for Cardiovascular Health. Many major fast food chains, particularly McDonald's, have protested the methods used in these studies and have responded with healthier menu options.

A fairly recent emphasis is on the link between low-grade inflammation that hallmarks atherosclerosis and its possible interventions. C-reactive protein (CRP) is a common inflammatory marker that has been found to be present in increased levels in patients at risk for cardiovascular disease.[8] Also osteoprotegerin which involved with regulation of a key inflammatory transcription factor called NF- $\kappa$ B has been found to be a risk factor of cardiovascular disease and mortality.

Some areas currently being researched include possible links between infection with *Chlamydia pneumoniae* (a major cause of pneumonia) and coronary artery disease. The

*Chlamydia* link has become less plausible with the absence of improvement after antibiotic use.[9]

Several research also investigated the benefits of melatonin on cardiovascular diseases prevention and cure. Melatonin is a pineal gland secretion and it is shown to be able to lower total cholesterol, very low density and low density lipoprotein cholesterol levels in the blood plasma of rats. Reduction of blood pressure is also observed when pharmacological doses are applied. Thus, it is deemed to be a plausible treatment for hypertension. However, further research needs to be conducted to investigate the side effects, optimal dosage and etc. before it can be licensed for use.[10]

### 1.4 Neural networks for complex medical diagnosis

In this article, an artificial intelligence alternative to the medical diagnosis is proposed. Neural networks are the tools that should be recalled for any classification job. They are developed enormously since the first attempts made modeling the perceptron architecture six decades ago [11].

The massive parallel computational structure of neural networks is what has contributed to its success in predictive tasks. It has been shown that the approach of using parallel networks is successful with respect to increasing the predictive accuracy of neural networks in robotics [12] and in disease diagnosis.

This work presents a parallel networks system which is bound together with a majority voting system in order to further increase the predictive accuracy of a cardiovascular diseases disease data set based on clinic recordings (reference).

For the proposed system it is shown with a case study of cardiovascular diseases. The type of network used is the standard feed forward back-propagation neural network, since they have proven useful in biomedical classification tasks [13]. The performance of the trained neural networks is evaluated according to the true positive, and true negative rate of the prediction task. Furthermore the area under the receiver operating characteristic curve and the Mean Squared Error are used as statistical measurements to compare the success of the different models.

The paper is organized as follows; first, the data used in this work is introduced in section 2. The neural network that is boosted by filtering is illustrated in section 3. Results of the research are shown in section 4 which followed by a conclusion.

## 2. DATA SET OF CARDIOVASCULAR DISEASES

### *Source of Data: Cleveland Clinic*

The Cleveland Clinic, formally known as the Cleveland Clinic Foundation, is a multispecialty academic medical center located in Cleveland, Ohio, United States. The Cleveland Clinic was established in 1921 by four physicians for the purpose of providing patient care, research, and medical education in an ideal medical setting.

The Cleveland Clinic Lerner Research Institute is home to all laboratory-based, translational and clinical

research at Cleveland Clinic. A new medical school, the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, was opened in 2004. The program's curriculum was devised by Cleveland Clinic staff physicians to train and mentor a new generation of physician-investigators.

Database contains 302 data with 76 attributes for each of them, but all published experiments refer to using a subset of 14 of them. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 no presence) to 4.

Table 1: Table describing the 14 attributes that are not used.

1.	age
2.	sex
3.	cp
4.	trestbps
5.	chol
6.	fbs
7.	restecg
8.	thalach
9.	exang
10.	oldpeak
11.	slope
12.	ca
13.	thal
14.	num (the predicted attribute)

Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence values 1, 2, 3, 4) from absence value 0).[14]

### 3. PRINCIPAL COMPONENT ANALYSIS

Principle component analysis (PCA) finds the linear combination of attributes that best accounts for the variations in the data. Two-dimensional plots of the first two principal components supply us with a means to inspect visually for trends, which occur as clusters of points. Later, cluster analysis may follow this step.

