Diagnosis of Cardiovascular Diseases by Boosted Neural Networks

Mehmet Can
mcan@ius.edu.ba

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

Abstract
A boosting by filtering 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%. Using a boosting by filtering technique prediction accuracy increased over 80%. 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.

Age: Age is an important risk factor in developing cardiovascular diseases. It is estimated that 87 percent of people who die of coronary heart disease are 60 and older. At the same time, the risk of stroke doubles every decade after age 55. Multiple explanations have been proposed to explain why age increases the risk of cardiovascular diseases. One of them is related to serum cholesterol level.[7] In most populations, the serum total cholesterol level increases as age increases. In men, this increase levels off around age 45 to 50 years. In women, the increase continues sharply until age 60 to 65 years. Aging is also associated with changes in the mechanical and structural properties of the vascular wall, which leads to the loss of arterial elasticity and reduced arterial compliance and may subsequently lead to coronary artery disease.

Sex: Men are at greater risk of heart disease than pre-menopausal women.[8] However, once past menopause, a woman's risk is similar to a man's. Among middle-aged people, coronary heart disease is 2 to 5 times more common in men than in women. In a study done by the World Health Organization, sex contributes to approximately 40% of the variation in the sex ratios of coronary heart disease mortality. Another study reports similar results that gender difference explains nearly half of the risk associated with cardiovascular diseases. One of the proposed explanations for the gender difference in cardiovascular disease is

hormonal difference. Among women, estrogen is the predominant sex hormone. Estrogen may have protective effects through glucose metabolism and hemostatic system, and it may have a direct effect on improving endothelial cell function. The production of estrogen decreases after menopause, and may change the female lipid metabolism toward a more atherogenic form by decreasing the HDL cholesterol level and by increasing LDL and total cholesterol levels. Women who have experienced early menopause, either naturally or because they have had a hysterectomy, are twice as likely to develop heart disease as women of the same age group who have not yet gone through menopause.

Among men and women, there are differences in body weight, height, body fat distribution, heart rate, stroke volume, and arterial compliance. In the very elderly, age related large artery pulsatility and stiffness is more pronounced in women. This may be caused by the smaller body size and arterial dimensions independent of menopause.[9]

Air pollution: Particulate matter has been studied for its short- and long-term exposure effects on cardiovascular disease. Currently, PM2.5 is the major focus, in which gradients are used to determine CVD risk. For every 10 μg/m3 of PM2.5 long-term exposure, there was an estimated 8-18% CVD mortality risk.[10] Women had a higher relative risk (RR) (1.42) for PM2.5 induced coronary artery disease than men (0.90) did. Overall, long-term PM exposure increased rate of atherosclerosis and inflammation. In regards to short-term exposure (2 hours), every 25 μg/m3 of PM2.5 resulted in a 48% increase of CVD mortality risk. Additionally, after only 5 days of exposure, a rise in systolic (2.8 mmHg) and diastolic (2.7 mmHg) blood pressure occurred for every 10.5 μg/m3 of PM2.5. Other research has implicated PM2.5 in irregular heart rhythm, reduced heart rate variability (decreased vagal tone), and most notably heart failure. PM2.5 is also linked to carotid artery thickening and increased risk of acute myocardial infarction.

Pathophysiology: Population based studies show that atherosclerosis the major precursor of cardiovascular disease begins in childhood. The Pathobiological Determinants of Atherosclerosis in Youth Study demonstrated that intimal lesions appear in all the aortas and more than half of the right coronary arteries of youths aged 7–9 years.

This is extremely important considering that 1 in 3 people will die from complications attributable to atherosclerosis. In order to stem the tide education and awareness that cardiovascular disease poses the greatest threat and measures to prevent or reverse this disease must be taken.

