Diagnosis of Vertebral Column Disorders Using Machine Learning Classifiers

Sana Ansari, Faria Sajjad, Zia-ul-Qayyum[1]
[1]Iqra University, Islamabad Campus

Nawazish Naveed[2], Imran Shafi[3]
[2]Foundation University, Islamabad [3]Abasyn University, Islamabad

***Abstract* – Medical science is characterized by the correct diagnosis of a disease and its accurate classification to avoid complexities at treatment/medication stage. This is often accomplished by a physician based on experience without much signal processing aids. It is envisioned that a sophisticated and intelligent medical diagnostic/classification system may be helpful in making right decisions especially at remote areas where specialist physicians are not present. With this in mind, this paper proposes diagnosis and classification of vertebral column disorders using machine learning classifiers including feed forward back propagation neural network, generalized regression neural network and support vector machine and evaluates their performance. The dataset is collected using the information from the magnetic resonance images (MRI) and is classified into three different classes which are disk hernia, spondylolisthesis and normal. The classifiers are trained using 50% ratio and 10 fold cross validation approaches and are comprehensively evaluated with different architectures, activation and kernel functions. Experimental results demonstrate that feed forward back propagation neural network is 93.87% accurate on unknown test cases and performs better than the other methods.**

Keywords - Artificial neural network; Feed forward NN; Support Vector Machine; Generalized regression; Disk hernia; Spondylolisthesis

I. INTRODUCTION

Vertebral column is an integral part of human body. It is a structure which consists of usually thirty three vertebrates, out of which twenty four are articulating and nine are fused vertebras. It is found in the dorsal aspect of the torso and is separated by number of intervertebral discs. These thirty-three vertebras are divided in **five** different groups which include seven vertebrae in cervical curve, twelve vertebrae in thoracic curve, five vertebrae in lumbar curve and nine vertebrae in sacral curve [1]. Intervertebral discs are interposed between the vertebral bodies, and serve not only as shock absorber for the column but also provide the normal mobility between the adjacent vertebrae. Each disc consists of a soft central portion of spongy material [2]. The two vertebral disorders discussed in this paper are disc hernia and spondylolisthesis. Disc hernia is an intervertebral disc protrusion that is produced by the effect of flexion force acting upon the most mobile portions of the spine. A sudden strain with the spine in an unguarded position will rupture the tough annulus, allowing portions of the torn annulus and soft nucleus to escape into the spinal canal [1]. Spondylolisthesis is a condition in which a lower lumbar vertebra, usually the fifth slips forward through the plain of the intervertebral disc below it and so carries with it the whole of the upper portion of the spine [2].

Artificial Intelligence (AI) is the study of designing machines and systems that are intelligent enough to perceive the environment and take decisions like humans. Artificial Neural Network (ANN) is a subset of AI. It is composed of a dense net of computing nodes called neurons and their connections which perform the computation in a parallel and distributed manner [3]. [symon hykin old 5]. ANNs acquire knowledge through the learning process and interneuron connections stores this acquired knowledge resembling human brain [4]. [Indranarain Ramlall old 3]. The neurons learn the pattern hidden in the samples (training set) provided to them and then classifies all the unseen samples (test set) on the basis of the learned pattern. We have used two types of ANNs the first one is feed forward neural network (FFNN), which is a layered structure of computational nodes. The input signal propagates through the network in the forward direction on a layer by layer basis. It uses a backpropagation training algorithm; calculating the error and adjusting the weights to give the correct output [3] and [5]. [Symon and zurada] The second ANN we used is generalized regression neural network (GRNN). It views the design of a neural network as curve fitting problem in a high dimensional space. Learning in this method means to find a surface in a multidimensional space that provides a best fit to the training data. The generalization is to use this space to interpolate the test data [3]. The Support Vector Machine (SVM) is relatively a newer technique with some very elegant properties. It is a highly principled learning method with a single layer of nonlinear units [3]. It divides the training data into two classes (binary classifier) by constructing a hyper plane as the decision surface, such as to maximize the margin between the classes and to minimize the generalization error. During the formation of hyper plane the points that lie on the margin are considered only and are known as support vectors. These vectors result in reduced complexity and are independent of the number of features, so capable of handling large data with multiple features [3], [6] and [7]. [simon and classification paper SVM 5, 17 and 18].

