

Diagnosis of Vertebral Column Pathologies using kNN Classifier

Vijayalakshmi G V

Associate Professor, Dept of ECE,
Dr T Thimmaiah Institute of Technology, VTU

Mohan Kumar M

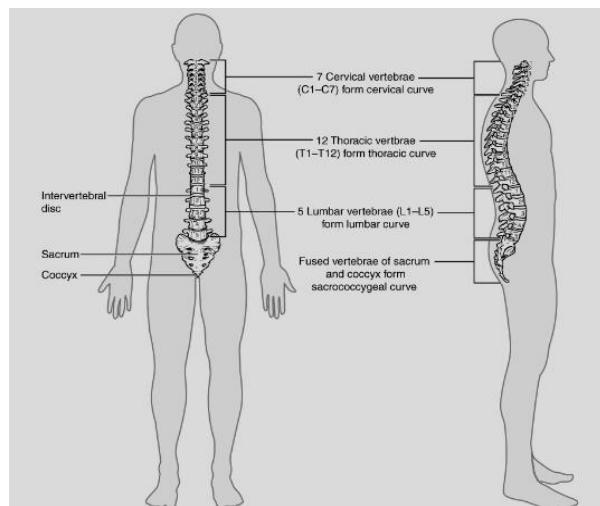
Assistant Professor, Dept of ECE
Yenepoya Institute of Technology, VTU

Abstract— The paper proposes a pattern recognition system to identify the pathologies in the vertebral column that can assist the medical experts in improving the health care systems by reducing the errors. In this work a system is developed using the kNN machine learning algorithm to identify the disc hernia and Spondylolisthesis. Also, the methodology utilized the dataset obtained from the UCI machine learning repository with the biomechanical features to evaluate the proposed system. The experimental results showed that the system was accurate in achieving a success rate of 88.31%.

Keywords— Pattern recognition, kNN, accuracy, pathologies, healthcare, machine learning

I. INTRODUCTION

Healthcare is an industry that offers care to millions of people while maintaining quality, value and outcome. Technology enabled healthcare allows the specialists to develop models to provide smart healthcare. Machine learning(ML) with its applications is receiving a gradual acceptance in the healthcare industry. It has the ability to analyze the data and have the potential to impact in a similar way as clinicians do. One of the widespread application of Machine learning in healthcare is diagnostics as it provides various methods and techniques that can help in diagnostic problems in medical domain. Machine Learning methods when implemented successfully can help in integrating the computer systems in the health care environment to enhance the work of medical experts and eventually improve the efficiency and quality of medical care. Thus this work focuses on developing a pattern recognition system, a subset of ML to diagnose the pathologies of the vertebral column.



Vertebral column(VC) is a vital part of the human body. It supports the head, neck and body allowing for their movements. VC comprises of 33 vertebrae of which 24 are articulating and the remaining 9 are fused. VC is found in the dorsal portion of the torso and is separated by numerous intervertebral discs. Of the 33 vertebrae, 7 vertebrae form the cervical curve, 12 form the thoracic curve, 5 form the lumbar curve and 9 the sacral curve. Each intervertebral disc present between the vertebrae consists of a soft central portion and the intervertebral discs provide the regular mobility between the adjacent vertebrae. The structure of the vertebral column of the human body is presented in Figure 1[1,2].

Ageing and trauma due to mechanical overload induces pathological changes in the VC. These changes generally tend to reduce the mechanical strength of the vertebral column and restricts the range of the movements. The pathologies considered in this work are disc hernia and spondylolisthesis. Disc hernia is a condition of intervertebral disc protrusion, where the soft centre pushes through a crack into the tougher exterior covering. This may bother the nerves in the close proximity that results in pain or numbness in leg or in an arm. Spondylolisthesis is the slipping of fifth lumbar vertebra forward through the plain of the intervertebral disc below it causing severe back pain. Both the VC disorders requires timely treatment that includes medication, physiotherapy or a surgery. Accordingly the work considers diagnosis of disc hernia and spondylolisthesis using machine learning algorithm and to forward the result to medical expert for further course of action. Rest of the paper is organized as follows. Section II reviews the literature. Section III provides the details of the proposed system. The experimental results are discussed in Section IV and finally Section V concludes the paper.

II. LITERATURE REVIEW

A review of the literature discloses the various methods that have been used in the diagnosis of the vertebral column disorders using image processing techniques and machine learning algorithms.

Neto et al., proposed a method for identifying the disorders on the vertebral column based on embedded reject option[3]. The method analyzed the proposed method with other state of the art techniques such as artificial neural network, general regression neural network and support vector machine with linear and kmod kernels. As a result it was found that inclusion of the reject option performed better as compared to other methods.

Sana ansari et al.,[4] proposed diagnosis and vertebral column pathologies using machine learning classifiers that included feed forward neural network, generalized regression neural network and support vector machine. The work considered the dataset from the UCI machine learning repository. The classifiers were evaluated using 10 fold cross validation and holdout methods using different kernel and activation functions with various architectures that provided an average classification accuracy of 88.56%.

