International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7, Issue-6, March 2019

Classification and Correlational Analysis on Lower Spine Parameters using Data Mining Techniques

Richie V Johny, R.Roseline Mary

Abstract*: The application of data mining in the field of medical science is slowly gaining popularity. This is due to the fact that enormous statistical inferences from data related to the human body and medicine was a possible with high accuracy rates which was a tedious task in the past. This had led to discoveries and breakthroughs which has saved thousands of lives. Lower back pain is one of the most common issues faced by majority of the population throughout the world. The early detection and treatment of LBP can avoid life threatening issues in the body. Objective: This study aims to create a classification model which can be used to detect an unhealthy spine using the lumbar and sacral parameters. Correlational analysis was performed between different attributes to find distinguishing factors between healthy and unhealthy spine. Method: Classification methods were used such as decision tree and SVM. Correlational analysis was performed using pearson method between each attribute. Results: After creating the model using the different classification methods it was found that Ctree produced the highest accuracy with 92.80% on average. It was also found that there were 6 attribute pairs that had high correlation coefficient to distinguish unhealthy and healthy spine observations.

Index Terms*: Data mining, Lower back pain, Classification, Correlation, Decision tree, Support vector machines, Pelvic incidence, Spondylolisthesis, Sacral slope.

I. INTRODUCTION

The human spine is one of the most important structure in the body. [1] The feature of human spine which is not seen in any other animal is that the spine is erect. This is a result of evolution in nature. With an erect spine the animal can now have more broader view of the environment. This enables it to move in the environment on two limbs and can manipulate anything it using the hands. This is a unique feature only seen in humans, apes and monkeys. [2] The spine consists of 33 vertebrae columns out of which 24 are presacral vertebrae followed by sacrum and coccyx. The presacral vertebrae is further subdivided into 7 cervical, 12 thoracic and 5 lumbar vertebrae columns.

The sacrum consists of 5 fused sacral vertebrae and the coccyx consists of 4 coccygeal vertebrae. [3] All the vertebrae columns are attached to each other with a type of a soft and spongy type of bone. It is called cancellous bone.[4] It is shaped like a flat disc and acts like a cushion between each vertebra. The entire spinal column takes the load of the upper body thus the cushion like bone in between each vertebra is very crucial in reducing the load on consecutive vertebrae columns and also to reduce friction during movement of the spine. There are 4 major curves or segments of the spine where there are points of significant change in degree of the curve. Starting from the bottom they are the sacral curve which is present right above the tailbone. Next is the lumbar curve and this is the curve we feel at the lower back. This part is present right behind the belly region. Next segment is the thoracic curve which is the segment of the spine at the heart and lung region. The final segment is the cervical curve and this part connects the head and the thoracic region of our body.

Apart from giving the skeletal structure to the body, [5]it conceals the spinal cord within the entire vertebral column. Each vertebra has a hollow center where the entire spinal cord runs through from the brain to the tailbone region. The spinal cord is very important part of the body. It acts as the information highway through which all other neural paths are established to every part of the body from the thoracic region till the foot.

Since the spine plays an important role in how the central nervous system functions, it is vital to have a healthy spine. This can be done by always maintaining the right posture and not stressing the back.[6] The right posture must be maintained for any kind of physical activity we perform, at the same time we should ensure that we do not put too much stress on the back. When these two factors are not taken care of, the person starts to experience pain in different segments of the spine. The pain is caused by various factors and these factors are mainly the result of wrong posture or applying to much stress on the back. As a result, the nerve roots that originate from the back are irritated or the muscles that support the back are strained or the bones, ligaments, joints or the intervertebral disc may be damaged.[7] According to a review done on 165 studies from 54 countries it has been found that lower back pain is one of the most common issues among the people and it is also the most common back pain. Apart from the pain, the main symptom of spinal issue is that the curvature of the spine is not healthy compared to healthy individual's spine.

Revised Manuscript Received on 30 March 2019.
* Correspondence Author
Richie V Johny*, Department of Computer Science CHRIST (Deemed To Be University) Bengaluru, Karnataka, India
Prof Roseline Mary. R Assistant Professor Department of Computer Science, CHRIST (Deemed To Be University) Bengaluru, Karnataka, India

The aim of this study is to create a classification model that classifies a person's spine as healthy or unhealthy based on the 12 parameters present in the lower back pain symptoms dataset. The models are created using decision tree and SVM algorithms. The performance analysis is done to show which algorithm yielded a better accuracy in classification.