In our study we have used UCI data set for the diagnosis and classification of vertebral column disorders using ANNs and SVM.

Rest of the paper is organized in six sections. Section II presents the literature review; Section III presents the

978-1-4799-0604-8/13/$31.00 ©2013 IEEE

proposed method for the diagnosis of vertebral column disorders. Section IV discusses the structure of the classifiers used for diagnosis and classification. Section V provides the discussion on the experimental results. Section VI finally concludes the paper.

II. LITRATURE REVIEW

ANNs and SVM are used for diagnosis of different bone diseases. One of which is hip osteoarthritis, where probabilistic neural network (PNN) based classification scheme is used [8]. This PNN is a combination of ANN and Bayes classification approach. In [9] lumbar stenosis is diagnosed with the help of Multilayer Perceptron (MLP). Initial work on the diagnosis of lumber disc hernia is done using a computer aided system based on formal analysis techniques [10]. In [11] a two-level probabilistic model is used for localizing and labeling inter-vertebral discs for the diagnosis of inter-vertebral disc degeneration (IDD). Abnormality detection of inter-vertebral discs with a probabilistic model is proposed in [12]. For diagnosis of lumber disc hernia, Bayesian classifier with Gibbs distribution is used, which on average provided 92.5% diagnosis accuracy [13]. Some other studies [14] [15] use ANN for the diagnosis of lower back pain and sciatica. SVM is used for classification of inter-vertebral column diseases into two classes [16]. It provided 84.67% of training accuracy and test accuracy of 90.23%. SVM with reject option is proposed by Ajalmar et al [17] attaining training accuracy of almost 86% and is called rejoSVM. Benefits of rejoSVM were simplicity and interpretability. It is compared with generalized regression and feed forward networks. Some of the ressearchers advised the use of ensembles of classifiers for the diagnosis of pathologies of the vertebral column [18].

III. PROPOSED METHOD FOR DIAGNOSING VETEBRAL COLUMN DISORDERS

A. Methodology

We have used two classifiers to diagnose vertebral column disorders, which includes two types of ANNs and SVM. The disorders are classified as disc hernia, spondylolisthesis and normal patients. The methodology used by both the classifiers is divided in to three phases. The detail of these phases is discussed below. Figure 1 presents the block diagram of these phases.

Phase I: Pre-processing

In this phase the data is first collected. During the collection the attributes are selected after discussing them with the doctor. We have incorporated all the attributes as they are significant and have a huge influence on the class labels. Next the data is analyzed and cleaned. In the cleaning process the missing and noisy values are handled. Afterwards the data is normalized. Subsequently, each sample in the data set is allotted its targets class. The entire data set is divided into two sets, the training and the test sets. In case of the validation set, the training set is further divided in to two sub groups, the training group and the validation group. After this the data set is ready to be provided to the classifiers.

Phase II: Design Classifier

The two types of classifiers we have used include two types of ANNs and SVM. After the pre-processing is complete the data is presented to these classifiers. Both the classifiers belong to the supervised category. For the supervised classifier the training set includes both the samples and their targets. These trained classifiers learn the pattern from the samples provided to them and diagnose the correct disorder. Each classifier is trained two times with two different datasets; i.e. the 50% ratio and the 10 fold cross validation. The performance of all six classifiers (three classifiers with two different data distributions) is evaluated. If the classifier is not trained the training process is repeated with different structure and functions.

Phase III: Post-processing

In this phase the results of the trained classifiers are converted into a form that is easily understandable. The results show whether a person is normal or has a vertebral column disorder and the type of disorder either disc hernia or spondylolisthesis.

B. Material

The data set used for classifying vertebral column disorders is taken from the UCI machine learning database. The dataset contains 310 samples out of which 60 samples are of disc hernia, 150 samples of spondylolisthesis and 100 normal samples. Each sample has six attributes. All the attributes are numerical. The second and sixth attribute contains some samples having negative values as well. All these attributes are used to classify between disc hernia, spondylolisthesis and normal patients. The attribute along with their minimum and maximum values are presented in table I and **figure 2**.