Yavuz unal and Erdinc Kocer, further investigated the identification of vertebral column disorders using the dataset from the UCI machine learning repository using back propagation neural network and naive bayesian classifiers[5]. The developed method was found to provide the classification accuracy of 84.35% on an average with both the classifiers.

Khaled Alawneh et al., developed a computer aided diagnostic system using imaging processing methods that diagnosed the lumbar disc herniation using MRI imaging [6]. The developed system worked on top down MRI spine view of 32 samples. The process included extraction of region of interest(ROI), enhancing ROI, feature extraction and classification. The experimental results showed up higher accuracy with zero misclassifications.

The work intends to present a computer aided diagnostic system to identify the pathologies of the VC that aids in medical decision.

III. PROPOSED SYSTEM

The methodology of the proposed system includes data collection, feature extraction, machine learning algorithm(classifier) and pathology classification as presented in Figure 2.

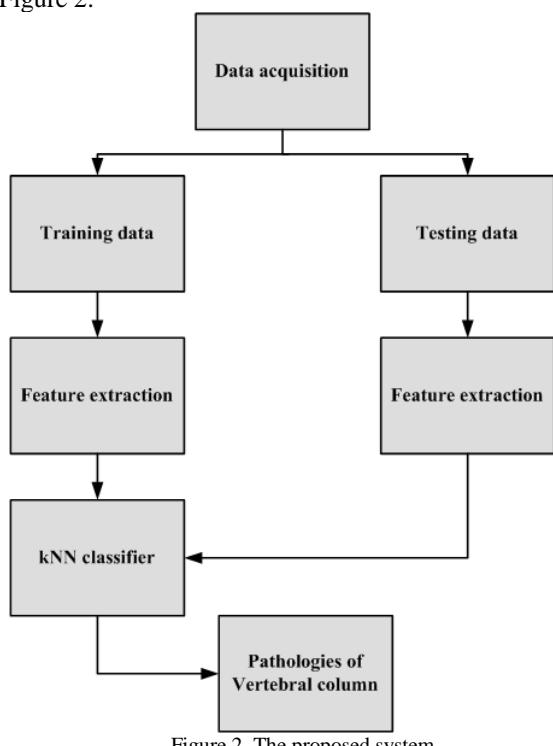


Figure 2. The proposed system

A. Data acquisition

The dataset is obtained from the UCI machine learning repository that contains data from 310 individuals. Of 310

persons, 100 are normal(NL) without any pathology in their spines, 60 are affected with disc hernia(DH) and the remaining 150 have spondylolisthesis(SL). To identify the abnormality using the proposed system, the data is divided into training set and the testing set with no overlap. The division is as shown Table I.

TABLE I. DATASET DIVISION

Dataset	Training(75%)	Test(25%)
No. of samples	233	77

B. Feature extraction

For each samples present in both training and testing set, 6 biomechanical attributes or features are extracted[3,7]. The features are as presented in Table II and the 2 dimensional feature space is displayed in Figure 3. These derived attributes decide the performance of the system. A good feature set will have strong class discrimination ability. The six features are integrated to form the feature vector FB.

$$FB = [P_i \ P_t \ L_a \ S_s \ P_r \ G_s] \quad (1)$$

The feature vectors in the training set are labeled with the three class labels $y = [0, 1, 2]$, where '0' indicate Disc hernia, '1' indicates Spondylolisthesis and '2' indicates Normal to form the training data set $[FB, y]$. The training dataset is now provided to kNN classifier for training to obtain the optimum classifier model. The classifier model is evaluated using the testing dataset.

TABLE II. BIO MECHANICAL FEATURES/ATTRIBUTES

No.	Attribute
1	Pelvic incidence(P_i)
2	Pelvic tilt(P_t)
3	Lordosis angle(L_a)
4	Sacral slope(S_s)
5	Pelvic radius(P_r)
6	Grade of slipping(G_s)

C. kNN Classifier

kNN classifier algorithm is one of the simplest and most frequently used learning algorithm for classification applications. It uses the labeled dataset, where the samples are separated into several categories and the output of the kNN classifier is the class membership. When a new sample is presented, the classifier predicts its class membership or label by considering the majority vote of the similarities between the new sample and its k nearest neighbors. Thus the algorithm checks how close is the new sample with the other samples of the training dataset using the similarity measures. The similarity measure used in general is the Euclidean distance, which is given as,

$$D_E = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (2)$$

The optimum value for k can be appropriately selected for best classification accuracies.

The training data set $[FB, y]$ along with the testing set is provided to the classifier. The training is fast here and the

classifier predicts the class label for the samples in the testing set using the majority voting and similarity score.

