DATA SCIENCE

EFFECTIVE DIAGNOSIS OF HEART DISEASE PRESENCE USING ARTIFICIAL NEURAL NETWORKS

Sandra Cako[1],
Angelina Njeguš[1],
Vladimir Matić[1]

[1]Singidunum University,
Belgrade, Serbia

Abstract:

Due to high complexity of decision making in medicine, it has been proven that usage of Neural Networks is in the cope with the aforementioned problem. Regarding the variety of the symptoms, one of the biggest challenges is heart disease. This research has shown that, depending on the symptoms, Multilayer Perceptron Classifier can effectively decide whether the patient is suffering from heart disease or not. Main goal of this paper is to determine the proper parameters setting for the Multilayer Perceptron algorithm in order to predict heart disease with higher accuracy. However, in order to compare the obtained results using MLP, the experiment is also done using kNN, and LDA algorithms. The results confirm that recognition rate of 96.67%, when using MLP, outperforms other methods when processing heart disease data.

Keywords:

Artificial Neural Network, Heart Disease, Multilayer Perceptron, Decision Making in Medicine, Deep Learning.

1. INTRODUCTION

Nowadays, human population (not mentioning all other spheres aside) is facing the enormous speed growing of digital age, and upcoming era of Big Data, Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL). Starting from the bottom, back in the 1950., the man who made fundament research about AI, Alan Turing, gave a chance to humanity to continue expanding this area and to encounter vast possibilities of this science [1]. Teaching machines to behave like humans, think like humans, act like humans, make decisions like humans but with little chance for failure, for sure is the best challenge and unavoidable bright future that is coming. Machines become smart, and tend to overcome humans. Therefore, AI is intertwined with computer science, finance, hospitals and medicine, heavy industry, transportations, games and toys, aviation and many other areas.

The big step ahead in AI applications is in healthcare, due to the fact that is the most important thing – to save other people lives. While intelligence-based medicine is growing, we are facing possibilities starting from healthcare systems with "online doctors" which instantly tell you which drug is most suitable for your problem, to predicting many

Correspondence:
Angelina Njeguš

e-mail:
anjegus@singidunum.ac.rs

DOI: 10.15308/Sinteza-2017-3-8

diseases, and at the end to detect gene mutation which are triggers to cancer occurrence and may save you years of life, and even get you healed [2].

Many researchers try to give contribution to this field. Durairaj et al. (2015) try to predict the existence of heart disease using neural network algorithm with back propagation [3]. On the around 13 medical attributes from Cleveland dataset, they applied four different feed-forwarded back propagation training functions, and obtained results are, 80.13%, 82.15%, 93.26%, and 96.29%, respectively.

Abushariah et al (2014) tried to develop heart disease diagnosis system based on two approaches: MLP, and Adaptive Neuro-Fuzzy Inference Systems. These algorithms were applied on Cleveland dataset. The best accuracy that system achieved for each approach, were 87.04%, and 75.93%, respectively [4].

Sunila et al (2012) have applied ordinary and improved MLP algorithm on different medical dataset in order to design decision support system for cardiovascular heart disease diagnosis. The results obtained with improved MLP on Cleveland, Hungarian, and Switzerland dataset were 82.8%, 80.73%, and 93.49%, respectively [5].

Olaniyi et al (2015), applied MLP, and Support Vector Machine (SVM) on medical dataset in order to develop an intelligent system that would prevent misdiagnosis in heart diseases. The obtained results are 85% with MLP and 87.5% with SVM [6].

Wadhonkar et al (2015) also used MLP on different medical datasets in order set the architecture for the classification of the heart disease. Datasets used are Cleveland, Hungarian, Switzerland, and Long Beach V.A. The obtained performances for the MLP classifier with 10% used of testing data were 96.29% [7].

In this paper we have applied MLP, k-Nearest Neighbors (kNN), and Linear Discriminant Analysis (LDA) on Cleveland dataset in order to find the best possible classifier for prediction the heart disease presence. Paper also discusses new trends in the field of deep learning algorithms. Furthermore, the adequate set of MLP parameters for best accuracy is given.

The remainder of the paper is structured as follows: In Section II, deep learning in neural network algorithms are discussed; In the Section III, the reasons why MLP gives the best results when predicting the presence of heart anomalies, are explained; Section IV represents the experiment results, and within Section V conclusion and final remarks are given.