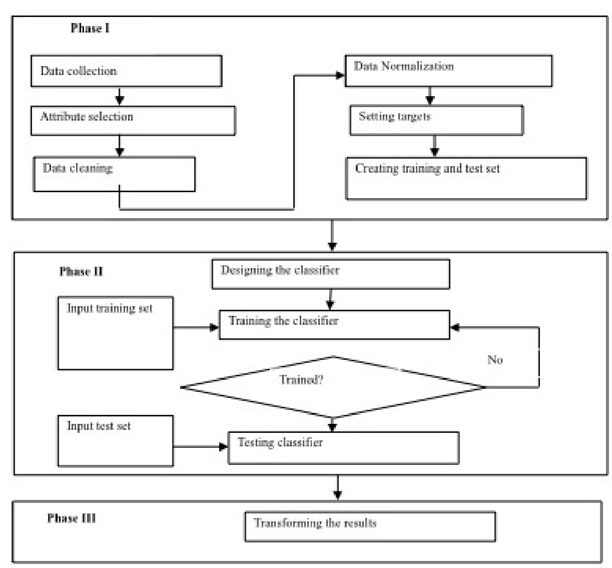

Figure 1: Proposed Methodology

TABLE I: ATTRIBUTES AND THEIR VALUES

Attributes	**Values**	
	Minimum	*Maximum*
Pelvic Incidence	26.1479	129.834
Pelvic Tilt	-3.7599	49.4319
Lumbar Lordosis Angle	14.00	125.7424
Sacral Slope	13.3669	121.4296
Pelvic Radius	70.0826	163.071
Degree Spondylolisthesis	-11.0582	418.5431

The dataset is divided into two sets the training and the test set. We applied two different approaches for data distribution. In the first method the entire data is divided into two sets with 50% ratio of samples. While in the second method 10-fold cross validation is done. In the first method the training set contains 159 samples which include data from all the three classes. The test set contains the rest of the 151 samples. These samples are not part of the training set and used to evaluate the performance of the classifiers. In the 10 fold cross validation the entire data set is divided into 10 subsets. Each subset is used as a test set in one of the 10 learning cycles. The subset used to test the

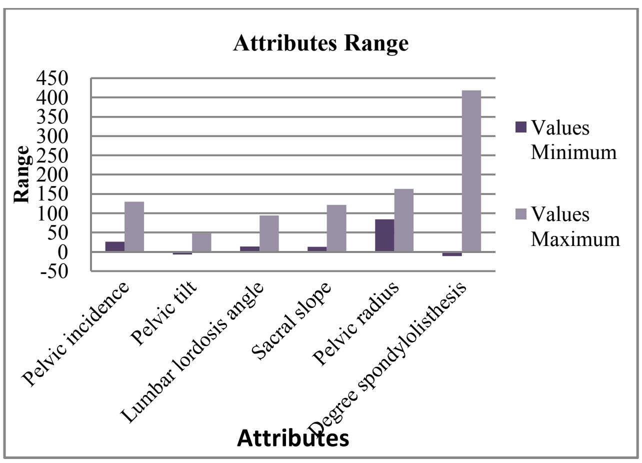

Figure 2: Attributes range

classifiers is different from the sets used to train them. The division of the samples in training and the test sets is presented in table II.

TABLE II: THE DATASET IN THE TRAINING AND TEST SET

Data Distribution	**Dataset**	**Total**
50% Ratio	Training Set	159
	Test Set	151
	Total	*310*
10-fold cross validation	Training Set	279
	Test Set	31
	Total	*310*

IV. CLASSIFIERS STRUCTURE

We have used two types of classifiers for the diagnosis and classification of vertebral column disorders which includes supervised ANNs and SVM. Within the ANN category we have used two different types of networks which are feed forward backpropagation network and the generalized regression network. The details of these classifiers are given below.

A. Feedforward Backpropagation Neural Network

The FFNN we used has an architecture 6-10-2. The input layer has 6 nodes because each sample has six features and so there are six values in each input record. The hidden and the output layers have 10 and 2 neurons respectively. These neurons perform all the computation. The activation or transfer function we used for both the hidden and output layer neurons is hyperbolic tangent sigmoid function. It is relatively faster than the other activation functions.

$$a = tansig(n) = \frac{2}{1 + e^{(-2*n)}} - 1$$

Where n is a value of the hidden and output layer neurons given to the hyperbolic tangent sigmoid activation function.