This simple but effective method continues to be used today, partly because of the ease with which the results are communicated and interpreted.

#### 3.1 Theory of Principal component Analysis

Multivariate statistics deals with the relation between several random variables. The sets of observations of the random variables are represented by a multivariate data matrix  $\mathbf{X}$ ,

Multivariate statistics deals with the relation between several random variables. The sets of observations of the random variables are represented by a multivariate data matrix  $\mathbf{X}$ ,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}. \quad (1)$$

Each column vector  $\mathbf{u}_k$  represents the data for a different variable. If  $\mathbf{c}$  is an  $p \times 1$  matrix, then

$$\mathbf{X}\mathbf{c} = c_1 \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{n1} \end{bmatrix} + c_2 \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \\ \vdots \\ x_{n2} \end{bmatrix} + \cdots + c_p \begin{bmatrix} x_{1p} \\ x_{2p} \\ x_{3p} \\ \vdots \\ x_{np} \end{bmatrix} \quad (2)$$

is a linear combinations of the set of observations.

Descriptive statistics can also be applied to a multivariate data matrix  $\mathbf{X}$ , the sample mean of the  $k$ th variable is

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}, k = 1, 2, \dots, p, \quad (3)$$

the sample variance is defined by

$$s_k^2 = \frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2, k = 1, 2, \dots, p. \quad (4)$$

Next we introduce a matrix that contains statistics that relate pairs of variables  $(x_i, x_k)$ , sample covariance  $s_{ik}$ :

$$s_{ik} = \frac{1}{n} \sum_{j=1}^n (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k), \quad i = 1, 2, \dots, p, k = 1, 2, \dots, p. \quad (5)$$

It follows that  $s_{ik} = s_{ki}$  and  $s_{ii} = s_i^2$ , the sample variance.

Matrix of sample covariances

$$\mathbf{S}_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ s_{31} & s_{32} & \cdots & s_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (6)$$

is symmetric.

**THEOREM** Let  $\mathbf{S}_n$  be the  $p \times p$  covariance matrix related to the multivariate data matrix  $\mathbf{X}$ . Let eigenvalues of  $\mathbf{S}_n$  be  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ , and corresponding orthonormal eigenvectors be  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$ . Then  $i$ th principal component  $\mathbf{y}_i$  is given by the linear combination of the original variables in the data matrix  $\mathbf{X}$ :

$$\mathbf{y}_i = \mathbf{X}\mathbf{u}_i, i = 1, 2, \dots, p. \quad (7)$$

The variance of  $\mathbf{y}_i$  is  $\lambda_i$ , and  $\text{cov}(\mathbf{y}_i, \mathbf{y}_j) = 0, i \neq j$ . The total variance of the data in  $\mathbf{X}$  is equal to the sum of eigenvalues:

$$\sum_{j=1}^p s_{jj} = \sum_{j=1}^p \lambda_j. \quad (8)$$

Proportion of the total variance covered by the " $k$ th principal component"

$$= \frac{\lambda_k}{\sum_{j=1}^p \lambda_j}. \quad (9)$$

If a large percentage of the total variance can be attributed to the first few components, then these new variables can replace the original variables without significant loss of information. Thus we can achieve significant reduction in data.

#### 4. PRINCIPAL COMPONENTS OF CARDIOVASCULAR DISEASES DATA

The information in the covariance matrix is used to define a set of new variables as a linear combination of the original variables in the data matrices. The new variables are derived in a decreasing order of importance. The first of them is called first principal component and accounts for as much as possible of the variation in the original data. The second of them is called second principal component and accounts for another, but smaller portion of the variation, and so on.

If there are  $p$  variables, to cover all of the variation in the original data, one needs  $p$  components, but often much of the variation is covered by a smaller number of components. Thus PCA has as its goals the interpretation of the variation and data reduction.