Obesity and diabetes mellitus are often linked to cardiovascular disease, as are a history of chronic kidney disease and hypercholesterolaemia. In fact, cardiovascular disease is the most life threatening of the diabetic complications and diabetics are two- to four-fold more

likely to die of cardiovascular-related causes than nondiabetics.[11]

Screening
Screening ECGs (either at rest or with exercise) are not recommended in those without symptoms who are at low risk. In those at higher risk the evidence for screening with ECGs is inconclusive.[12]
Some biomarkers may add to conventional cardiovascular risk factors in predicting the risk of future cardiovascular disease; however, the clinical value of some biomarkers is still questionable.[13] Currently, biomarkers which may reflect a higher risk of cardiovascular disease include:

- Coronary artery calcification
- Carotid intima-media thickness
- Carotid total plaque area
- Higher fibrinogen and PAI-1 blood concentrations
- Elevated homocysteine
- Elevated blood levels of asymmetric dimethylarginine
- Inflammation as measured by C-reactive protein
- Elevated blood levels of brain natriuretic peptide (also known as B-type) (BNP)

Prevention
Currently practiced measures to prevent cardiovascular disease include:

- A low-fat, high-fiber diet including whole grains and plenty of fresh fruit and vegetables (at least five portions a day)
- Tobacco cessation and avoidance of second-hand smoke;
- No alcohol consumption; excessive alcohol intake increases the risk of cardiovascular disease.
- Lower blood pressures, if elevated, through the use of antihypertensive medications;
- Decrease body fat (BMI) if overweight or obese;
- Increase daily activity to 30 minutes of vigorous exercise per day at least five times per week;
- Decrease psychosocial stress. Stress however plays a relatively minor role in hypertension. Specific relaxation therapies are not supported by the evidence.

Routine counseling of adults to advise them to improve their diet and increase their physical activity has not been found to significantly alter behavior, and thus is not recommended.

Diet
Evidence suggests that the Mediterranean diet may improve cardiovascular outcomes. On February 25, 2013, medical researchers at the University of Barcelona, based on a five year study of 7,447 people, reported in the New England Journal of Medicine that the Mediterranean diet reduced the risk of heart disease in people at high risk by "about 30 percent".[14]

In clinical trials the DASH diet (high in fruits and vegetables, low in sweets, red meat and fat) has been shown to reduce blood pressure, lower total and low density lipoprotein cholesterol [15]and improve metabolic syndrome;[16] but the long term benefits outside the context of a clinical trial have been questioned.

The link between saturated fat intake and cardiovascular disease is controversial (see Saturated fat and cardiovascular disease controversy) and scientific studies, both observational and clinical, show conflicting results.[17] Dietary substitution of polyunsaturated fats for saturated fats may reduce risk, substitution with carbohydrates does not change or may increase risk. Increased dietary intake of Trans fatty acids significantly increases the risk of cardiovascular disease.[18]

The effect of a low salt diet is unclear with any benefit in either hypertensive or normal tensive people being small if present. A low salt diet may be harmful in those with congestive heart failure.[19]

Supplements
Evidence to support omega-3 fatty acid supplementation is lacking.[20] As is evidence to support antioxidants and vitamins.[21]

Medication
Aspirin has not been found to be of benefit overall in those at low risk of heart disease as the risk of serious bleeding is equal to the benefit with respect to cardiovascular problems.[22] Statins are effective in preventing further cardiovascular disease in those with a history of cardiovascular disease.[23] A decreased risk of death however seems to only occur in men.

Management
Cardiovascular disease is treatable with initial treatment primarily focused on diet and lifestyle interventions. Medication may also be useful for prevention.

Mortality
According to the World Health Organization, cardiovascular diseases are the leading cause of death. In 2008, 30% of all global death is attributed to cardiovascular diseases. Death caused by cardiovascular diseases are also higher in low and middle-income countries as over 80% of all global death caused by cardiovascular diseases occurred in those countries. It is also estimated that by 2030, over 23 million people will die from cardiovascular diseases annually.

Research
The first studies on cardiovascular health were performed in 1949 by Jerry Morris using occupational health data and were published in 1958.[24]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.[25] Also osteoprotegerin which involved with regulation of a key inflammatory transcription factor called NF-κB has been found to be a risk factor of cardiovascular disease and mortality.

Some areas currently being researched include possible links between infection with Chlamydophila 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.[26]

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.[27]

1.3 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 [28].