2. DEEP LEARNING IN NEURAL NETWORKS

When thinking about the future trends and consequences of deep learning algorithm applications, the first question that appears in our minds is "What is the best possible way to train the machines to learn in order to predict with the higher accuracy?". Inputting huge collections of structured data into machine, applying specific algorithms, and letting machine to mine the data and learn form history of patterns in order to predict, adapt, and act autonomously, is the primary idea of ML, as the part of AI.

Speaking about medicine, not so inappreciable fact is if we give a machine a large amount of patient's disease symptoms, and data collection about all possible diseases, machine will take time and learn by applying diverse algorithms on mentioned data. Imagine the great outcome it could be. Is it way better than, e.g. cardiologist reacting by intuitive for patient experiencing heart attack when there is no time, or struggling with just simple ventricular tachycardia caused by heart disease, such as a congenital heart defect? [8]

Take one step further. By far, we have one missing puzzle, to complete the described scenario. The breakout in 2016. is taking the lead over all, it is fast-growing and called Deep Learning. DL is chasing the AI using the benefits of ubiquitous ML applied with diverse algorithms that attempt to extract abstractions from data [3]. DL faces the challenge to simplify every complex input in manner that the outcome depicts the same result as it would human reproduction. DL trains Neural Networks (NN) using certain set of techniques [9].

From a biological point of view, human brain is composed of neurons, which are strongly interconnected at synapses. As they accept impulse to one entry, process it and carry out on the other side, we can do many tasks such as recognize picture, move body or hear voices. Seeing picture over and over again, including emotions within, make our synaptic points stronger which further implies that we got our neurons trained to react on a certain picture producing related emotion [10]. That's where NN idea is originating – simulating human brain (Fig. 1).

Artificial Neural Networks (ANN) tend to replace neurons and dendrites, with computer powerful tools such as CPUs and GPUs. Training NN means to apply specific algorithms by oscillating with synaptic weights in order to get realistic output [1]. Our NNs can have different number of neurons, synapses and the impulse directions can differ one from another. Thus, NN architecture is numerous and various. First, basic architecture

is Feed Forward Single Layer, followed by Feed Forward Multilayer and Recurrent NNs. Each of these architectures are characterized by layers. First, input layer, second, output layer and the layers between called hidden layers where diverse operations occur. Every layer has, synaptic weight of each neuron, and bias. Synaptic weight refers to connection intensity of nodes, while Bias is a skew that is added to input and every other hidden layer, and allows us to better predict data fit, giving a flexibility to the model. Considering problem where all input values are equal to zero, without bias any NN would not be able to fit the data. The parameter between inputs and activation function is Sumator. Sumator is used to calculate all parameters before they are proceeded to Activation function. Activation function is triggered if value exceeds threshold, so the signal is going to be sent [8]. Figure 2 illustrates the principle that has been described.

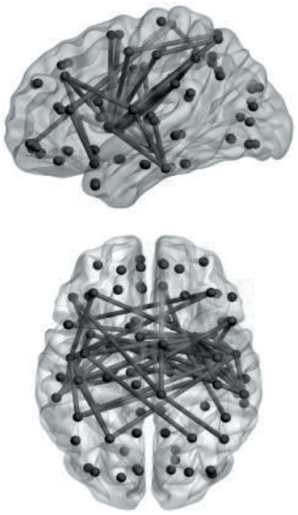

Fig. 1. Interconnections between neurons in human brain

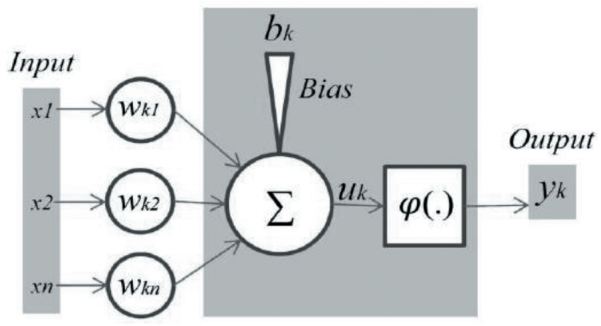

Fig. 2. Nonlinear perceptron model

First activation function – step or heaviside functions are used for binary representations, since they can have values [1,0]. For biological neurons we need values among 0 and 1, where Sigmoid function takes the case.