In fact PCA is nothing but the spectral decomposition of the covariance matrix. The fourteenth component of the data which is related to the absence and presence of the disease removed, and the principal transformation of the data is realized. The first two principal components of these two types of data are intermingled as seen in Figure 1. In our classification perceptron, first five principal components are found to be satisfactory.

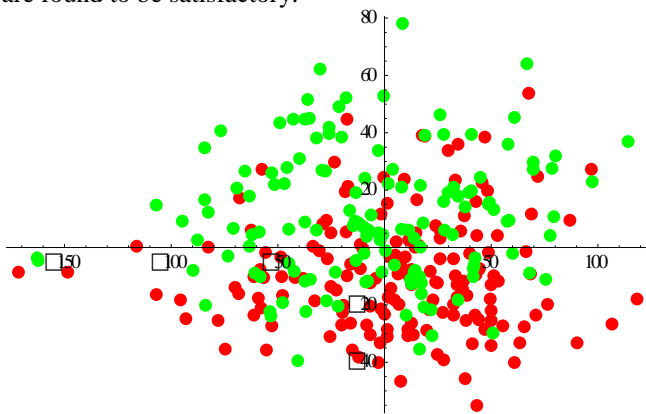


Fig. 1. The distribution obtained by the use of first two principal components of absence and presence data

#### 5. ARTIFICIAL NEURAL NETWORKS

Nervous systems existing in biological organism for years have been the subject of studies for mathematicians who tried to develop some models describing such systems and all their complexities. Artificial neural networks emerged as generalizations of these concepts with mathematical model of artificial neuron due to McCulloch and Pitts [15] described in 1943 definition of unsupervised learning rule

by Hebb [16] in 1949, and the first ever implementation of Rosenblatt's perceptron [17] in 1958. The efficiency and applicability of artificial neural networks to computational tasks have been questioned many times, especially at the very beginning of their history the book "Perceptrons" by Minsky and Papert [18], published in 1969, caused dissipation of initial interest and enthusiasm in applications of neural networks.

It was not until 1970s and 80s, when the back-propagation algorithm for supervised learning was documented that artificial neural networks regained their status and proved beyond doubt to be sufficiently good approach to many problems. Artificial Neural Network can be looked upon as a parallel computing system comprised of some number of rather simple processing units (neurons) and their interconnections. They follow inherent organizational principles such as the ability to learn and adapt, generalization, distributed knowledge representation, and fault tolerance. Neural network specification comprises definitions of the set of neurons (not only their number but also their organization), activation states for all neurons expressed by their activation functions and offsets specifying when they fire, connections between neurons which by their weights determine the effect the output signal of a neuron has on other neurons it is connected with, and a method for gathering information by the network that is its learning or training rule.

##### 5.1. Architecture

From architecture point of view neural networks can be divided into two categories: feed-forward and recurrent networks. In feed-forward networks the flow of data is strictly from input to output cells that can be grouped into layers but no feedback interconnections can exist. On the other hand, recurrent networks contain feedback loops and their dynamical properties are very important.

The most popularly used type of neural networks employed in pattern classification tasks is the feedforward network which is constructed from layers and possesses unidirectional weighted connections between neurons. The common examples of this category are Multilayer Perceptron or Radial Basis Function networks, and committee machines.

Multilayer perceptron type is more closely defined by establishing the number of neurons from which it is built, and this process can be divided into three parts, the two of which, finding the number of input and output units, are quite simple, whereas the third, specification of the number of hidden neurons can become crucial to accuracy of obtained classification results.

The number of input and output neurons can be actually seen as external specification of the network and these parameters are rather found in a task specification. For classification purposes as many distinct features are defined for objects which are analyzed that many input nodes are required. The only way to better adapt the network to the problem is in consideration of chosen data types for each of

selected features. For example instead of using the absolute value of some feature for each sample it can be more advantageous to calculate its change as this relative value should be smaller than the whole range of possible values and thus variations could be more easily picked up by artificial neural network. The number of network outputs typically reflects the number of classification classes.