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 [29] 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 [30]. 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).[31]

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 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 X,

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

Each column vector u_k represents the data for a different variable. If c is an $p \times 1$ matrix, then

$$Xc = c_1 \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{n1} \end{bmatrix} + c_2 \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \\ \vdots \\ x_{n2} \end{bmatrix} + \cdots + c_p \begin{bmatrix} x_{1p} \\ x_{2p} \\ x_{3p} \\ \vdots \\ x_{np} \end{bmatrix} \tag{2}$$

is a linear combinations of the set of observations.

Descriptive statistics can also be applied to a multivariate data matrix X, the sample mean of the kth variable is

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^{n} x_{ik}, k = 1,2,\ldots,p, \tag{3}$$

the sample variance is defined by

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

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

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

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

Matrix of sample covariances

$$S_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ s_{31} & s_{32} & \cdots & s_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \tag{6}$$

is symmetric.

THEOREM Let S_n be the $p \times p$ covariance matrix related to the multivariate data matrix X. Let eigenvalues of S_n be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, and corresponding orthonormal eigenvectors be $u_1, u_2, \ldots, u_p$. Then ith principal component y_i is given by the linear combination of the original variables in the data matrix X:

$$y_i = Xu_i, i = 1,2,\ldots,p. \tag{7}$$

The variance of y_i is λ_i, and $\text{cov}(y_i, y_j) = 0, i \neq j$. The total variance of the data in X is equal to the sum of eigenvalues:

$$\sum_{j=1}^{p} s_{jj} = \sum_{j=1}^{p} \lambda_j. \tag{8}$$

Proportion of the total variance covered by the "kth principal component"

$$= \frac{\lambda_k}{\sum_{j=1}^{p} \lambda_j}. \tag{9}$$

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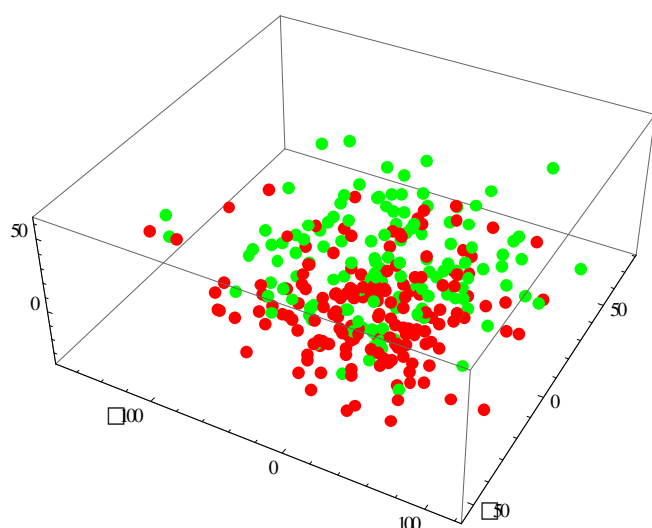
Fig. 1. The distribution obtained by the use of first three principal components of absence and presence data

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 McCuloch and Pitts [32] described in 1943 definition of unsupervised learning rule by Hebb [33] in 1949, and the first ever implementation of Rosenblatt's perceptron [34] 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 [35], 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 trains 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 thresh-old 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:
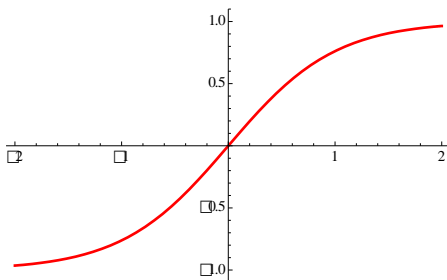

Fig. 1. Antisymmetric sigmoid tangent hyperbolic activation function

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 [34].

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 η 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 [24] 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 [36] (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."

5.5 Boosting

Boosting is a method that belongs to the "static" class of committee machines. Boosting is quite different from ensemble averaging. In a committee machine based on ensemble averaging, all the experts in the machine are trained on the same data set; they may differ from each other in the choice of initial conditions used in network training. By contrast, in a boosting machine the experts are trained on data sets with entirely different distributions; it is a general method that can be used to improve the performance of any learning algorithm.

Boosting' can be implemented in three fundamentally different ways:

1. Boosting by filtering. This approach involves filtering the training examples by different versions of a weak learning algorithm. It assumes the availability of a large (in theory, infinite) source of examples, with the examples being either discarded or kept during training. An advantage of this approach is that it allows for a small memory requirement compared to the other two approaches.