Apart from creating a classification model, correlational analysis was performed to find attribute pairs that can give more insight in differentiating between an unhealthy and healthy spine.

II. LITERATURE REVIEW

In [7] a review on the significance of lower back carried out by analyzing 165 studies from 54 countries all around the world. These studies were carried out on general population between the year 1980 and 2009. The results from the review included, women between the age 40 -80 were the highest number of individuals from the study who suffered lower back pain. It was clearly seen that as a human being ages the likeliness to suffer from lower back pain is higher. This number is going to increase over time as mentioned in the paper.The current study is based on the work done [8] by Gaonkar who performed an analysis on lower back pain symptoms dataset. This dataset contained 12 attributes related to lumbar vertebrae and sacral segments of the spine. In this paper the author performs PCA in order to detect the outliers and also to perform feature selection. After performing PCA on the data, the number of observations were reduced from 381 observations to 310 observations. This data was then used to train classification models such as K nearest neighbor, Random forest, J48 Decision tree and Support vector machines. The accuracy percentage obtained from each of the models are 85.48%, 87.09%, 90.29%, 85.48% respectively. It was found that degree spondylolisthesis to be the most prominent factor in determining if an observation should be classified under healthy or unhealthy spine.

In 2002 [9] a study carried out by Darrel S Hanson, forty patients with spondylolisthesis were considered for analysis. These patients were divided into two groups, one that had low grade and the other one with high grade spondylolisthesis. Parameters like lumbar sagittal alignment, sacral inclination, slip angle and pelvic incidence were measured using radiographic methods. Pearson correlational analysis was done between the measurements taken and it was found that there was a high positive correlation between pelvic incidence and both low-grade and high-grade spondylolisthesis. This analysis was also done with pediatric and adult age groups where the results were consistent irrespective of the age group.In a study [10] carried out by Hubert Labelle in the year 2004 on the sagittal alignment in patients with developmental spondylolisthesis. This study included 214 subjects with spondylolisthesis between L5 and S1 vertebra column. The radio graphs of their spine and pelvis was taken which included parameters like pelvic incidence, sacral slope, pelvic tilt, lumbar lordosis, thoracic kyphosis and degree of spondylolisthesis. Pearson correlation was done on all of the measurements taken. It was found that pelvic incidence, sacral slope, pelvic tilt and lumbar lordosis were significantly higher in patients with high degree of spondylolisthesis and also the thoracic kyphosis is decreased in these patients. There was high linear correlation between pelvic incidence with sacral slope, pelvic tilt and lumbar lordosis in patients with high degree of spondylolisthesis. This was also compared to 160 healthy subjects and it was found that the differences between the two sets of subjects in terms of correlation between the measurements increase proportional to the degree of spondylolisthesis.

In a retrospective clinical study [11] conducted by author Zhenjiang Ma, new parameters were discovered in order to describe the pelvic anatomy and lumbosacral segmental deformity in children with high degree of spondylolisthesis. This is more specific to the L5 vertebra. The modified parameters were pelvic incidence and lumbosacral angle. the study included 24 children with high degree of spondylolisthesis. The spine parameters were recorded using radiography. The modified parameters of pelvic incidence and lumbosacral angle were taken and compared with 152 children without spondylolisthesis. The results concluded that the modified pelvic incidence and modified lumbosacral angel can clearly describe the pelvic morphology and the deformity in the spine in L5 vertebra where high degree of spondylolisthesis was observed.

A study [12] done by author Piotr Janusz clearly states the importance of the parameters namely Pelvic incidence, pelvic tilt and sacral slope. It was found that the orientation of the pelvis determines the accuracy measure of the important parameters. This indicates the importance of the pelvis and all parameters related to it when we are considering any kind of study on the spine.

A study published by [13] Maybin K Muyeba in the year 2013, talks about how lower back pain affects many people. The paraspinal muscles are observed to be overloaded or strained in lower back patients. There is a clear distinction between the paraspinal muscle activity in healthy and lower back pain patients. This study included parameters like physical activity, psychological factors which were quantified using numerical methods and questionnaires. Using fuzzy association rule mining techniques, the relationship between the values were determined. This is due to the fact that the data recorded were numerical and there was a certain degree of randomness within the data. The experiments were carried out on authentic lower back pain clinical data which contained the relationship between different aspects of the data and also rules which were used for interpretation.