Gradient descent (GD) or method of steepest descent is an optimization algorithm to find local minimum of function, in a way of going one step proportional of positive gradient at the point, till reaching local minimum (Fig. 3).

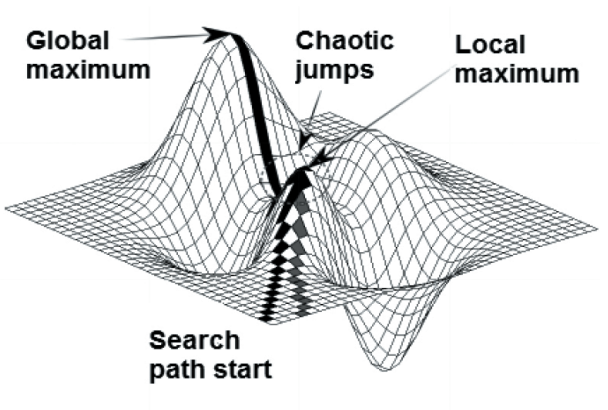

Fig. 3. Local minimum and maximum of a function

Negative gradients are orthogonal to curves at every step. The opposite method which counts local maximum is Gradient Ascent (GA). Let's take one well-known example towards minimum and maximum. Hill Climbing (HC) explains that if we start climbing on the mountain, and reach the first peak or foothill, GD and GA helps us not to be blinded about the real top or downhill which in fact, we are seeking for [1].

These methods are combined with Back propagation algorithm (BP) which is algorithm for fast training NNs. Since there was always a problem how to fit the values of bias and synaptic weights, BP actually calculates and changes biases and synaptic weights, in order to get closer to realistic outcome, thus learning the NN. It can also give us possibility to see the behavior of network, by setting other values [9].

3. HEART DISEASE PREDICTION USING MULTILAYER PERCEPTRON CLASSIFIER

Multilayer Perceptron (MLP) fits the best to problem that we are researching because it is of great importance to properly classify presence of any kind of anomalies, with respect to symptoms affecting people. It represents

form of supervised learning where function $f(\cdot) : R^m \rightarrow R^o$ learns from input which is determined by parameter *m*, that stands for number of dimensions for input, and give the output presented by parameter o that is number of dimensions for output. MLP can learn depending on input features $X=x_1,x_2,\ldots,x_m$ and output targets y , either to classify or make a regression of non-linear function. Subsequently, it has a capability to learn models in real time (online learning) [11]. Classification here identify the targets or class label of an object [12]. Class MLPClassifier implements MLP algorithm which trains NN using BP.

Considering multiple possibilities of adapting MLP object, each parameter gives variety of solutions regarding your desirable output. It consists of hidden_layer_sizes, activation, solver, alpha, batch size, learning rate, max_iter, random state, shuffle, tol, learning_rate_init, power_t, verbose, warm_start, momentum, nestrovs_momentum, early_stopping, validation_fraction, beta_1, beta_2 and epsilon, respectively.

Parameter hidden_layer_sizes represents matrix of values for hidden layers and for number of neurons in each hidden layer. Important part of every NN is choosing the most suitable algorithm for weight optimization, which we can define with solver parameter. Algorithm can be LBFGS, Stochastic Gradient Descent or Adam [11]. In order to prevent over fitting, regularization is a very meaningful method in ML [13]. From mathematical point of view, it adds a regularization term in order to avert the coefficients to overfit. Following, alpha represents the sum of the weights. To determine state for random number generator we use aforementioned parameter random state.

4. EXPERIMENTS AND RESULTS

For the purpose of this research, we have used Cleveland database that has 303 instances and 76 attributes, where 13 of them are actually utilized, and 14th represents the predicted value. Table 1. describes 14 attributes, respectively.