2. Boosting by subsampling. This second approach works with a training sample of fixed size. The examples are "resampled" according to a given probability distribution during training. The error is calculated with respect to the fixed training sample.

3. Boosting by reweighting. This third approach also works with a fixed training sample, but it assumes that the weak learning algorithm can receive "weighted" examples. The error is calculated with respect to the weighted examples.

In this paper Boosting by filtering is used. This algorithm is due to Schapire [36] (1990). The original idea of boosting described in Schapire (1990) is rooted in a distribution free or probably approximately correct (PAC) model of learning. To be more specific, the goal of the learning machine is to find a hypothesis or prediction rule with an error rate of at most ε, for arbitrarily small positive values of ε, and this should hold uniformly for all input distributions.

In boosting by filtering, the committee machine consists of three experts or subhypotheses. The algorithm used to train them is called a boosting algorithm. The three experts are arbitrarily labeled "first," "second," and "third." The three experts are individually trained as follows:

1. The first expert is trained on a set consisting of N_2 examples.

2. The trained first expert is used to filter another set of examples by proceeding in the following manner:
Flip a fair coin; this in effect simulates a random guess.
If the result is heads, pass new patterns through the first expert and discard correctly classified patterns until a

pattern is misclassified. That misclassified pattern is added to the training set for the second expert.

If the result is tails, do the opposite. Specifically, pass new patterns through the first expert and discard incorrectly classified patterns until a pattern is classified correctly. That correctly classified pattern is added to the training set for the second expert.

Continue this process until a total of N_1 examples has been filtered by the first expert. This set of filtered examples constitutes the training set for the second expert.

By following this coin flipping procedure, it is ensured that if the first expert is tested on the second set of examples, it would have an error rate of 1/2. In other words, the second set of N_1 examples available for training the second expert has a distribution entirely different from the first set of N_2 examples used to train the first expert. In this way the second expert is forced to learn a distribution different from that learned by the first expert [37].

3. Once the second expert has been trained in the usual way, a third training set is formed for the third expert by proceeding in the following manner:
• Pass a new pattern through both the first and second experts. If the two experts agree in their decisions, discard that pattern. If, on the other hand, they disagree, the pattern is added to the training set for the third expert.
• Continue with this process until a total of N_1 examples have been filtered jointly by the first and second experts. This set of jointly filtered examples constitutes the training set for the third expert.

The third expert is then trained in the usual way, and the training of the entire committee machine is thereby completed.

Let N_2 denote the number of examples that must be filtered by the first expert to obtain the training set of N1 examples for the second expert. Note that N_1 is fixed, and N_2 depends on the generalization error rate of the first expert. Let N_3 denote the number of examples that must be jointly filtered by the first and second experts to obtain the training set of N_1 examples for the third expert [38].
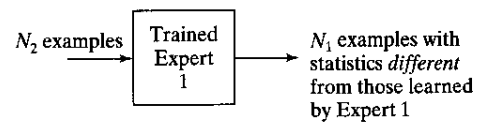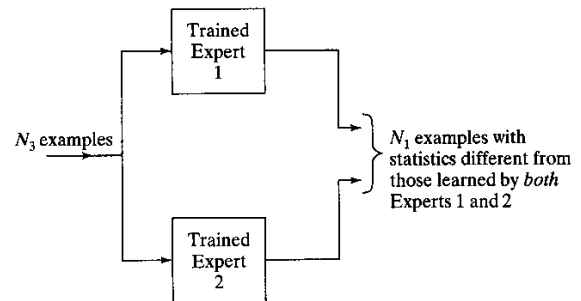

Fig. 2a.Filtering of data by Expert 1

b. Filtering of data by Expert 2 and 3
Fig. 2.The three-point filtering procedure

With N_1 examples also needed to train the first expert, the total size of data set needed to train the entire committee machine is $N = N_1 + N_2 + N_3$. However, the computational cost is based on 3N1 examples because N1 is the number of examples actually used to train each of the three experts. We may therefore say that the boosting algorithm described herein is indeed "smart" in the sense that the committee machine requires a large set of examples for its operation, but only a subset of that data set is used to perform the actual training.