In the paper published in conference [14] in the year 2009 by Ori Hay, the paper talks about modeling spine curvature of a healthy spine and of a spine that is not healthy. There could be lot of factors by which the curvature of the spine can change to cause spinal issue and the main factor is posture. At any given time, the person should have proper posture so that there is no strain on the back. During any given activity like sitting, standing, walking, sleeping and other physical activities, the person should make sure there is no strain on the back. This paper proposes a method for analyzing spine curvature in 3D using CT imaging. The subjects in the paper are healthy individuals, scoliosis patients and patients with chronic lower back pain. The assessment was carried out by first performing spinal canal segmentation where the dimensions of the spinal canal is recorded using several axial slices each of width 10mm which are put together to form a 2D image.

Spine curve extraction was performed so that irrespective of the height of the person the curvature value was found so that analysis on these values is not affected by the height of the person. Finally, the curvature values were analyzed to find curvature deformities that causes lower back pain.

III. DATASET

The dataset used in this research is obtained from Kaggle, called the lower back pain symptoms data set. It contains 310 observations with 13 attributes. The 13th attribute is the class label which has 2 class labels namely "healthy" and "unhealthy". All the 12 attributes contain numerical physical spine data. The first attribute is [15] pelvic incidence which is the angle between the line perpendicular to the sacral plate at its midpoint and the line connecting this point to the femoral head axis. Next is Pelvic Tilt which is the orientation of the pelvis with respect to the thigh bones and the rest of the body. Next is Lumbar Lordosis Angle which is the angle of curvature of the lumbar region of the spine. The Sacral Slope is measured between the tangent line to the superior endplate of S1 and the horizontal plane. Pelvic radius is the line drawn from hip axis to the posterior point of the Sacral Endplate. Degree Spondylolisthesis is the degree by which a particular vertebral body has slipped forward with respect to the vertebral body below it. The dataset obtained was already preprocessed. This was done in the previous study which was based on this dataset. The original dataset contained 381 observations and it was reduced to 310 using [16] PCA analysis. The outliers were removed in the process. It was also found that all 12 attributes were required for the analysis after performing PCA.

IV. METHODOLOGY

The classification models used in this research are Decision Tree and Support Vector Machines. Decision Tree is used for classification and regression. In this paper Decision Tree is used for classification of patient spine curvature data as unhealthy or healthy. [17] Decision Tree works by first calculating information gain of every attribute in the dataset. The attribute with the highest information gain will be the root node of the tree. The succeeding levels in the tree has attributes with information gain in descending order. There are two Decision Tree algorithms used in the paper for performance comparison namely Ctree and C50. Out of these algorithms [18] Ctree and [19] C50 performs automatic pruning based on information gain values of each attribute. The accuracy of classification depends on the concept of underfitting and overfitting. In the case of underfitting, if fewer attributes are chosen for the tree then the rules generated for classification may not be able to classify certain observations accurately. This is because all attributes that influences which class label, the observation belongs to is not considered while training the model. In case of overfitting, due to the fact that we have chosen more or all of the attributes, the attributes which do not really have any influence on the observation may induce an error while training the model. This can again lead to misclassification. When the right set of attributes are chosen to create the decision tree then the accuracy of the model increases.

The information gain [20] is calculated using the following steps Firstly, the amount of information needed to classify a tuple in dataset D is given by

$$info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (1)$$

Where p_i is the probability that a tuple in D belongs to class C_i and this is calculated by the formula $|C_{i,D}|/|D|$. This is calculated for all observations belonging to a particular class. Once the information value is calculated we now use this to calculate the information of each attribute in the dataset. This is calculated as follows

$$info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times info(D_j) \quad (2)$$

$\frac{|D_j|}{|D|}$ is calculated to differentiate each partition of observations that belong to a particular class label from another partition. This value is multiplied with the information value calculated using the previous formula. This gives the amount of information required to classify an observation from D based on the attribute that we are calculating for.