The third factor in specification of the multilayer perceptron is the number of hidden neurons and layers and it is essential to classification ability and accuracy. With no hidden layer the network is able to properly solve only linearly separable problems with the output neuron dividing the input space by a hyperplane. Since not many problems to be solved are within this category, usually some hidden layer is necessary.

With a single hidden layer the network can classify objects in the input space that are sometimes and not quite formally referred to as simplexes, single convex objects that can be created by partitioning out from the space by some number of hyperplanes, whereas with two hidden layers the network can classify any objects since they can always be represented as a sum or difference of some such simplexes classified by the second hidden layer.

Apart from the number of layers there is another issue of the number of neurons in these layers. When the number of neurons is unnecessarily high the network easily learns but poorly generalizes on new data. This situation reminds auto-associative property: too many neurons keep too much information about training set rather "remembering" than "learning" its characteristics. This is not enough to ensure good generalization that is needed.

On the other hand, when there are too few hidden neurons the network may never learn the relationships amongst the input data. Since there is no precise indicator how many neurons should be used in the construction of a network, it is a common practice to build a network with some initial number of units and when it learns poorly this number is either increased or decreased as required. Obtained solutions are usually task-dependant.

## 5.2 Activation Functions

Activation or transfer function of a neuron is a rule that defines how it reacts to data received through its inputs that all have certain weights.

Among the most frequently used activation functions are linear or semi-linear function, a hard limiting threshold function or a smoothly limiting threshold such as a sigmoid or a hyperbolic tangent. Due to their inherent properties, whether they are linear, continuous or differentiable, different activation functions perform with different efficiency in task-specific solutions. For classification tasks antisymmetric sigmoid tangent hyperbolic function is the most popularly used activation function:

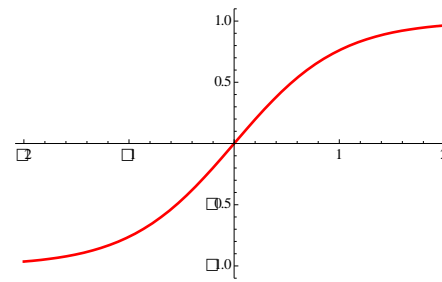


Fig. 2. Antisymmetric sigmoid tangent hyperbolic activation function

## 5.3 Learning Rules

In order to produce the desired set of output states whenever a set of inputs is presented to a neural network it has to be configured by setting the strengths of the interconnections and this step corresponds to the network learning procedure. Learning rules are roughly divided into three categories of supervised, unsupervised and reinforcement learning methods.

The term supervised indicates an external teacher who provides information about the desired answer for each input sample. Thus in case of supervised learning the training data is specified in forms of pairs of input values and expected outputs. By comparing the expected outcomes with the ones actually obtained from the network the error function is calculated and its minimization leads to modification of connection weights in such a way as to obtain the output values closest to expected for each training sample and to the whole training set.

In unsupervised learning no answer is specified as expected of the neural network and it is left somewhat to itself to discover such self-organization which yields the same values at an output neuron for new samples as there are for the nearest sample of the training set.

Reinforcement learning relies on constant interaction between the network and its environment. The network has no indication what is expected of it but it can induce it by discovering which actions bring the highest reward even if this reward is not immediate but delayed. Basing on these rewards it performs such re-organization that is most advantageous in the long run [22].

The modification of weights associated with network interconnections can be performed either after each of the training samples or after finished iteration of the whole training set.

The important factor in this algorithm is the learning rate  $\eta$  whose value when too high can cause oscillations around the local minima of the error function and when too low results in slow convergence. This locality is considered the drawback of the backpropagation method but its universality is the advantage.

## 5.4 Architecture of artificial neural networks, Committee Machines

As the base topology of artificial neural network committee machines with the feed-forward multilayer perceptron with

sigmoid activation function trained by backpropagation algorithm is used.