The error is calculated using the mean squared error performance function:

$$Eav = 1/N \sum_{n=1}^{N} E(n)$$

Where Eav is the average error of the network, N is the total number of training samples and E(n) is the network error.

The learning algorithm used is the gradient descent with momentum weight and bias learning function. It changes the weights in such a way that it reduces the error between the output and the targets. The learning function is:

$$dW = mc * dWprev + (1 - mc) * lr * gW$$

Where dW is the weight change for a particular neuron from its input and error, dWprev is previous weight change, lr is the learning rate in our case it is 0.01, mc is momentum constant which is 0.9 in our situation and W is either the weight or bias. The weights are changed based on the gradient decent method.

The training algorithm we used is a variant of backpropagation algorithm which is levenberg-marquardt backpropagation. It is based on the error correction rule. The backpropagation algorithm uses the jacobian matrix which contains first derivative of network error with respect to the weights and bias. The Levenberg–Marquardt algorithm is a combination of steepest descent and the gauss–newton algorithm. Its updation rule is:

$$w_{k+1} = w_k - [J^T J + \mu I]^{-1} J^T e$$

Where w is the weight, J is the jacobian matrix, T means transpose, μ is the combination coefficient which is always positive, I is the identity matrix, and e is the error. The J^Te is the gradient. The combination coefficient allows the Levenberg–Marquardt algorithm to switch between the steepest decent and netwon algorithms during the training process.

The training set provided to FFNN is divided in two parts the training set and the validation set. The validation set monitors the training process and stops training when the error starts to increase instead of decrease. Both the data distribution (50% ratio and 10-fold cross validation) used the same FFNN architecture, training, learning, activation and performance functions.

The weights are converged after 120 epochs in case of 50% ratio data distribution approach whereas in 10-fold approach weights are converged after 70 epochs. Figure 3 and 4 shows the training graphs of FFNN with 50% ratio and 10 fold cross validation approaches respectively. The results of FFNN in the two methods are discussed in the next section.

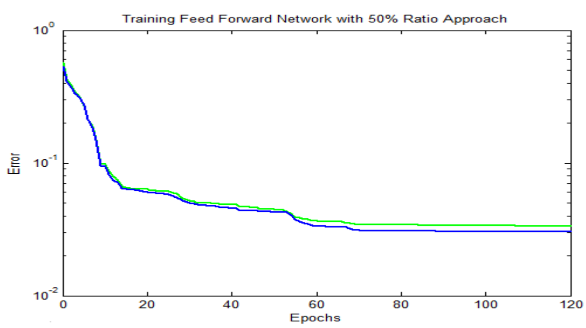

Figure 3: Training of FFNN with 50% ratio method.

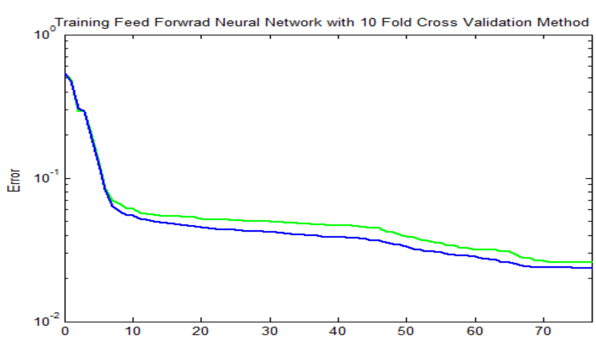

Figure 4: Training of FFNN with 10 fold cross validation method.

B. *Generalized Regression Neural Network*

GRNN also have a layered structure with input, hidden and output layers. It uses the normal distribution probability density function.

$$Y_i' = \frac{\sum_{i=1}^{n} yi * \exp\left(^{-D_i^2}/_{2\sigma^2}\right)}{\sum_{i=1}^{n} exp\left(^{-D_i^2}/_{2\sigma^2}\right)}$$

$$D_i^2 = (X - X_i)^T - (X - X_i)$$

Where yi is the weight connection between the neurons, n is the number of training patterns, D is the distance between the training sample and the point of prediction, σ is the spread or the smoothness parameter and X is the training samples.