Attribute name	Description
Age	Age in years
Sex	1 = male; 0 = female
Cp	Chest pain type: Value 1 = typical angina Value 2 = atypical angina Value 3 = non-anginal pain Value 4 = asymptomatic
Trestbps	Resting blood pressure (in Hg)
Chol	Serum cholestoral in mg/dl
Fbs	Fasting blood sugar > 120mg/dl; 1 = true, 0 = false
Restecg	Resting electrocardiographic results: Value 0 = normal Value 1 = having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) Value 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria
Thalac	Maximum heart rate achieved
Exang	Exercise induced angina (1 = yes; 0 = no)
Oldpeak	ST depression induced by exercise relative to rest
Slope	The slope of the peak exercise ST segment: Value 1: upsloping Value 2: flat Value 3: downsloping
Ca	Number of major vessels (0-3) colored by flourosopy
Thal	3 = normal; 6 = fixed defect; 7 = reversable defect
Num	The predicted attribute

Table 1. Description of dataset attributes

Number of instances that were tested are 261, from which 76% is used for training set and 10% for test set. Training set was used for learning NN. Process of the experiment had this flow: Firstly, two arrays were generated, one of them is 2D array that represent training set which has 231 elements with 13 features. The second 2D array consists of 30 elements and stands for test set. Targets were defined in another 2D array with 2 features, for the purpose of mapping training set with his target pair. First feature is coded with [0,0] and gives the result of a patient which has no disease, where second feature coded [1,1] gives the result of a patient suffering from disease. Following, MLP Classifier instance was set with certain parameters in order to train NN. Training set was mapped with targets using the fit function. After fitting, function predict was used for foreseeing outputs. The output field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 1 (has presence) [14].

For this experiment we have used open source scikit-learn [15] from which we have used machine learning algorithms developed in Python Notebook. For running the code we have used Anaconda platform [16].

In order to fully confirm good performances that MLP has achieved, we have also applied Linear

Discriminant Analysis (LDA) and k-Nearest Neighbors (kNN) algorithms to the same dataset. LDA is most commonly used statistical technique in data classification and dimensionality reduction [17]. Fundamental concept of LDA is finding a linear combination of predictors (targets) that separates two classes in a best way. After applying LDA on training set we got results giving 88,25% accuracy. Test set with randomly chosen values gave 93,33% accuracy, which confirmed effectiveness for classification problems. Following, we continue our research targeting kNN algorithm. Entire training set is model for kNN. For unseen data, kNN search through existing data finding the k-most nearest neighbors. Accordingly, the major feature of kNN is distance measuring, without guessing anything about input data. Hence, value of parameter k could vary depending on approximation of input data. Therefore, we set *k* to value 1 which appears to be the most suitable, and got result with 74,49% of accuracy.

Based on obtained results we can conclude that used classification algorithms show their full potential on smaller datasets with smaller number of targets. Regarding MLP precision on larger datasets, it gives the best results even if you have more targets. The percentage of accuracy that we got according to dataset that we use for the purpose of this paper is 96,67%. Classification and results are shown in Figure 4, and 5, respectively.

Figure 4 represents training data which represents patients that have disease or not. There are also imported data from test set which represent patients for whom we need to determine whether they are suffering from heart disease or not.

Figure 5 predicted and classified data. From the previous parameter "indeterminate" served for deriving new results which were correctly classified as healthy or sick. Also, there were points of wrongly classified patients, which showed where algorithm had made a mistake.

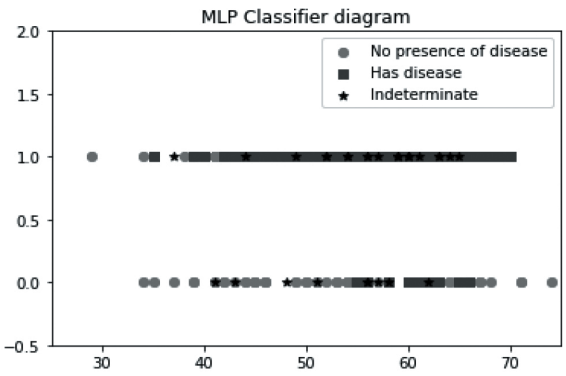

Fig. 4. MLP Classifier diagram representing training set data and test data without mapping values with targets

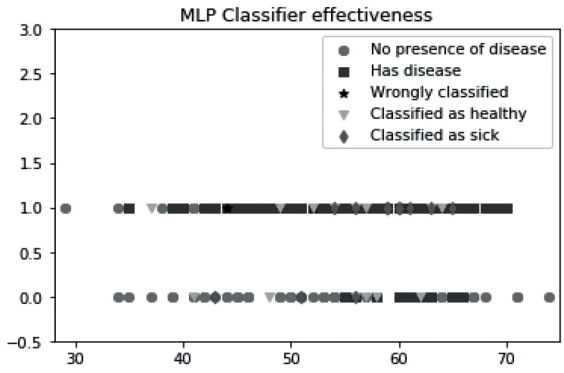

Fig. 5. MLP Classifier effectiveness with all classified data shown

5. CONCLUSION

Lately, the field of deep learning is progressively advancing. Researchers try to find the best possible algorithm, that will achieve best performance and provide even better accuracy. It seams that goal is to develop algorithm that is better of human mind. This fact can scare us, but when we take into account that medical errors are the third-leading cause of death, then we are more then thrilled with a thought of perfect mind (machine) that never makes mistakes. In this paper we tried to contribute in the field of heart disease diagnosis, therefore we have applied several algorithms on the Cleveland dataset. In order to compare our results, in the first part of the paper we have analyzed previous studies in this domain. So far the best accuracy that is achieved is 96.29% with MLP when applied on Cleveland dataset. In this paper we also try to define the best parameters for MLP algorithm in order to achieve best performance. According to the obtained results we can confirm that the neural network algorithm MLP outperformed previous results, and other methods, with the recognition rate of 96.67%.

REFERENCES

[1] Milosavljević, M. (2015) Artificial Intelligence. Singidunum University, Belgrade, Serbia.

[2] Russel, A., Norvig, P. (1995) Artificial Intelligence - A modern Approach. Prentice Hall, Englewook Cliffs, New Jersey.

[3] Durairaj, M., Revathi, V. (2015) Prediction of heart disease using back propagation MLP algorithm. International Jourlan of Scientific and Technology Research, 4(8), 235-239.

[4] Abusharah, M.A.M, Alqudah, A.A.M., Adwan, O.Y., Yousef, R.M.M. (2014) Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches. Journal of Software Engineering and Applications, Vol. 07, No. 12.

[5] Sunila, Panday, P., and Godara, N. (2012) Decision Support System for Cardiovascular Heart Disease Diagnosis using Improved Multilayer Perceptron. International Journal of Computer Applications 45(8):12-20, May 2012.

[6] Olaniyi, E.O., Oyedotun, O.K., Adnan, K. (2015) Heart Diseases Diagnosis Using Neural Networks Arbitration. International Journal of Intelligent Systems and Applications (12), 75-82.

[7] Wadhonkar, M.B., Tijare, P.A., Sawalkar, S.N. (2015) A data mining approach for classification of heart disease dataset using neural network. International Journal of Application or Innovation in Engineering and Management. Vol. 4, Iss. 5, 426-433.

[8] Doorn, J. (2014) Analysis of Deep Convolutional Neural Network Architectures. 21th Twente Student Conference on IT. University of Twente. Enschede, The Netherlands.

[9] Nielsen, M.A. (2015) *Neural Networks and Deep Learning*. Determination Press.

[10] Drakos, N. (2004) Computer Based Learning Unit. University of Leeds.

[11] Pedregosa et al. (2011) Scikit-learn: Machine Learning in Python,, JMLR 12, pp. 2825-2830, 2011.

[12] Nieminen, P. (2010) Classification and Multilayer Perceptron Neural Networks. Department of Mathematical Information Technology University of Jyväskylä , Data Mining Course (TIES445), Nov 23, 2010.

[13] Nasef, A. Marjanovic-Jakovljevic, M., Njegus, A. "Stochastic gradient descent analysis for the evaluation of a speaker recognition", *Analog Integrated Circuits and Signal Processing* (2017) 90:389-397, Springer, 2017.

[14] Janosi et al., Hungarian Institute of Cardiology. Budapest (Available at: http://archive.ics.uci.edu/ml/datasets/Heart+Disease)

[15] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

[16] Anaconda. Continuum analytics. (Available at: https://www.continuum.io)

[17] S. Balakrishnama, A. Ganapathiraju, Linear Discriminant Analysis - A Brief Tutorial, Institute for signal and information processing, Department of Electrical and Computer Engineering Mississippi State University.