A Machine Learning Approach on Classifying Orthopedic Patients Based on Their Biomechanical Features

Kamrul Hasan, Safkat Islam, Md. Mehfil Rashid Khan Samio, Amitabha Chakrabarty,
Department of Computer Science and Engineering
BRAC University, Dhaka, Bangladesh
Email: {khasan550, sazid439, mehfil01}@gmail.com, amitabha@bracu.ac.bd

***Abstract*—A person's orthopedic health condition can be detected from his biomechanical features. Now a days, disease prediction can be done automatically. Application of machine learning algorithms in medical science is not new. Different algorithms are applied to detect diseases and classify patients accordingly. This paper aims to assist specialists to predict the type of orthopedic disease. In this paper we have applied various machine learning algorithms to find out how each algorithm performs to detect and classify orthopedic patients. Each of the patient in the dataset is represented by six biomechanical attributes derived from the shape and orientation of pelvis and lumbar spine. We performed our operation in two stages and got an average accuracy of more than 90 percent for most of the algorithms, whereas Decision Tree (DT) algorithm stood out from the rest providing 99 percent accuracy.**

***Keywords*—Algorithm, Biomechanical, Classification, Disease, Health, Feature, Orthopedic, Machine Learning, Prediction.**

I. INTRODUCTION

Machine learning has been implemented in various medical fields and proven to be very accurate in classifying and predicting diseases. Use of machine learning is spreading widely with the growth of medical data in medical field to improve medical service and diagnosis of diseases. Biomechanics is the study of the movement of living beings using the science of mechanics. According to the science of mechanics, motion is created by force. Living beings create motion using force. Biomechanics essential to understand how living things make moves [1]. The condition when the gel like material (Nucleus Pulposus) get squeezed out through fractures in outer wall of intervertebral disc is known as Lumbar Disc Herniation [2]. Fig. 1 shows an example of Disk herniation. Spondylolisthesis is a medical condition in which one of the bones of a person's back (vertebrata) slides forward over the bone below it. Most of the time it occurs in lower spine or lumbosacral area [3]. Fig. 2 shows an image of Spondylolisthesis, where it can be seen that upper disk has moved forward with respect to lower disk. Both conditions can squeeze the spinal cord or nerve roots and cause pain. Both of these diseases may cause similar types of pain but they are different. Depending on the changes they make in a normal person's biomechanical characteristics, the disease can be predicted. This paper shows the classification of orthopedic patients based on these biomechanical features which are represented in the dataset we used. The classification is done in two stages where firstly, the system checks if the patients are in normal condition or not. If they are not in normal condition, then the system classifies them in two different categories of diseases. The objective of this paper is to find a suitable and accurate algorithm to enhance medical diagnosis by predicting diseases in an automated way.