Finally, we calculate the gain by the formula:

$$Gain(A) = info(D) - info_A(D) \quad (3)$$

At each level in the tree, the attribute chosen for branching is determined by gain value of that attribute. The attribute with the highest gain value is chosen as root node and the attributes in successor nodes are chosen in decreasing order of information gain. The other supervised learning algorithm used is support vector machines. SVM uses a discriminant function that attempts to separate or distinguish the two sets of observations using a hyperplane. It generates this hyperplane by finding the support vectors in the dataset. Support vectors are observations that lie close to the hyperplane. These observations do not follow the same trend or they do not have similar characteristics than other observations that are present among the same class label. They are not erroneous but they are extreme cases within the same class label. The support vectors are the most difficult to classify. So, the idea here is to clearly distinguish the support vectors or the extreme cases in each of the Class Labels. Then the hyperplane can easily distinguish the other data points or observations which follow a similar trend. In training set of k number of samples, it is interpreted by

$$\{x_i, y_i\}, i = 1..k, \quad (4)$$

where $x \in R^N$, belonging in N- dimensional space and $y \in \{-1, +1\}$ which is used to represent the two class labels. If there exists a vector w and a scalar value b such that

$$y_i(w.x_i + b) - 1 \geq 0 \quad (5)$$

then these observations are said to be linearly separable.

V. IMPLEMENTATION

The tool used for performing analysis on the data is R. The aim was to create a classification model with higher accuracy than existing research done. Out of 310 observations, 70% of the data was used for training and 30% of the data was used for testing. The sample function from "Base" package was used to obtain random observations for training set. The sample function was run three times on the entire dataset in order to generate three sets of training and testing data. This was done in order to get an average performance score of each model generated using all of the classification methods used in this research. The results were plotted using the "plot" function from "graphics" package was used to plot the decision tree generated. Furthermore, the correlation of each attribute with every other attribute was calculated using the function "cor" in "ggpubr" package. The results were then plotted using scatter plot for each pair of attributes that have high correlation.

VI. RESULTS AND DISCUSSION

After training the model for each algorithm, the testing data set was passed to it in order to compute the accuracy. As mentioned earlier the average accuracy was calculated by passing three sets of training data. The accuracy measures are shown below.

Table 1: Accuracy score of each model

Classifier Model	Accuracy rate			Average Accuracy score
	Training Set1	Training set 2	Training set 3	
Ctree	90.37%	95.52%	92.83%	92.80%
C50	85.11%	80.30%	82.96%	82.79%
SVM	89.03%	88.66%	87.62%	88.43%

It is observed that the Ctree algorithm produces highest accuracy. This algorithm automatically prunes the tree considering the top most significant attributes. The tree generated from the model is shown below.

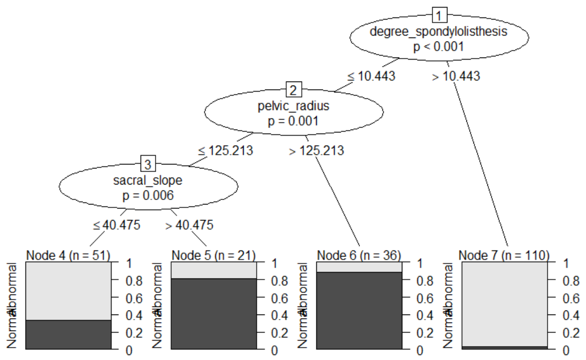

Figure 1: Results of Ctree Decision tree algorithm

As mentioned in earlier studies the degree spondylolisthesis is the most significant factor that contributes to the recognition of an unhealthy spine. Next comes pelvic radius, followed by the sacral slope. In order to obtain more information about the significant factors that are observed in patients with healthy and unhealthy spines, correlation analysis was performed for each attribute with every other attribute. Correlational analysis was performed separately for unhealthy spine data and healthy spine data and also on the entire dataset. This was done to see if there is any difference in how strongly attributes are correlated between both the class labels.

It was observed that there were 6 pairs of attributes that had high correlation among all other attribute pairs. Out of them, 3 pairs of attributes are taking into account all of the observations in the dataset since there was no significant changes from unhealthy and healthy spine data. The rest of the attribute pairs had significant difference in the correlation coefficient values between healthy and unhealthy spine observations. The results are plotted using ggscatter plot function as shown below.