The structure we used in 50% ratio approach has 6 nodes in the input layer (as each sample has six different attributes so for each attribute we have a separate node). The hidden layer has 159 dimensions. It uses by-weight-and-bias layer initialization function which calculates new weights and bias values according to their initialization function. The net input is calculated using product net input function. It calculates net input of a layer by combining its weighted inputs and biases. The weighted input is calculated using the euclidean distance weight function. For the hidden layer neurons hexagonal layer topology function is used which arrange the neurons in 159 dimensional hexagonal patterns. The transfer function used at hidden layer is the radial basis function.

$$a = radbas(n) = \exp(-n^2)$$

Where n is the value of the hidden layer neurons.

The output layer on the other hand has the dimension 2. It uses the same weight bias initialization and topology function that is by-weight-and-bias layer initialization function and hexagonal layer topology function respectively. However the net input is calculated using sum net input function. The weighted input is calculated using normalized dot product weight function. The activation function used at the output layer is linear transfer. The spread that we used is 4.5.

In the second approach (10-fold cross validation) the input layer has **again** six nodes. The hidden layer has the dimension 279. The rest of the function which includes initialization function, net input function, distance function, topology and transfer function were same as used in the first approach. The output layer in the second method is same as the first methods output layer. However, the spread used in the second method is increased to 5.5.

With the above mentioned functions and spread the GRNN gave good results which are presented in the next section.

C. *Support Vector Machine*

In case of SVM we divided three class data into binary form- the abnormal and normal class and disorder classes. The SVM we implemented is also trained and tested using the two data distribution approaches (50% ratio and 10-fold cross validation).

We have used two types of functions for implementing SVM. The first one generates and trains the SVM structure with train support vector machine classifier function (SVMtrain). It uses the training data along with the group and the kernel function for constructing the hyper plane and segregates the data into the particular classes. The train function used:

$$c = \sum_i a_i \, k(s_i, x) + b$$

Where a_i are the weights, k is the kernel function, s_i are the support vectors, x are the vectors to be classified or the training samples and b is the bias.

The kernel function we used is polynomial for both the data distributions and the normal and disorder classes.

$$k(x_i, x_j) = (x_i, x_j + z)^d$$

Where z is the constant and d is the degree of the polynomial.

The second function is the classify support vector machine. It takes the SVM structure created by SVMtrain and the test data to classify the unseen samples.

In the first approach the SVM designed has 72 support vectors. The bias used is negative 0.2193. The support vector indices are 72.

The SVM designed for the disorders class has 17 support vectors. The bias in this case is negative 1.3571. The support vector indices are 17.

For the second data distribution approach the SVM we designed for the abnormal-normal class has 887 support vectors. The bias in this case is negative 1.3571. The support vector indexes are 887 as well.

The SVM we designed for the disorder class has 21 support vectors. The bias in this case is negative 1.0898. The support vector indexes are 21 as well. Its results are discussed in section IV.

V. RESULTS AND DISCUSSION

The three classifiers we have used belong to two categories one is the neural networks and the other is the SVM. Within the ANN category we have used two supervised ANNs the feed forward ANN with the backpropagation algorithm and the other is the generalized regression. The results of the classifiers with two datasets distributions are discussed. In the first case the data is equally divided between the training and the test set. The results attained through this distribution are suitable and are shown in Table III. The second case is the 10 fold cross validation. The results achieved through this data distribution are better than the first case. This method is more detailed and produced better classifiers **as compared to the first approach**. Its results are presented in Table IV and figure 4.

Table III: RESULTS OF THE CLASSIFIERS ON 50% DIVISION OF DATA

Dataset	**Number of samples**	**Classifier Results**		
		Artificial Neural Network		***Support Vector Machine***
		Feedforward Network	*Generalized Regression*	
Training set	159	86.79%	100%	100%
Test set	151	86.76%	92.05%	86.87%

The above table concludes that the generalized regression network performed better than the feed forward ANN and SVM with the accuracy of 92.05% when the data is 50-50 divided between the train and the test set. Whereas the accuracy of the feed forward ANN and the SVM is almost similar for the test set. The figure 5 depicts the results of the above table.