In committee machines approach, a complex computational task is solved by dividing it into a number of computationally simple tasks and then combining the solutions to those tasks. In supervised learning, computational simplicity is achieved by distributing the learning task among a number of experts, which in turn divides the input space into a set of subspaces. The combination of experts is said to constitute a committee machine. Basically, it fuses knowledge acquired by experts to arrive at an overall decision that is supposedly superior to that attainable by anyone of them acting alone. The idea of a committee machine may be traced back to Nilsson [19] (1965); the network structure considered therein consisted of a layer of elementary perceptrons followed by a vote-taking perceptron in the second layer.

Committee machines are universal approximators. They may be classified into two major categories:

1. *Static structures.* In this class of committee machines, the responses of several predictors (experts) are combined by means of a mechanism that does not involve the input signal, hence the designation "static." This category includes the following methods:

- Ensemble averaging, where the outputs of different predictors are linearly combined to produce an overall output.
- Boosting, where a weak learning algorithm is converted into one that achieves arbitrarily high accuracy.

2. *Dynamic structures.* In this second class of committee machines, the input signal is directly involved in actuating the mechanism that integrates the outputs of the individual experts into an overall output, hence the designation "dynamic."

In this research ensemble averaging category of committee machines will be used.

### 5.5 Ensemble averaging

Figure 1 shows a number of differently trained neural networks (i.e., experts), which share a common input and whose individual outputs are somehow combined to produce an overall output  $y$ . In this research the outputs of the experts are scalar-valued. Such a technique is referred to as an ensemble averaging method. The motivation for its use is two-fold:

- If the combination of experts in Fig. 1 were replaced by a single neural network, we would have a network with a correspondingly large number of adjustable parameters. The training time for such a large network is likely to be longer than for the case of a set of experts trained in parallel.
- The risk of overfitting the data increases when the number of adjustable parameters is large compared to cardinality (i.e., size of the set) of the training data.

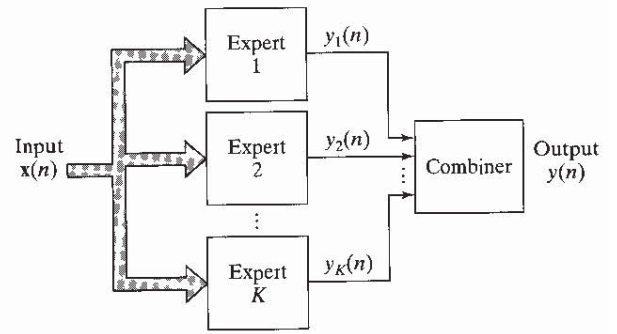


Fig. 3. Block diagram of a committee machine based on ensemble-averaging.

In any event, in using a committee machine as depicted in Fig. 3, the expectation is that the differently trained experts converge to different local minima on the error surface, and overall performance is improved by combining the outputs in some way.

The number of input terminals equaled the number of attributes in the human voice data, thus it is eleven. There are two hidden layers with eleven neurons within each of three neural networks in the committee machine for preserving generalization properties but achieving as shown in the signal flow graph in Figure 4.

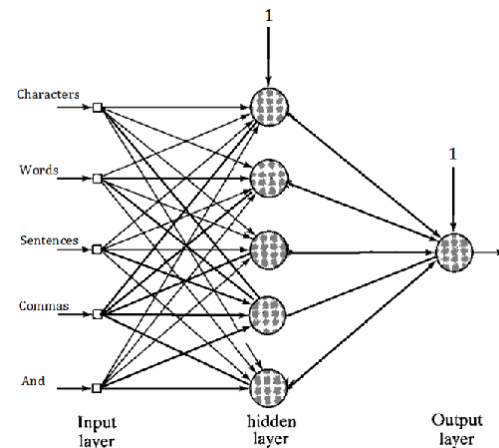


Fig. 4. Signal flow graph of an expert neural network

For preserving generalization properties but achieving convergence during training with tolerance at most 0.15 for all training samples recognized properly. Algorithm results in a decision about attribution of paragraphs whose textual description entered as inputs.

In this research two hidden layer, feed forward, back propagation artificial neural networks are used as the eleven committee machines. The Cardiovascular disease data was 14 dimensional. Therefore fourteen input ports equaled the number of fourteen attributes used, thus it is fourteen. There is one hidden layer with fourteen neurons within each of eleven neural networks in the committee machines for preserving generalization properties but achieving

convergence during training with tolerance at most 0.14 for all training samples recognized properly.

For all structures of artificial neural networks, only one output is produced. Actually, it was possible to use a single output and by interpretation of its active state as one class and inactive output state the second class the task would have been solved as well.

## 6. RESULTS AND DISCUSSION

The data is divided into two parts. The first part, training data consists of 75 absence and 75 presence data. In the training stage the eleven committee machines performances were as in Table 2.

Table 2: Performances of eleven committee machines in training.

	Committee Machines										
%	1	2	3	4	5	6	7	8	9	10	11
TP	91	92	87	91	91	89	91	89	88	89	89
TN	89	88	91	91	88	89	91	91	91	91	91
FP	11	12	9	9	12	11	9	9	9	9	9
FN	9	8	13	9	9	11	9	11	12	11	11

Then test data which consists of the remaining 88 absence and 64 presence data sent to eleven committee machines. Results from the testing experiment can be seen in table 2.

Table 3: Performance measurements of eleven committee machines and majority vote.

	Committee Machines											
%	1	2	3	4	5	6	7	8	9	10	11	MV
TP	78	78	77	80	78	77	77	77	75	80	80	86
TN	90	89	89	91	89	90	89	92	90	91	92	91
FP	10	11	11	9	11	10	11	8	10	9	8	9
FN	12	22	23	20	22	23	23	23	25	20	20	14

It has been shown in this study that parallel neural networks in combination with a majority rule based system increase performance of true recognition rates in an imbalanced data set. In both conducted experiments all measurement parameters are improved compared to single network predictions. From the two experiments it is proven the parallel system with forward propagation of untrained data samples increases the robustness and decrease the variability as seen in the system which does not have this feature.

Despite the advantages of having an accurate system prediction, the training time and complexity of the parallel network algorithm do increase as the number of parallel networks increases [20-21]. The data set is very unbalanced with regard to the class distribution. This, in combination with the small sample size, makes it difficult to train any type of classifier to predict the presence of Cardiovascular disease's disease. Out of 302 samples, 139 are with cardiovascular disease type and the remainder is of healthy character.

It implies that the baseline prediction is 46% and any prediction accuracy less than the baseline is not relevant. A common problem with imbalanced data sets is that they can increase to high false positive rates. Traditionally, the problem with false positive predictions is dealt with over- or undersampling [22]. However techniques to adjust the sample distribution sometimes overweight the benefits of generalizing the classifier. Any modification to the data set is merely artificial alternatives to the problem of inadequate training data. In this paper, it has been demonstrated that parallel neural networks are strong at adjusting the imbalanced data set problem.

False positive rates up to 25 - 30% of the positive class have been reported [23] in the literature. It has been demonstrated in this study that a true positive rate up to 86% of positive class is achieved by using eleven parallel networks. This is a significant improvement compared to previously demonstrated results. It has also evident that networks with forward propagation of untrained data do increase the robustness of the parallel system. However, it can be seen from table 2 and 3, that after a certain number of parallel networks, the overall accurate prediction does not increase which is a direct implication of an insignificant improvement of the MSE.

## 7. CONCLUSIONS

A system has been presented consisting of eleven parallel neural networks and a majority voting system. An empirical investigation demonstrates that it is possible to achieve >90% true positive rate for each class in the Cardiovascular Disease data set.

## REFERENCES

- [1]. M. Anthea, Hopkins J., C. W. McLaughlin, S. Johnson, M. Q. Warner, D. LaHart, J. D. Wright (1993) *Human Biology and Health*, Englewood Cliffs, New Jersey: Prentice Hall.
- [2]. B. B. Kelly (2010) *Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health*. Washington, D.C: National Academies Press.
- [3]. A. P. Dantas, F. Jimenez-Altayo, E.Vila (2012) Vascular aging: facts and factors. *Frontiers in Vascular Physiology* 3 (325): 1–2.
- [4]. B. B. Kelly, editor (2010) *Promoting cardiovascular health in the developing world: a critical challenge to achieve global health*. Washington, D.C.: National Academies Press. Chapter 2.
- [5]. S. Mendis, P. Puska, and B. Norrving (editors) (2011) *Global Atlas on cardiovascular disease prevention and control*.
- [6]. H.C. McGill, C.A. McMahan, S.S. Gidding (2008). Preventing heart disease in the 21st century: implications of the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study, *Circulation* 117 (9) pp. 1216–27.



- [7]. J. N. Morris, M.D. Crawford (1958) Coronary Heart Disease and Physical Activity of Work. *British Medical Journal* 2 (5111): 1485–1496.
- [8]. M. Karakas, W. Koenig (2009). CRP in cardiovascular disease. *Herz* 34 (8): 607–13.
- [9]. R. Andraws, J.S. Berger, D.L. Brown (2005). Effects of antibiotic therapy on outcomes of patients with coronary artery disease: a meta-analysis of randomized controlled trials. *JAMA* 293 (21): 2641–7.
- [10]. A. Dominguez-Rodriguez, (2012). Melatonin and Cardiovascular Disease: Myth or Reality? *Rev Esp Cardiol* 65: 215-218.
- [11]. M. L. Minsky, and S. A. Papert, (1988) *Perceptrons*, Expanded Edition. Cambridge, MA: MIT Press. Original edition, 1969.
- [12]. R. Koker (2008) Reliability-based approach to the inverse kinematics solution of robots using Elman's networks, *Engineering Applications of Artificial Intelligence* (18) 685 – 693.
- [13]. A. J. Lee (1997) Parallel neural networks for speech recognition, *Neural Networks*, 1997., International Conference on 4 (9-12) 2093–2097.
- [14]. R. Detrano ,V.A. Medical Center, Long Beach and Cleveland Clinic Foundation.
- [15]. W. S. McCulloch, and W. Pitt's. (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*, 5:115-133. Reprinted in Anderson & Rosenfeld [1988], pp. 18-28.
- [16]. B. O. Hebb, (1949). *The Organization of Behavior*. New York: John Wiley & Sons. Introduction and Chapter 4 reprinted in Anderson & Rosenfeld 1988, pp. 45-56.
- [17]. E. Rosenblatt, (1958) The Perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, vol. 65, pp. 386-408.
- [18]. H. Tang, K. C. Tan, and Z. Yi, (2007) *Neural Networks: Computational Models and Applications*, Springer-Verlag Berlin Heidelberg.
- [19]. N. J. Nilsson (1965) *Learning Machines: Foundations of Trainable Pattern-Classifying Systems*, New York: Macgraw-Hill.
- [20]. S. Haykin (1999) *Neural Networks A Comprehensive Foundation*, Second Edition, Prentice-Hall, Inc., Simon & Schuster, A Viacom Company Upper Saddle River, New Jersey.
- [21]. R. E. Schapire (1990) The strength of weak learnability, *Machine Learning*, vol. 5, pp.197-227.
- [22]. R. E. Schapire (1997) Using output codes to boost multiclass learning problems, *Machine Learning: Proceedings of the Fourteenth International Conference*, Nashville, TN.
- [23]. H. Drucker, C. Cortes, L.D. Jackel, and Y. LeCun (1994) Boosting and other ensemble methods. *Neural Computation*, vol. 6, pp.1289-1301