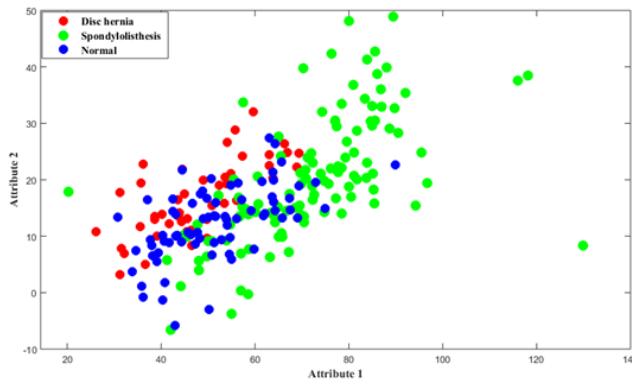


Figure 2. 2 Dimensional feature space

IV. RESULTS

The experiment to diagnosis the vertebral column pathologies is conducted on the 310 samples that constitute both normal and abnormal conditions using kNN classifier. The classifier was trained using the training dataset by varying the values of k from the initial value of 1($k=1$) to obtain the best model. With each variation the training error was recorded, eventually the least error was obtained for $k=3$. Later the performance metrics: accuracy of classification and misclassification error was calculated for the training phase using the confusion matrix[8]. The confusion matrix of the training process is presented in Table III.

TABLE III. CONFUSION MATRIX OF THE TRAINING PHASE

Actual class	Predicted class		
	DH	SL	NL
DH	35	0	10
SL	2	107	4
NL	7	2	66

The elements of the principal diagonal in Table III indicate the true classifications and the elements off the diagonal show the misclassifications or errors. From the entries of the above table, the training accuracy(Accuracy) and the misclassification error is calculated as shown,

$$\text{Accuracy} = \frac{\sum \text{Elements of the principal diagonal}}{\sum \text{Elements of the matrix}} = 89.27\%$$

and the

$$\text{Misclassification error} = 1 - \text{Accuracy} = 10.72\%$$

Further, the model is evaluated using the testing dataset. The confusion matrix of the testing process is presented in Table IV.

TABLE IV. CONFUSION MATRIX OF THE TESTING PHASE

Actual class	Predicted class		
	DH	SL	NL
DH	13	0	2
SL	2	35	0
NL	3	2	20

From the Table 4, the Accuracy and the error is found to be 88.31% and 11.68% respectively. The classification accuracies of the individual classes is as displayed in Figure 4.

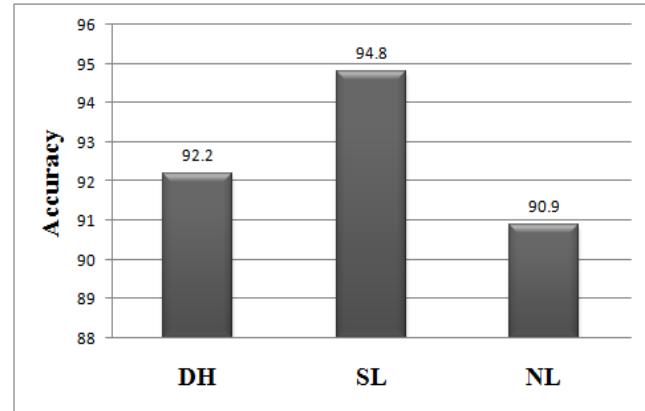


Figure 4. Classification accuracies of individual classes

From the result, it is found that the proposed system was able to identify the pathology Spondylolisthesis with a higher accuracy of 94.8% and the overall accuracy of the system is 88.31% with least errors.

V. CONCLUSION

The proposed work focused on developing a pattern recognition system to diagnose the pathologies of the vertebral column that when implemented successfully can help in integrating the computer systems in the health care environment to add to the work of medical experts in improving the efficiency and quality of medical care. Accordingly the results proved that the system was efficient in providing a better accuracy of 88.31% with less false alarms.

REFERENCES

- [1] Smink, D. S. (2015). Schwartz's principles of surgery. Annals of Surgery, 261(5), 1026.
- [2] Marieb, E. N., & Hoehn, K. (2007). Human anatomy & physiology. Pearson Education.
- [3] da Rocha Neto, A. R., Sousa, R., Barreto, G. D. A., & Cardoso, J. S. (2011, June). Diagnostic of pathology on the vertebral column with embedded reject option. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 588-595). Springer, Berlin, Heidelberg.
- [4] Ansari, S., Sajjad, F., Naveed, N., & Shafi, I. (2013, June). Diagnosis of vertebral column disorders using machine learning classifiers. In 2013 International Conference on Information Science and Applications (ICISA) (pp. 1-6). IEEE.
- [5] Unal, Y., & Kocer, H. E. (2013, May). Diagnosis of pathology on the vertebral column with backpropagation and Naïve Bayes classifier. In 2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECEE) (pp. 276-279). IEEE.
- [6] Alawneh, K., Al-dwiekat, M., Alsmirat, M., & Al-Ayyoub, M. (2015, April). Computer-aided diagnosis of lumbar disc herniation. In 2015 6th International Conference on Information and Communication Systems (ICICS) (pp. 286-291). IEEE.
- [7] Berthonnaud, E., Dimmet, J., Roussouly, P. & Labelle, H. (2005). 'Analysis of the sagittal balance of the spine and pelvis using shape and orientation parameters', Journal of Spinal Disorders & Techniques, 18(1):40-47.
- [8] Fawcett, T. (2003) ROC Graphs: Notes and Practical Considerations for Researchers