Table 2. Number of samples used at each stage of the training-testing processes.

	N_1	N_2	N_3	Test
Sick	47	50	37	50
Healthy	53	50	39	50

Another noteworthy point is that the filtering operation performed by the first expert and the joint filtering operation performed by the first and second experts make the second and third experts, respectively, focus on "hard-to-learn" parts of the distribution.

During the training stage, the performances of committee members, are shown in Table 3.

Table 3. Positives, and Negatives in training stage.

	First	Second	Third
True positive %	98	94	97
True negative %	98	96	97
False positive %	2	4	3
False negative %	2	6	3

In the theoretical derivation of the boosting algorithm originally presented in Schapire (1990)[36], simple voting was used to evaluate the performance of the committee machine on test patterns not seen before. Specifically, a test pattern is presented to the committee machine. If the first and second experts in the committee machine agree in their respective decisions, that class label is used. Otherwise, the class label discovered by the third expert is used. However, in experimental work presented in Drucker et al.[39-40] (1993,1994), it has been determined that addition of the respective outputs of the three experts yields a better performance than voting. For example, in the optical character recognition (OCR) problem, the addition operation is performed simply by adding the "digit 0" outputs of the three experts, and likewise for the other nine digit outputs.

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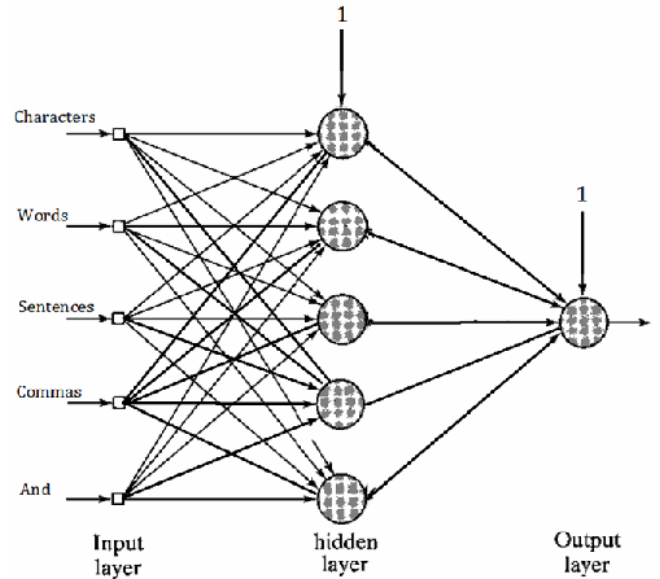

Fig. 3. Signal flow graph of each of the three expert machines with one hidden layer.

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, but with such approach the text is attributed to either one or another author and classification is binary. Algorithm results in a decision about attribution of paragraphs whose textual description entered as inputs.

6. RESULTS AND DISCUSSION

To perform the boosting by filtering technique, we the training data are chosen in a special way described in Section 3.5. A balanced set of 50-50 positive and negative members are chosen from available data for testing. During the testing stage, the performances of committee members and success in the final decision are shown in Table 4.

Table 4. False positives, and false negatives in testing stage.

	First	Second	Third	Majority
True positive %	74	80	88	82
True negative %	88	81	72	84
False positive %	12	19	28	16
False negative %	26	20	12	18

It has been shown that parallel neural networks, when boosted by filtering, in combination with a majority voting increase performance of true recognition rates in an imbalanced data set.

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.

Out of 302 samples, 46% are cardiovascular disease type and the remainder is of healthy character.

False positive rates up to 25 - 40% of the positive class have been reported [41-50] in the literature. It has been demonstrated in this study that a true negative rate up to 90% can be achieved by using three parallel networks, and majority voting. This is a significant improvement compared to previously demonstrated results

7. CONCLUSIONS

A system has been presented consisting of parallel distributed neural networks with one hidden layer, boosted by the use of filtering, and a majority voting system. The different expertise of the committee members increases the robustness of the system. An empirical investigation demonstrates that it is possible to achieve >90% true positive rate for each class in a 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. Vartiainen, T. Puska (1999) Sex, age, cardiovascular risk factors, and coronary heart disease, *Circulation* 99: 1165–1172.