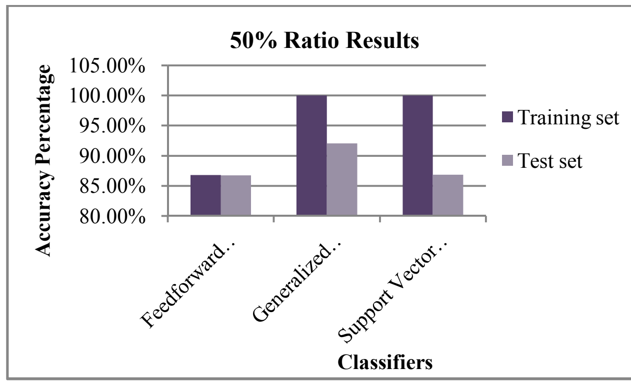

Figure 5: 50% ratio results

Table IV: RESULTS OF 10-FOLD CROSS VALIDATION

Number of folds	**Test samples**	**Classifier Results**		
		Artificial Neural Network		***Support vector Machine***
		Feedforward Network	*Generalized regression*	
1	1-31	90.32%	80.65%	86.33%
2	32-62	90.32%	80.65%	81.49%
3	63-93	87.10%	77.42%	86.33%
4	94-124	90.32%	74.19%	82.34%
5	125-155	96.77%	93.55%	91.94%
6	156-186	93.55%	87.10%	87.10%
7	187-217	100%	83.87%	89.43%
8	218-248	100%	96.77%	88.79%
9	249-279	96.77%	93.55%	92.78%
10	280-310	93.55%	80.65%	86.33%
Total	**Average**	**93.87%**	**84.84%**	**87.29%**

From the above table we can comprehend that in case of 10 fold cross validation the feed forward ANN outperformed the GRNN and the SVM with the accuracy of 93.87%. The SVM in both the cases remained the second best classifier. Figure 6 shows the results for the 10 fold cross validation method.

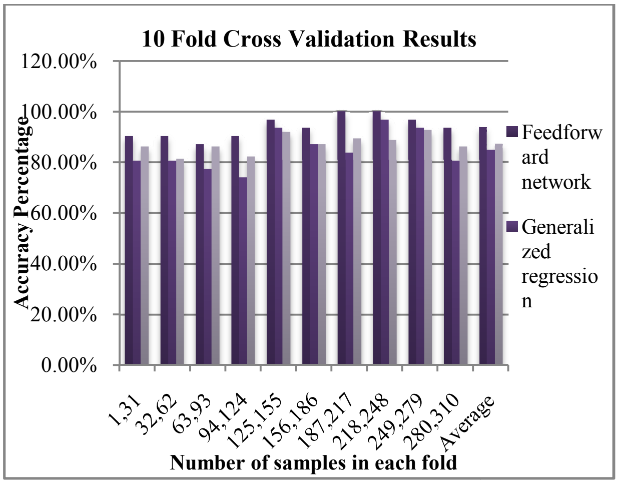

Figure 6: 10 fold cross validation results

From Table III and IV we can conclude that as the number of training examples increase, the FFNN and SVM produced better results. The GRNN works on the one pass learning rule and so produced good results even when the number of examples are less. We have deduced that the 10 fold cross validation method is better than the 50% ratio approach as it involves thorough training and provided better and more accurate results for the two classifiers FFNN and SVM.

The GRNN uses the one pass learning concept and thereby its training time is less than the FFNN and SVM. The SVM has an advantage over the FFNN as its takes less training time compared to FFNN. But in our case the results of the FFNN are better making us prefer the FFNN over SVM even when FFNN is more time consuming.

VI. CONCLUSION

During the last decade artificial intelligence has been used extensively for medical applications. Intelligent systems are applied for diagnosis, classification, monitoring, drug development etc. This paper presents the machine learning classifiers for the diagnosis and classification of vertebral column disorders. The classifiers used are neural networks which includes feed forward network and generalized regression network and SVM. They classify the disorders in three classes namely disk hernia, spondylolisthesis and normal. The classifiers are trained using two approaches 50% ratio and 10-fold cross validation. It is concluded that in 50% ratio approach the GRNN outperformed both the FFNN and SVM whereas in 10-fold cross validation the FFNN performed better than SVM and GRNN. From these results we have found that GRNN overfitted in case of 10-fold cross validation and so did not perform well as compared to the 50% ratio approach. The FFNN gives best results but takes more training time as compared to SVM.

REFERENCES

[1] Hamilton Bailey and Robert John McNeill Love, Bailey & Love's Short Practice of Surgery , 25th ed., 2008.

[2] Dana Andersen, Timothy Billiar, David Dunn, John Hunter, Jeffrey Matthews and Raphael E. Pollock F. Brunicardi, Schwartz's Principles of Surgery, 9th edition, Ed.: McGraw Hill, 2009.

[3] Simon Haykin, Neural Networks and Learning Machines. New Dehli: Pearson Education, 2010.

[4] Indranarain Ramlall, "Artificial Intelligence: Neural Networks Simplified," no. 39, 2010.

[5] J. M. Zurada, Introduction to Artificial Neural Systems. Mumbai, India: Jaico Publishing House, 1994.

[6] S. B. Kotsiantis · I. D. Zaharakis · P. E. Pintelas, "Machine learning: a review of classification and combining techniques," Artificial Intelligence Review, no. 26, pp. 159-190, 2006.

[7] John Shawe-Taylor & Nello Cristianini, Support Vector Machines and other kernel-based learning methods.: Cambridge University Press, 2000.

[8] L. Costaridou, D. Cavouras, I. Kalatzis, E. Panagiotopoulos and G.Panayiotakis I. Boniatis, "Assessing hip osteoarthritis severity utilizing a probabilistic neural network based classification scheme," Medical and Enginnering and Physics, pp. 1-11, 2006.

[9] Kathleen Hua, Kien A. Hua and Jintavaree Srisomboon Soontharee Koompairojn, "Computer-Aided Diagnosis of Lumbar Stenosis Conditions," in Society of Photo-Optical Instrumentation Engineers, San Diego, USA, 2010, pp. 76241C-76241C-12.

[10] W.R Hudgins, "Computer-aided diagnosis of lumbar disk herniation," Spine , vol. 8, pp. 604-615, 1983.

[11] Raja' S. Alomari, and Vipin Chaudhary Jason J. Corso, "Lumbar Disc Localization and Labeling with a Probabilistic Model on both Pixel and Object Features," in Medical Image Computing and Computer Assisted Intervention, 2008, pp. 202-210.

[12] J. J. Corso bf, B. Chaudhary and G. Bhillon R. S. Alomari, "Abonormality Detection in Lumbar Discs from Clinical MR Images with Probablistic Model," in Computer Assisted Radiology and Surgery, 2009.

[13] Vipin Chaudhary and Gurmeet Dhillon Raja' S. Alomari · Jason J. Corso, "Toward a clinical lumbar CAD: herniation diagnosis," International Journal of Computer Assisted Radiology and Surgery, vol. 6, pp. 119–126, 2011.

[14] S. Kul and M. Ozates B. Karlik, "Diagnosis of Lumber Disc Hernia from Images using Artificial Neural Network," in International Conference on Electrical Systems Design and Technology, 2008.

[15] H. E. Kocer and H. E. Y. Unal, "A comparison of feature extraction techniques for diagnosis of lumbar intervertebral degenerative disc disease," in IEEE International Sympossium on Innovations in Intelligent Systems and Applications, Konya, Turkey, 2011, pp. 490-494.

[16] Margarita Kotti and Konstantinos I. Diamantaras, "Towards minimizing the energy of slack variables for binary classification," in European Signal Processing Conference, Bucharest, Romania, 2012.

[17] Ricardo Sousa, Guilherme de A. Barreto, and Jaime S. Cardoso Ajalmar R. da Rocha Neto, "Diagnostic of Pathology on the Vertebral Column with Embedded Reject Option," in Proceeeding of 5th Iberian Conference on Pattern Recognotion and Image Analysis, Berlin, Germany, 2011, pp. 588-595.

[18] D.Joannes, R. Pierre and L.hurert B.Eric, "Analysis of the Sagittal Balance of the Spine and Pelvis Using Shape and Orientation Parameters," Journal of Spinal Disorders and Techniques, vol. 18, no. 1, pp. 40-47, Februray 2005.