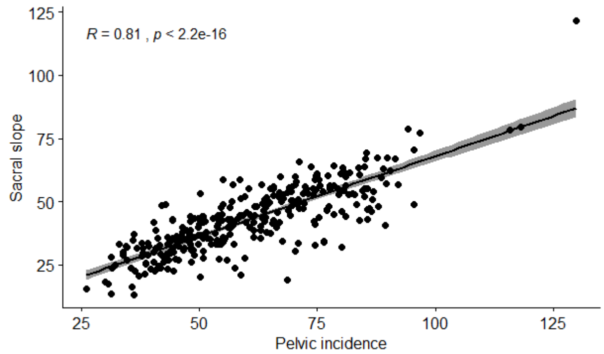

Figure 2: Correlation between pelvic incidence and sacral slope

As observed in Figure 2, the highest positive correlation was found between pelvic incidence and sacral slope with a correlation coefficient of 0.814. We see that as pelvic incidence increases, sacral slope also increases.

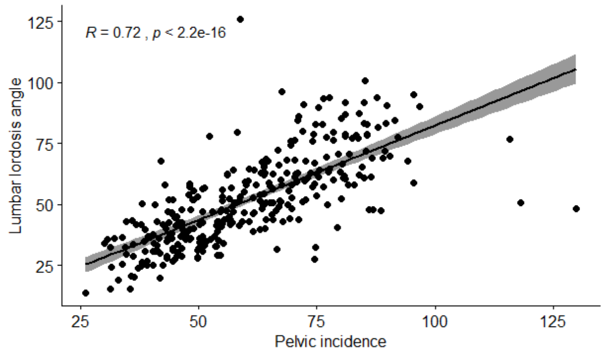

Figure 3: Correlation between Pelvic Incidence and Lumbar Lordosis

As observed in Figure 3, the second highest positive correlation was found between Pelvic Incidence and Lumbar Lordosis angle with a correlational coefficient of 0.717 and also as Pelvic incidence increases the Lumbar Lordosis angle also increases.

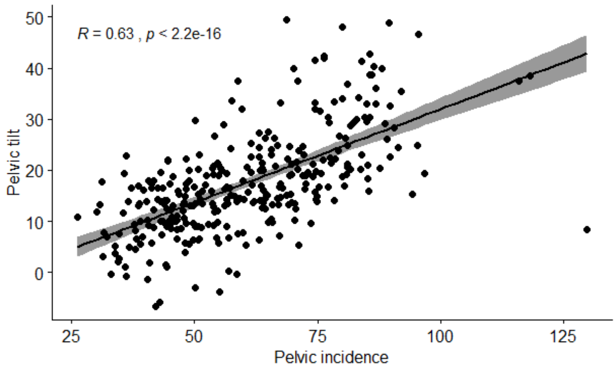

Figure 4: Correlation between Pelvic incidence and pelvic tilt

As observed in Figure 4, Pelvic incidence and pelvic tilt were found to have high correlation with a correlation coefficient of 0.629. As pelvic incidence increases, the pelvic tilt also increases.

It was observed there were attribute pairs that had difference in correlation with respect to a healthy and an unhealthy spine data. Out of which three pairs of attributes has significant difference between the sets of data.

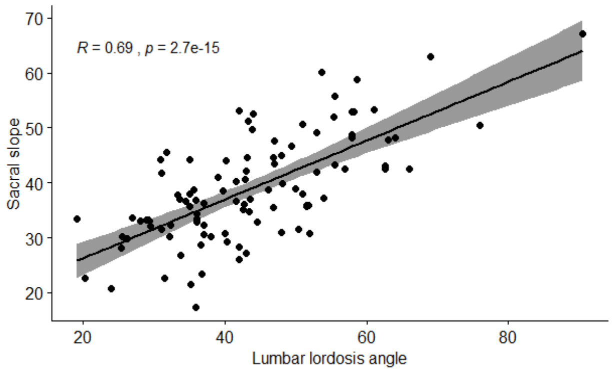

Figure 5: Correlation between Lumbar lordosis angle and Sacral slope for healthy spine data

As observed in Figure 5, the correlation coefficient between Lumbar lordosis angle and Sacral slope was found to be 0.687.