[8]. http://www.world-heart-federation.org/cardiovascular-health/cardiovascular-disease-risk-factors

[9]. J. B, Rajkumar (2006) Ageing and vascular ageing, *Postgrad Med J* 82: 357–362.

[10]. M. Khallaf, (2011) The Impact of Air Pollution on Health, Economy, Environment and Agricultural Sources. *InTech.* pp. 69–92.

[11]. E. Kvan, K.I. Pettersen, L. Sandvik, and A. Reikvam (2007) High mortality in diabetic patient with acute myocardial infarction: cardiovascular co-morbidities contribute most to the high risk, *Int J Cardiol* 121: 184–188.

[12]. V.A. Moyer, U.S. Preventive Services Task, Force (2012) Screening for coronary heart disease with electrocardiography: U.S. Preventive Services Task Force recommendation statement. *Annals of internal medicine* 157 (7): 512-8.

[13]. T.J. Wang, P. Gona, M.G. Larson, G.H. Tofler, D. Levy, C. Newton-Cheh, P.F. Jacques, N. Rifai, J. Selhub, S.J. Robins, E.J. Benjamin, R.B. D'Agostino, and R.S. Vasan (2006) Multiple biomarkers for the prediction of first major cardiovascular events and death, *N. Engl. J. Med.* 355 (25): 2631–9.

[14]. R. Estruch, et al (2013) Primary Prevention of Cardiovascular Disease with a Mediterranean Diet. *New England Journal of Medicine.*

[15]. E. Obarzanek, F.M. Sacks, W.M. Vollmer et al. (2001) Effects on blood lipids of a blood pressure-lowering diet: the Dietary Approaches to Stop Hypertension (DASH) Trial, *Am. J. Clin. Nutr.* 74 (1): 80–9.

[16]. L. Azadbakht, P. Mirmiran, A. Esmaillzadeh, T. Azizi, and F. Azizi (2005) Beneficial effects of a Dietary Approaches to Stop Hypertension eating plan on features of the metabolic syndrome. *Diabetes Care* 28 (12): 2823–31.

[17]. P.W. Siri-Tarino, Sun Qi, Hu, B. Frank, and M. Krauss Ronald (2010) Saturated fat, carbohydrate, and cardiovascular disease, *American Journal of Clinical Nutrition* 91 (3): 502–509.

[18]. D. Mozaffarian, A. Aro. W.C. Willett (2009) Health effects of trans-fatty acids: experimental and observational evidence. *European journal of clinical nutrition 63 Suppl 2*: S5-21.

[19]. R.S. Taylor, K.E. Ashton, T. Moxham, L. Hooper, and S. Ebrahim, (2011) Reduced dietary salt for the prevention of cardiovascular disease, *Cochrane database of systematic reviews* (Online) (7)

[20]. S.M. Kwak, S.K. Myung, Y.J. Lee, H.G. Seo, for the Korean Meta-analysis Study, Group (2012) Efficacy of Omega-3 Fatty Acid Supplements (Eicosapentaenoic Acid and Docosahexaenoic Acid) in the Secondary Prevention of Cardiovascular Disease: A Meta-analysis of Randomized, Double-blind, Placebo-Controlled Trials, *Archives of Internal Medicine.*

[21]. S.K. Myung, W. Ju, B. Cho, S.W. Oh, S.M. Park, B.K. Koo, and B.J. Park, for the Korean Meta-Analysis (KORMA) Study, Group (2013) Efficacy of vitamin and antioxidant supplements in prevention of cardiovascular disease: systematic review and meta-analysis of randomised controlled trials, *BMJ (Clinical research ed.)* 346.

[22]. J.S. Berger, A. Lala, M.J. Krantz, G.S. Baker, and W.R. Hiatt, (2011) Aspirin for the prevention of cardiovascular events in patients without clinical cardiovascular disease: a meta-analysis of randomized trials, *American heart journal* 162(1): 115–24.

[23]. J. Gutierrez, G. Ramirez, T. Rundek, and R.L. Sacco (2012) Statin Therapy in the Prevention of Recurrent Cardiovascular Events: A Sex-Based Meta-analysis Statin Therapy to Prevent Recurrent CV Events, *Archives of Internal Medicine* 172 (12): 909–19.