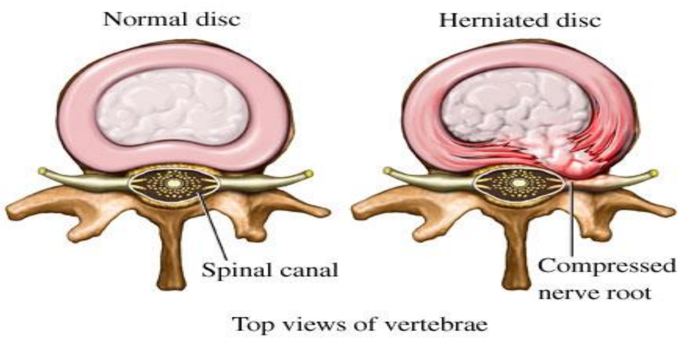

Fig. 1 Disk Herniation

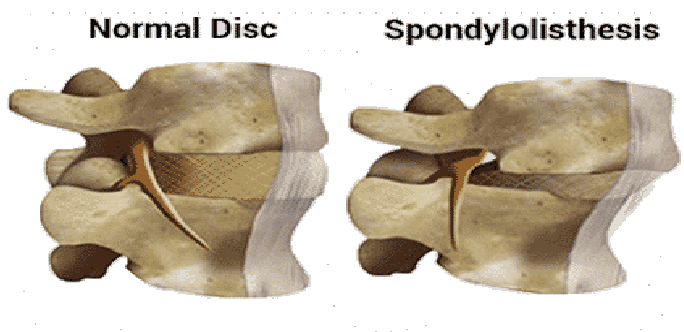

Fig. 2 Spondylolisthesis

II. LITERATURE REVIEW

For the past few years, different algorithms of machine learning has been applied to predict and classify different sorts of diseases. Algorithms like K-Nearest Neighbor (k-NN), Logistic Regression (LR), Random Forest (RF) and Support Vector Machine (SVM) were used to classify cardiac arrhythmia where SVM out performed others [4]. Again, Random Forest was used to classify Pulmonary Tuberculosis and Sarcoidosis where it provided a very good accuracy [5]. SVM has been used to classify the types of Leukemia [6]. Machine learning techniques are being used extensively to detect diseases in easier way. Algorithms like Artificial Neural Network (ANN) and Probabilistic Neural Network (PNN) have been compare to predict osteoporosis where ANN provided better results than PNN [7]. Application of ANN can be found in many medical fields including radiology [8], oncology [9], urology [10] - [13], cardiology [14]. ANN has been proven to be very useful to develop existing medical technique [15] - [18].

III. METHODOLOGY

A. Supervised Learning

Maximum practical machine learning uses supervised learning. In supervised learning we have input variable(X), output variable (Y) and then we use an algorithm to learn the mapping function from input and output.

$$Y = f(X) \tag{1}$$

The goal is to approximate the mapping function so well that when there is new input data (x), the algorithm can predict the output variables (Y) for that data. It is called supervised learning because the learning process of an algorithm from the training dataset can be thought of as a teacher supervising the learning process, where the teacher knows the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance [19].

B. Learning Algorithms

Decision Tree is a widely used classifier. It partitions data into subsets. The partition continues until there is no partition possible. The partition is done with Binary Split. The main purpose of Decision Tree is to shrink the training dataset in the smallest tree [20]. DT uses the entropy function for characterizing impurity of a dataset.

$$entropy(dataset) = -(p_+ * \log_2(P-)) \tag{2}$$

here, P= is the ratio of elements of each level in the set.

The equation for information gain given below,

$$IG(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - remainder(A) \tag{3}$$

Here, p= positive examples and n= negative examples.

$$remainder(A) = \sum_{i=1}^{v} \frac{p_i+n_i}{p+n} I(\frac{p_i}{p+n}, \frac{n_i}{p+n}) \tag{4}$$

A has ‘v’ distinct values.

Most of the algorithms in machine learning are parametric but K-Nearest Neighbor is non-parametric. It classifies object based on majority of vote it gets from its closest neighbors [21]. This algorithm predicts the class by using the Euclidean distance. The value of ‘K’ influences result. Smaller value of k means noise will have higher influence on result. On the other hand large value of ‘K’ increases the computation cost. K is selected using the following equation,

$$K = n^{\frac{1}{2}} \tag{5}$$

Here, n= number of features.

Quadratic Discriminant Analysis is closely related to Linear Discriminant Analysis. For the estimation of parameters needed in quadratic discriminant analysis, more computation and data is required. QDA can be derived from simple probabilistic models which model the class conditional distribution of the data $P(X|y = k)$ for each class k. Predictions can then be obtained by using Bayes’ rule,

$$P(y = k|X) = \frac{P(X|y=k)P(y=k)}{P(X)} \tag{6}$$

Naïve Bayes is a probabilistic classifier based on Bayes' Theorem. It utilizes the independent attributes of the data set to make main assumption [21]. Independent attributes is also the main point to make predictions. Bayes’ theorem states that,

$$P\,(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{7}$$

Here, P(A|B)= posterior probability of class given predictor, P(B|A)= likelihood which is the probability of predictor given class , P(A)= class prior probability of the class , P(B)= predictor prior probability.

Adaptive boosting classifier combines weak classifier algorithms in order to form strong classifier. It is an ensemble classifier. It can be used in conjunction with many other types of learning algorithms to improve performance. It uses the following equation for choosing the final classifier,

$$h_s(x) = sign(\sum_{t=1}^{T} w_t h_t(x)) \tag{8}$$

Here, h_t=the output of the weak classifier't' and w_t=weight applied to classifier't'. Linear combination of all the weak classifiers is the final output. The final decision is made according to the 'sign' of the sum. The classifiers are trained one at a time. After that output weight is computed using the equation given below [22],

$$w_t = \frac{1}{2}\log(\frac{1}{\epsilon_t} - 1) \tag{9}$$

ϵ_t= is the number of misclassification over the training dataset divided by size of training set.

After weight computation the training example weights are updated using the formula given below [22],

$$D_i^{(t+1)} = \frac{D_i^{(t)}\exp(-w_t y_i h_t(x_i))}{\sum_{j=1}^{m} D_j^{(t)}\exp(-w_t y_j h_t(x_j))} \tag{10}$$

D_i= is a vector of weights, t= is the training example number.

Multilayer perceptron is a supervised learning algorithm. MLP has multiple nodes arranged in interconnected layers named input, hidden and output layers [20]. MLP utilizes back propagation for training.

Support Vector Machine is a supervised Machine Learning algorithm. Basic concept of this algorithm is finding a hyper plane in order to classify the datasets [21]. The training algorithm of SVM builds a model that assigns new examples to one category or the other which makes SVM a non-probabilistic binary linear classifier.

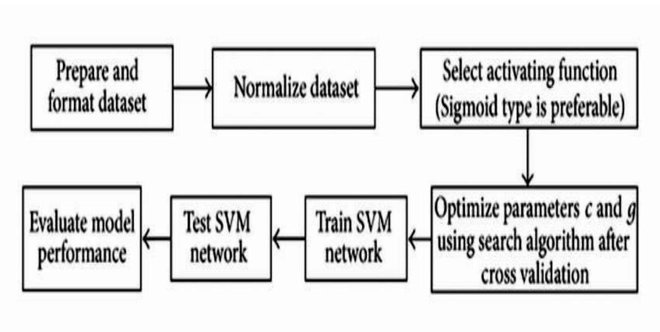

Fig. 3. SVM workflow

Random Forests minimizes the variance which might cause error in decision tree. A set of decision trees are created from randomly selected subset of training set. After creating the set, it aggregates the votes from different decision trees to choose the final class of the test object. Random forest uses Gini index for deciding the final class of each tree. If data sets T contains examples from n classes Gini index, Gini (T) is defined as,

$$Gini(T) = 1 - \sum_{j=1}^{n}(Pj)^2 \tag{11}$$

Here, T= is the dataset, n= number of classes, Pj= is the relative frequency of class j in T

Among all the algorithms borrowed from the field of statistics, Logistic Regression is the most popular. For classification problems, it is the must use algorithm. Logistic Regression searches the whole datasets to find the hyper plane which fits the most for identifying the classes [19].

Gaussian Process is a non-parametric classification method. GPs can be applied to integration, global optimization, and mixture of expert model, unsupervised learning models and more.

$$P(y(x)|t_{n,}\ X_n) = \frac{P(t_n|y(x),(X_n)P(y(x))}{P(t_n,X_n)} \tag{12}$$

Here, P (y (x) | t_n, X_n) is the probability of the target values which is given in the function y(x). P(y(x)) is the prior distribution on function which is assumed by the model. For predicting the future values of 't' only assumed prior P(y(x)) is needed.

C. Dataset

The dataset we used in this study contains 620 instances, each containing six features named pelvic incidence, pelvic tilt numeric, lumber lordosis angle, sacral slope, pelvic radius and degree spondylolisthesis. All of these six features are necessary to determine three dimensional positions. Pelvic tilt, Pelvic incidence, Sacral slope, Degree Spondylolisthesis are important parameters to identify Spondylolisthesis. The rest are needed to diagnose Disk herniation. Moreover, Pelvic incident is the result of the addition of pelvic tilt and sacral slope, which is again needed to measure herniation. Each feature has been used as the column of the dataset and the dataset has been converted into a Comma Separated Values (CSV) file. The whole dataset was divided into two parts, both of these parts has been used as the input of the system for both of our classification analysis. Table I and II shows a sample of the dataset used in the system. The class contains four values, "Normal", "Abnormal", "Hernia" and "Spondylolisthesis".

Table I. Sample of Dataset (i)

Sl.	pelvic incidence	Pelvic tilt numeric	lumber lordosis angle
1.	63.02782	22.55259	39.60912
2.	39.05695	10.06099	25.01538
3.	68.83202	22.21848	50.09219
4.	38.50527	16.9643	35.11281
5.	54.92086	18.96843	51.60146
6.	44.36249	8.945435	46.9021
7.	49.71286	9.652074	28.31741
8.	40.2502	13.92190	25.12495
9.	53.43293	15.86433	37.16593
10.	44.52905	9.433234	52.28371
11.	77.69058	21.38064	64.42944
12.	72.56070	17.38519	52.00056

Table II. Sample of Dataset (ii)

sacral slope	pelvic radius	degree spondylolisthesis	Class
40.47523	98.67292	-0.2544	Abnormal
28.99596	114.4054	4.564259	Abnormal
46.61354	105.9851	-3.53032	Abnormal
21.54098	127.6329	7.986683	Normal
35.95243	125.8466	2.001642	Normal
35.41706	129.2207	4.994195	Normal
40.06078	108.1687	7.918500	Hernia
26.32829	130.3279	2.230651	Hernia
37.56859	120.5675	5.988550	Hernia
35.09582	134.7118	29.10657	Spondylolisthesis
56.30993	114.8188	26.93184	Spondylolisthesis
55.17551	119.1937	32.10854	Spondylolisthesis

IV. RESULT

In our study, we have used ten classifiers named Decision Tree (DT), K-Nearest Neighbor (k-NN), Naive Bayes (NV), Adaptive Boosting (ADB), Support Vector Machine (SVM), Random Forest (RF), Quadratic Discriminant Analysis (QDA), Multi-Layer Perception (MLP), Logistic Regression (LR) and Gaussian Process (GP). We have used python and its related packages for classification. To train the system, we have used about 30% of the data. The whole dataset was divided into two parts. For the first analysis we used a dataset with 434 instances and for our second analysis we used a dataset with 186 instances. Firstly, we tested the system to check if the patient is in normal state or not. If the patient is not in normal condition, then we tested what the disease is. Accuracy score and Confusion matrix were used as metrics to calculate the performance of the algorithms.

Table III. Confusion matrix.

n	No (Prediction)	Yes (Prediction)
No (Actual)	True Negatives (TN)	False Positives (FP)
Yes (Actual)	False Negatives (FN)	True Positives (TP)

Where, True Positives (TP) = these are cases in which algorithms predicted yes (they have the disease), and they do have the disease. True Negatives (TN) = these are cases in which algorithms predicted no, and they don't have the disease. False positives (FP) = these are the cases in which algorithms predicted yes, but they don't actually have the disease. False Negatives (FN) = these are the cases in which algorithms predicted no, but they actually do have the disease. n = total number of instances. From each algorithms confusion matrix, we used the following equation to derive the accuracy of each algorithm,

$$\text{Accuracy} = (TP + TN)/n \qquad (13)$$

We have performed our both analysis for ten machine learning techniques mentioned before. In this part, we reveal the output of each techniques. We have used Matplotlib library to visually represent the results of each algorithm. Table IV shows the accuracy in percentage for each algorithm for both of our analysis as well as the time complexity of the algorithms. Fig. 4 & 5 shows the predictions of each algorithm for both of our analysis respectively. For fig. 4, blue and red block represents the number of patients predicted to be in normal and abnormal state respectively. In fig. 5, green and red block represents the number of patients predicted to be diagnosed with Hernia and Spondylolisthesis respectively. For both of our analysis, we compared the predictions of each algorithm with the actual results from the test set. In both of our analysis, we found that all the algorithm works very accurately for our data set. In Table III, the first column is the algorithms and the second and third columns are the outcomes for first and second analysis correspondingly.

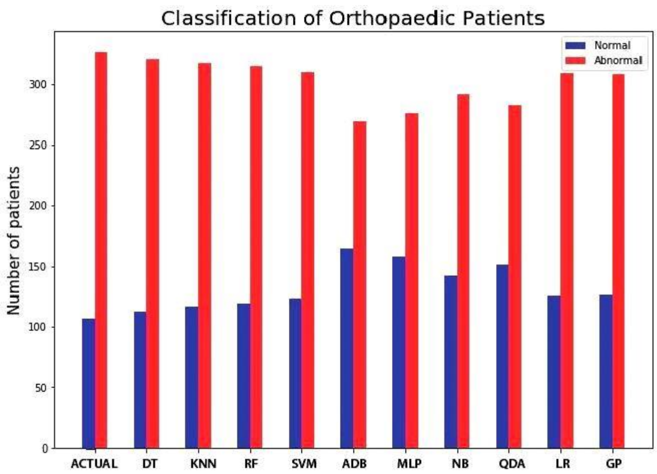

Fig. 4 Output histogram for first analysis

From fig. 4 we can see that some of the algorithms prediction was close to the actual result, whereas decision tree was only few units away from the actual result, providing the best result among all the algorithms for our first analysis.

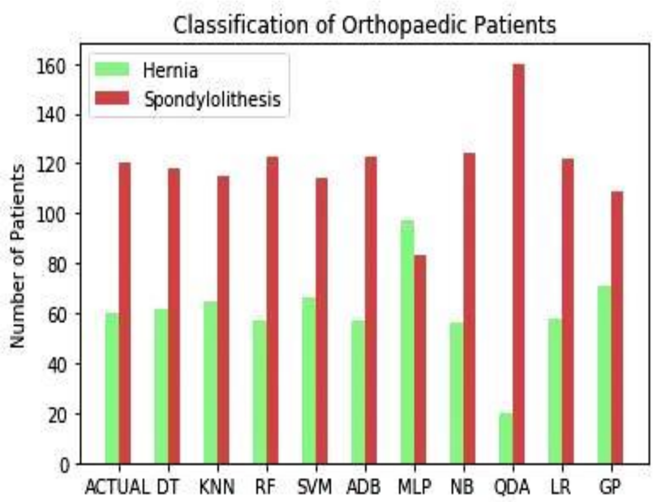

Fig. 5 Output histogram for second analysis

From fig. 5 we can see that unlike the result of first analysis, decision tree was only 2 units away from the actual result, providing the best result among all the algorithms for our second analysis as well.

From the result of both of our analysis, we can state that Decision Tree algorithm provides the best result for this kind of analysis given the similar type of dataset.

Table IV. Accuracy Distribution

Name of the Algorithm	**Normal/ Abnormal (%)**	**Hernia/ Spondylolisthesis (%)**	**Time complexity**
Adaptive Boosting	84	98	O(Tf)
Decision Tree	92	99	O(h)
Gaussian Process	89	93	O(N^3)
K-Nearest Neighbor	90	97	O(nd)
Logistic Regression	86	94	O(N)
Multi-Layer Perception	75	85	O(2^n)
Naïve Bayes	85	97	O(Np)
Quadratic Discriminant Analysis	88	77	O(log(MN)/ epsilon^3)
Support Vector Machine	89	96	O(N^2)
Random Forest	90	97	O(v * n log(n))

V. CONCLUSION

A person can be categorized based on his orthopedic condition and to do that we have used a number of machine learning techniques. Since there has not been enough application of machine learning in orthopedic field, we tried to implement a number of algorithms to analyze the comparative performance. Six biomechanical features have been used as parameters for the algorithms. Among all the ten algorithms, Decision Tree has provided the most accuracy for our dataset. It has given 92% accuracy for the first analysis where it detected whether the person is normal or not. And for the second analysis it has provided 99% to detect if the person has Disk Herniation or Spondylolisthesis. This analysis can assist doctors in identifying diseases in a faster and easier way. In future our work can be extended with the help of deep learning and neural network.

VI. REFERENCES

[1] D. Knudson, Fundamentals of Biomechanics, 2nd ed, New York: Springer, 2007, pp.3.

[2] K. Alawneh, M. Al-dwiekat, M. Alsmirat and M. Al-ayyoub, "Computer-Aided Diagnosis of Lumbar Disc Herniation," in 6th International Conference on Information and Communication System (ICICS 2013), 2013.

[3] S. Liao et al., "Automatic Lumbar Spondylolisthesis Measurement in CT Images," IEEE Trans. Med. Imag., vol. 35, no. 7, pp.1658-1669, July 2016.

[4] P. Shimpi, S. Shah, M. Shroff and A.Godbole, "A Machine Learning Approach for the Classification of Cardiac Arrhythmia," in Proceeding of the IEEE International Conference on Computing and Communication (ICCMC), pp. 603-607, 2017.

[5] Y. Wu, H. Wang and F. Wu, "Automatic Classification of pulmonary Tuberculosis and Sarcoidosis based on Random Forest," in 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017.

[6] P. Jagadev and H.G. Virani, "Detection of Leukemia and its Types using Image Processing and Machine Learning," in International Conference on Trends in Electronics and Informatics (ICEI), IEEE, 2017.

[7] D. H. Mantzaris, G. C. Anastassopoulos and D. K. Lymberopoulos, "Medical Disease Prediction Using Artificial Neural Networks,"

[8] L. Lanzarini, M. V. Camacho, A. Badran, and I. D. G. Armando, "Images compression for medical diagnosis using neural networks," *J. Computer Science and Technology*, Vol.2, No. 1, 1999, pp. 78-80.

[9] A. Taktak, A. Fisher, and B. Damato, "Modelling survival after treatment of intraocular melanoma using artificial neural networks and Bayes theorem," *Phys. Med. Biol.*, vol. 49, pp. 87-98, 2004.

[10] D. H. Mantzaris, G. C. Anastassopoulos, A. D. Tsalkidis, and A. V. Adamopoulos, "Intelligent prediction of vesicoureteral reflux disease," *WSEAS Trans. Systems*, Issue 9, vol. 4, pp. 1440-1449, 2005.

[11] D. Tasoulis, P. Spyridonos, N. Pavlidis, D. Cavouras, P. Ravazoula, G. Nikiforidis, et all, "Urinary bladder tumor grade diagnosis using on-line trained neural networks," in *Proc. Knowledge Based Intelligent Information Eng. Systems Conf.*, Heidelberg, 2003, pp. 199-206.

[12] M. Tanthanuch and S. Tanthanuch, "Prediction of upper urinary tract calculi using an artificial neural network," *J. Medical Association of Thailand*, vol. 87, No. 5, pp. 515-518, 2004.

[13] F. Dieterlea, S. Muller-Hagedorn, H. Liebich, and G. Gauglitza, "Urinary nucleosides as potential tumor markers evaluated by learning vector quantization," *Artificial Intelligence in Medicine*, vol. 28, issue 3, pp. 265–279, 2003.

[14] R. Silipo and C. Marchesi, "Artificial neural networks for automatic ECG analysis," *IEEE Trans. Signal Processing*, vol. 46, pp. 1417-1425, 1998.

[15] K. Kambouri, and S. Gardikis, "Selective clinical estimation of childhood abdominal pain based on pruned artificial neural networks," in *Proc. 3rd WSEAS Int. Conf. on Cellular and Molecular Biology, Biophysics and Bioengineering*, Athens, 2007, pp. 50-55.

[16] D. Mantzaris, G. Anastassopoulos, A. Adamopouos, I. Stephanakis, K. Kambouri, and S. Gardikis, "Abdominal pain estimation in childhood based on artificial neural network classification," in *Proc. the 10th Int. Conf. on Engineering Applications of Neural Networks*, Thessaloniki, 2007, pp. 129-134.

[17] K. Papik, B. Molnar, R. Schaefer, Z. Dombovari, Z. Tulassay, and J. Feher, "Application of neural networks in medicine - a review," *Medical Scince Monitor*, vol. 4, no. 3, pp.538-546, 1998.

[18] M. Lundina, J. Lundina, H. Burked, S. Toikkanenb, L. Pylkkanenc, and H. Joensuua, "Artificial neural networks applied to survival prediction in breast cancer," *Oncology*, vol.57 no.4 pp.281-286, 1999.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE:synthetic minority over-sampling technique", *Journal of artificial intelligence Research*, 2002, pp. 321-357.

[20] H. M. M. G. T. Herath, J. R. S. S. Kumara, M. A. R. M. Fernando, K. M. K. S. Bandara and I. Serina, "Comparison of supervised machine learning techniques for PD classification in generator insulation," *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, Peradeniya, 2017, pp. 1-6.

[21] S. Pouriyeh, S. Vahid, G. Sannino, G. D. Pietro, H. Arabnia, J.Gutierrez, "A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", *IEEE Symposium on Computers and Communications (ISCC)*, Heraklion, Greece, 2017.

[22] C.McCormic (2013,Dec.), "AdaBoost Tutorial"[online], http://mccormickml.com/2013/12/13/adaboost-tutorial