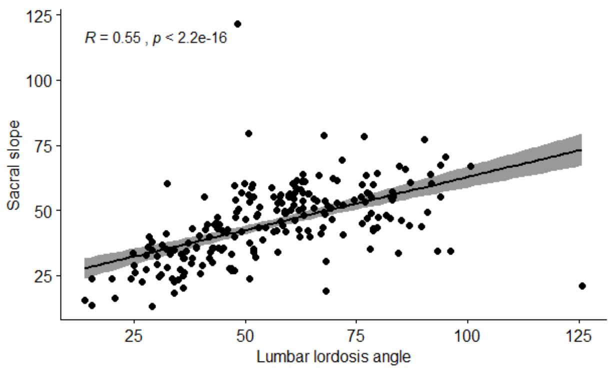

Figure 6: Correlation between Lumbar lordosis angle and Sacral slope for unhealthy spine data

Whereas in Figure 6 it is observed that there is a decrease in correlation coefficient by 0.14. This is observed in unhealthy spine data.

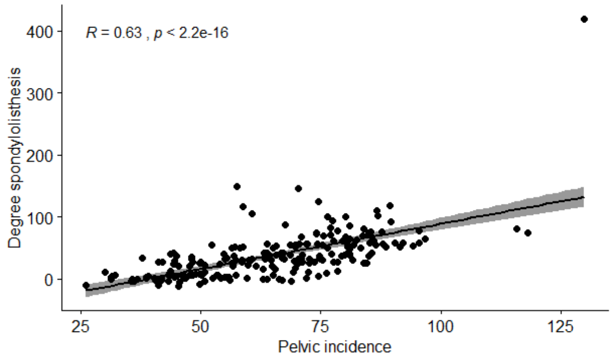

Figure 7: Correlation between pelvic incidence and degree spondylolisthesis for unhealthy spine data.

In Figure 7 there is a high positive correlation between pelvic incidence and degree spondylolisthesis with a correlation coefficient of 0.631. This observation was made in unhealthy spine observation

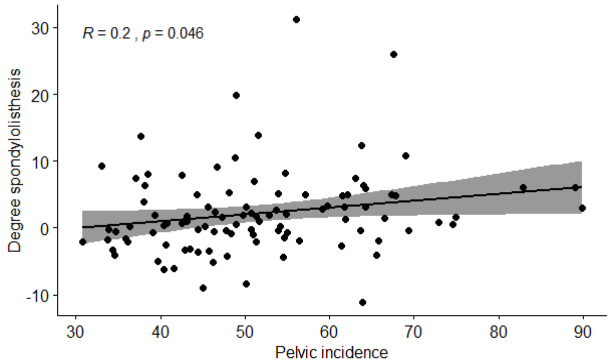

Figure 8: Correlation between pelvic incidence and degree spondylolisthesis for healthy spine data.

Whereas in healthy spine observations, there is very less correlation between pelvic incidence and degree spondylolisthesis with a correlation coefficient of 0.20 as seen in Figure 8.

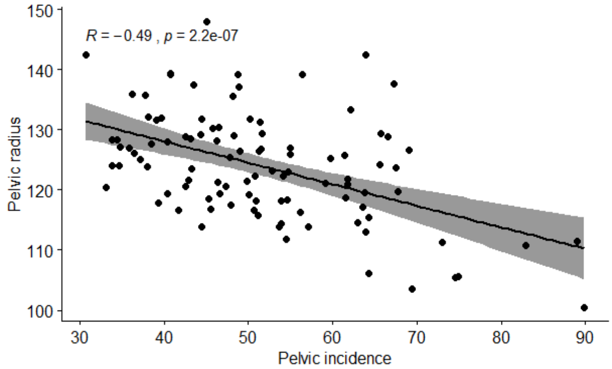

Figure 9: Correlation between Pelvic incidence and pelvic radius in healthy spine data.

In Figure 9 a negative correlation between Pelvic incidence and pelvic radius with a correlational coefficient of -0.49 is observed. This observation was made in healthy spine data.

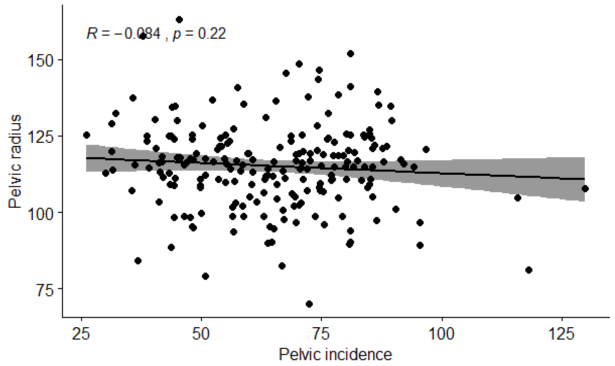

Figure 10: Correlation between Pelvic incidence and pelvic radius in unhealthy spine data.

In Figure 10, we see a significant change in the correlational coefficient between pelvic incidence and pelvic radius is unhealthy spine data. The correlation coefficient was found to be -0.084 which is significantly less correlated when compared with healthy spine observations.

VII. CONCLUSION

The human body is built to experience pain for a specific reason.[21] It is the mechanism of the physical body to signal us that the part of the body experiencing pain needs to be given attention. It is a survival mechanism to ensure that we do not ignore any problem in that part of the body so that it doesn't lead to life threating consequences. The human spine is one of most important parts of the body. Any kind of damage to the spine can lead to permanent unhealthy functioning of the body depending on the part of the spine that is affected. As seen in previous studies lower back pain is one of the most common types of back pain experienced by people all around the world. If any abnormality in a person's spine is detected at an early stage then the patient can be treated to completely eliminate the problem before it leads to further complications. With the help of computer science, we can create technology that can detect these abnormalities at an early stage.

Apart from finding the attribute with highest significance (degree spondylolisthesis) while creating the classification model, we observe that in Figure 2,3 and 4 Pelvic incidence seems to have a direct impact on sacral slope, lumbar lordosis and pelvic tilt values. This indicates that pelvic incidence has a significant impact on the lower spine parameters. This is with respect to the entire dataset. When we observe the correlation values for healthy and unhealthy spine data separately, we see that there is a clear difference in correlation between Lumbar lordosis and sacral slope (Figure 5 and 6), Pelvic incidence and degree spondylolisthesis (Figure 7 and 8), Pelvic incidence and pelvic radius (Figure 9 and 10). These pair of attributes can be used to distinguish unhealthy and healthy spine much more accurately than considering only degree spondylolisthesis when considering the classification model.

VIII. FUTURE ENHANCEMENT

This study was performed on 310 observations. Since we are dealing with a medical problem, there is a requirement of higher number of observations to create more accurate models to classify an unhealthy or healthy spine. The current research yielded an accuracy percentage of 92.80%. This rate can be increased by implementing different appropriate algorithms

REFERENCES

1. C. Owen Lovejoy, "The natural history of human gait and posture Part 1. Spine and pelvis" Gait & Posture Volume 21, Issue 1, January 2005, Pages 95-112
2. Manohar M. Panjabi, Thomas Oxland, Koichiro Takata, Vijay Goel, Joanne Duranceau, Martin Krag, "Articular Facets of the Human Spine Quantitative Three-Dimensional Anatomy" SPINE Volume 18, Number 10, pp 1298-1310
3. Raja S Alomari, Jason J Corso, Vipin Chaudhary,Gurmeet Dhillon, "Lumbar Spine Disc Herniation Diagnosis with a Joint Shape Model" Computational Methods and Clinical Applications for Spine Imaging pp 87-98
4. L.J.Gibson, "The mechanical behaviour of cancellous bone" Journal of Biomechanics Volume 18, Issue 5, 1985, Pages 317-328
5. Frank H.Netter,"Atlas of Human anatomy" Saunders 24th March 2014
6. Marwan El-Rich, Aboulfazl Shirazi-Adl, Navid Arjmand, "Muscle Activity, Internal Loads, and Stability of the Human Spine in Standing Postures: Combined Model and In Vivo Studie" SPINE Volume 29, Number 23, pp 2633–2642
7. Damian Hoy, Christopher Bain, Gail Williams, Lyn March, Peter Brooks, Fiona Blyth, Anthony Woolf, Theo Vos, Rachelle Buchbinder, "A Systematic Review of the Global Prevalence of Low Back Pain" ARTHRITIS & RHEUMATISM Vol. 64, No. 6, June 2012, pp 2028–2037
8. Abhijitsingh Putu Gaonkar, Ritwik Kulkarni, Ronnie D. Caytiles, N.Ch.S.N Iyengar , "Classification of Lower Back Pain Disorder Using Multiple Machine Learning Techniques and Identifying Degree of Importance of Each Parameter", International Journal of Advanced Science and Technology Vol.105 (2017), pp.11-24
9. Darrell S. Hanson, Keith H. Bridwell, John M. Rhee,Lawrence G. Lenke, "Correlation of Pelvic Incidence With Low- and High-Grade Isthmic Spondylolisthes" SPINE Volume 27, Number 18, pp 2026–2029
10. Hubert Labelle,Pierre Roussouly, E´ric Berthonnaud, Ensor Transfeldt, Michael O'Brien, Daniel Chopin, Timothy Hresko, Joannes Dimnet, "Spondylolisthesis, Pelvic Incidence, and Spinopelvic Balance" SPINE Volume 29, Number 18, pp 2049–2054
11. Zhenjiang Ma, Changqing Zhao, Kai Zhang, Lin Du, Jianping Tian, Xiaojiang Sun, MD, Jie Zhao, "Modified Lumbosacral Angle and Modified Pelvic Incidence as New Parameters for Management of Pediatric High-grade Spondylolisthesis" Clin Spine Surg Volume 31, Number 2, March 2018
12. Piotr Janusz, Marcin Tyrakowski, Jad Bou Monsef, Kris Siemionow, "Influence of lower limbs discrepancy and pelvic coronal rotation on pelvic incidence, pelvic tilt and sacral slope" Eur Spine J. 2016 Nov;25(11):3622-3629
13. Maybin K. Muyeba, Sandra Lewis, Liangxiu Han, "Understanding Low Back Pain using Fuzzy Association Rule Mining", in Proceedings of International Conference on Systems, Man, and Cybernetics.
14. Ori Hay, Israel Hershkovitz, Ehud Rivlin, "Spine Curve Modeling for Quantitative Analysis of Spinal Curvature" in Proceedings of International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009
15. Paul C. Celestre, John R. Dimar, Steven D. Glassman, "Spinopelvic Parameters: Lumbar Lordosis, Pelvic Incidence, Pelvic Tilt, and Sacral Slope What Does a Spine Surgeon Need to Know to Plan a Lumbar Deformity Correction?" Neurosurgery Clinics of North America, Volume 29,Issue 3, July 2018, Pages 323-329
16. Tao Chen, Elaine Martin, Gary Montague, "Robust probabilistic PCA with missing data and contribution analysis for outlier detection" Computational Statistics and Data Analysis 53(2009) 3706–3716
17. Witold Pedrycz, Zenon A. Sosnowski, "C–Fuzzy Decision Trees" IEEE Transactions on systems, man, and cybernetics—part c: applications and reviews, Vol. 35, no. 4, November 2005
18. Tim Häring, Elke Dietz, Sebastian Osenstetter, Thomas Koschitzki, Boris Schröder, "Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils" Geoderma 185–186 (2012) 37–47

19. Rafael Muñoz-Mas, Shinji Fukuda, Paolo Vezza, Francisco Martínez-Capel, "Comparing four methods for decision-tree induction: A case study on the invasive Iberian gudgeon (Gobio lozanoi; Doadrio and Madeira, 2004)" Ecological Informatics 34 (2016) 22–34
20. Jiawei Han, Micheline Kamber, Jian Pei Data Mining:Concepts and Techniques, 3rd Edition
21. MJDV Good, P Brodwin, B Good, A Kleinman Pain as Human Experience: An Anthropological Perspective 1994

AUTHORS PROFILE

Richie V Johny is a PG scholar in the Department of Computer Science in CHRIST (Deemed To Be University).His research interests include Artificial Intelligence, Machine Learning, Data Mining and Data Analytics.

Prof Roseline Mary. R currently serves as an Assistant Professor in the Department of Computer Science, CHRIST (Deemed To Be University). She obtained her BSc (Mathematics), MCA from Bharathidasan University, MPhil from Alagappa University and MTech from Karnataka State Open University and currently pursuing her doctorate from Bharathidasan University in Deep Learning. Her Research interests include Database Systems, Big Data, Data Mining and Deep Learning..