[24]60. J.N. Morris, M.D. Crawford (1958) Coronary Heart Disease and Physical Activity of Work. *British Medical Journal* 2 (5111): 1485–1496.

[25]61. M. Karakas, and W. Koenig (2009) CRP in cardiovascular disease, *Herz* 34 (8): 607–13.

[26]4. 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.

[27]65. A. Dominguez-Rodriguez (2012) Melatonin and Cardiovascular Disease: Myth or Reality?. *Rev Esp Cardiol* 65: 215-218.

[28]. M. L. Minsky, and S. A. Papert, (1988) *Perceptrons,* Expanded Edition. Cambridge, MA: MIT Press. Original edition, 1969.

[29]. 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.

[30]. A. J. Lee (1997) Parallel neural networks for speech recognition, *Neural Networks*,1997., International Conference on 4 (9-12) 2093–2097.

[31]. R. Detrano ,V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/heart-disease.names

[32]. 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.

[33]. 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.

[34]. E, Rosenblatt, (1958) The Perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, vol. 65, pp. 386-408.

[35]. M. L. Minsky, and S. A. Papert, (1988) *Perceptrons,* Expanded Edition. Cambridge, MA: MIT Press. Original edition, 1969.

[36]. R. E. Schapire (1990) The strength of weak learnability, *Machine Learning*, vol. 5, pp.197-227.

[37]. R. E. Schapire (1997) Using output codes to boost multiclass learning problems, Machine Learning: *Proceedings of the Fourteenth International Conference,* Nashville, TN.

[38] Simon Haykin. Neural Network A Comprehensive Foundation, Pearson Education, Inc, 1999,ch.4, pp.156-226.

[39] H. Drucker, C. Cortes, L.D. Jackel, and Y. LeCun, Boosting and other ensemble methods. Neural Computation, vol. 6, 1994, pp.1289-1301

[40]H. Drucker, R.E. Schapire, and P. Simard, Improving performance in neural networks using a boosting algorithm, Advances in Neural Information Processing Systems, vol. 5, 1993, pp. 42-49, Cambridge, MA: MIT Press.

[41]. W. David Aha, and Dennis Kibler, "Instance-based prediction of heart disease presence with the Cleveland database"

[42]. J. W. Everhart, J. E. Dickson, W. C. Knowler, W. C. Johannes (1988) Using the ADAP learning algorithm to forecast the onset of diabetes mellitus, Proc. Symp. on Computer Applications and Medical Care, pp. 261–5.

[43] R. Setiono (1998) Techniques for extracting rules from artificial neural networks, 5th International Conference on Soft Computing and Information Systems, Iizuka, Japan.

[44] W. H. Wolberg, and O.L. Mangasarian (1990) Multisurface method of pattern separation for medical diagnosis applied to breast cytology, Proceedings of the National Academy of Sciences, USA, Volume 87, pp 9193-9196.

[45]. L. Prechelt (1994) Prolem1 - A Set of Neural Network Benchmark Problems and Benchmarking Rules, Technical Report 21/94, 76128 Karlsruhe, Germany.

[46] R. Parekh, J. Yang, and V. Honavar (2000) Constructive Neural Network Learning Algorithms for Pattern Classification, *IEEE Transactions On Neural Networks*, Vol. 11,No. 2.

[47]. R. Setiono, and L.C.K. Hui(1995) Use of quasi-Newton method in a feedforward neural network construction algorithm, *IEEE Transactions on Neural Networks*, vol. 6, no. 1, pp. 273-277.

[48] Rüdiger W. Brause,"Medical Analysis and Diagnosis by Neural Networks", J.W. Goethe-University, Computer Science Dept., Frankfurt a. M., Germany.

[49] R. Setiono and H. Liu (1996) Improving Back-propagation Learning with Feature selection", *Applied Intelligence*, Vol. 6, No. 2, pp. 129-140.

[50]. S. M. Kamruzzaman, A.R. Hasan, A.B. Siddiquee, and Md. E. H. Mazumder (2004) Medical diagnosis using neural network, 3rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh.