Diagnosis of Vertebral Column Disorders using A Novel Sprint Algorithm

K N Nithya, P. Suresh

Abstract**:** ***Data mining in the medical field has witnessed huge popularity and their applications gained remarkable impact. In the past, disease diagnosis is considered as a tedious task because of the inaccuracy and time consumptions. This raises the necessity of a prominent disease analysis system which would save thousands of lives. To achieve this especially in data mining data mining is vital, its classification accuracy will be an efficient remedy for proper diagnosis of diseases. In the world, people are affected by various kinds of diseases, in which Lower back pain has gained attention in recent years. The main challenge is the detection of the healthy and unhealthy spine. In this paper, we proposed a SPRINT algorithm for achieving better classification results. The proposed concept is a key basis of Decision Tree which considers lumbar and sacral parameters that perform effectively on detecting unhealthy spines. The experimental result is carried out with three sets of datasets on WEKA, a perfect and popular data mining suite. The obtained results are compared with K-NN and rep-tree on the basis of several parameters. It is proven that on this comparison classification accuracy obtained by SPRINT algorithm is far better than k-nn and rep-tree thus ensuring its overall performance.***

Keywords: Data mining, classification, disease diagnosis, and Lower back pain.

I. INTRODUCTION

The vertebral column is the backbone or spine, considered as the major part of the axial skeleton. The Vertebral column is structured with 33 vertebrae, in which 7 cervical, 12 thoracic, 5 lumbar, and 5 fused sacral vertebrae with 4 frequently fused coccygeal vertebrae. The vertebral column is responsible for the human flexibility this reason makes its importance at the peak on the medical field. This vertebral column is comprised of other parts such as nerves, inter-vertebral discs, medulla, muscles, and group of vertebras. Apart from body movement, it is responsible for medulla spinals, human body support and nerve center protection [1, 2]. There are several reasons for vertebral column disorders in which anatomic or life condition is the major one [3]. In the medical field, recognized vertebral column disorders are disc herniation and spondylolisthesis [4]. In disc hernia, the reason for severe pain is the disposal of inter-vertebral discs from its place and nerve compresses [5]. The forward slippage of the vertebra and Nerval presses are the reason for spondylolisthesis [6]. The Vertebral column pathologies are placed mostly in the lumbar region; the MRI or other radiology images are best for detecting those pathologies [7]. An experienced doctor can easily identify the disorders visually by means of axial or sagittal images. For a perfect precise diagnosis, the pelvis, lumbar spine's shape along with orientation is more important. These part clearly states the pelvic tilt, pelvic radius, pelvic incidence, sacral slope, lumbar lordosis angle and grade of spondylolisthesis [8]. An expert doctor can easily identify the disorders with these parameters at normal range. It's not to finalize that these parameters itself enough to confirm the diagnosis there is some exceptions also [9]. To overcome this problem several computer-aided automatic detection methods are evolved in recent years [10 – 16]. The existing method use the classification process for separating normal, hernia and spondylolisthesis cases. But still these parameters not sufficient for diagnosing both disorders, hence the classification need to done for each disorder separately.

This research work is organized as follows, section 2 with

II. RELATED WORKS

describing various method and ideas taken to overcome these issues. In this, we describe the problem statement, which paves a way for the research contribution. Section 3, states the proposed methodology with our contributions. It also describes the data sets, tools, proposed architecture along with the proposed algorithm. Section 4 describes the results and discussions. Section 5 completes the work with the conclusion and future scope. 2. Related Works Damian Hoy [17] et al. proposed a systematic review of the global prevalence of low back pain. In this work, the author detects the main symptom spinal issue as curvature on the spinal. The author observed the results by conducting 165 studies from 54 countries. The experiment is carried out with patients suffering from lower back pain between the ages of 40 – 80. The main aim of this work is separating healthy and unhealthy spines. His observation shows the highest prevalence is found among female individuals. The author uses methodologic variation for judging the highest prevalence. Abhijitsingh Putu Gaonkar [18] et al. proposed Classification of Lower Back Pain Disorder by means of Multiple Machine Learning Techniques. In this work, the author aimed to address the disorder origin and setting of treatment for that diagnosis.

Manuscript published on 30 September 2019
* Correspondence Author

K N Nithya* , PG and Research Department of Computer Science, Shri Sakthi Kailassh Women's College1 Salem, Tamil Nadu India knnithya.10@gmail.com

P. Suresh, Department of Computer Science, Sowdeswari College Salem Tamil Nadu India. surbhoo71@gmail.com

The experiment is carried out on the dataset having 12 attributes which are connected to lumbar vertebrae and sacral segments of the spine. In the proposed work author applies Principal Component Analysis (PCA) for detecting feature selection. This work is best for dealing with spondylolisthesis on healthy and unhealthy spines.

Jan Egger [19] et al. proposed a vertebral body segmentation with GrowCut. The author aimed to develop a better Imaging and analysis of a vertebral column. In this work, the author proposes a cellular automata-based approach for segmenting the vertebral bodies. It segments the spine portion by applying T2-weighted magnetic resonance imaging acquisitions. The main advantage of this method is free from wrong interpretations, reducing interpretation errors and minimum time consumptions.

Piotr Janusz [20] et al, discussed the influence of Pelvic incidence (PI), pelvic tilt (PT) and sacral slope (SS) in spine alignment evaluation. In this experiment, the author proves that pelvis orientation determines the accuracy measure of important parameters. As the pelvis orientation plays a vital role in studying what kind of the spine. The major work on determining the coronal plane (CPR) on radiographic accuracy of PI, PT, and SS measurements.

Ori Hay [21] et al proposed Spine curve modeling for quantitative analysis of spinal curvature. The major contribution of the work is describing the modeling spine curvature of healthy and unhealthy spines. The author states the maintaining of healthy Spine curvature and posture is very important. Incorrect spine configuration leads to low back pain (LBP). The author detects spine curvature in 3D by using CT imaging. The proposed work based on two concepts such as deriving spine curvature from spinal canal centerline and evaluating the curve.

2.1 Problem Statement

Diagnosis of Lower back pain by means of data mining is the most popular topic in the research filed. Even though there are several methods are developed still a prominent accuracy in classification is a challenging one.

The major problems are described below;

- ✓ Time consumption
- ✓ Lack of classification accuracy
- ✓ Determining of spine curvature with details
- ✓ Accuracy measure of important parameters
- ✓ The need for automatic segmenting of vertebral bodies
- ✓ The need for method delivering highest prevalence on healthy and unhealthy spines

Let's discuss our proposed methodology in overcoming these issues, algorithms key concept and idea as below;

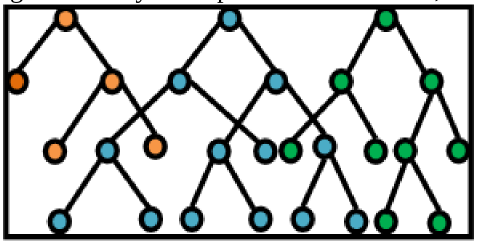

Fig-1 SPRINT structure

2.1.2 The Basic Idea of Improved SPRINT

The improved SPRINT algorithm states two data structures known as attribute table and histogram. The property sheet consists of three parts such as attribute value, class identification and indexing of data records. In the memory section, all the recorded data are not stored, on the hard discs, only the list of attributes can be kept. Attribute table separated with node expansion and associate with each corresponding child nodes. Histogram are connected with nodes to represent the type of attribute node distribution. On class distribution according to the aspect of numerical properties nodes correspond with two histograms Cbelow and Cabove. The former is responsible for stating samples type which is dealing with distribution. It also describes untreated samples distribution type and the value of both samples are taken with updates. In the class of discrete distribution, the attribute is stated by only one node on the histogram. For effective performance, the minimum description length principle is applied on SPRINT pruning.Table- II: Name of the Table that justify the values

The implementation details such type of data set used and tool for executing the task are stated in the below section. The reason for incorporating Data preprocessing as the major task was also provided below.

3.Dataset For this research, we have taken the dataset from UCI, which contains a huge collection of databases with domain theories and data generators. The UCI machine learning repository is widely used by the machine learning community for experimental study of machine learning algorithms. The main reason for using this dataset is it has the most required diagnostic parameters and best for students, educators, and researchers to do their comparison studies [12]. It contains the subject of 150 spondylolisthesis disorders, 100 healthy subjects, and 60 disk hernia disorders. Each dataset one consist of six attributes such as pelvic tilt, pelvic radius, pelvic incidence, sacral slope, lumbar lordosis angle and grade of spondylolisthesis. These attributes consist of six different biomechanical feature vectors which are perfect for diagnosing vertebral column pathologies

III. DATA PREPROCESSING

In a machine learning process, data pre-processing is more important especially in computational biology [18]. It separates unnecessary noisy and irregular information from the images at the training stage. This preprocessing makes the process free from troublesome and time-consuming, as data creation and filtering stages consume more processing time. Data pre-processing includes filtering, feature selection, feature extraction and transformation, standardization and instance determination. Most of the existing methods prove that data preprocessing is a popular technique in the machine learning process [6].

3.1 Data mining Tools

Similar to SAS Enterprise Miner, WEKA is one of the popularly known open source data mining tool. WEKA is a tool to utilize which enables the researcher to modify the algorithm code as per the requirement. The makes possible of re-implementations of multiple data mining algorithms including C4.5 known as J48.

Retrieval Number: C5630098319/2019©BEIESP
DOI:10.35940/ijrte.C5630.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Comparing to SAS Enterprise Miner WEKA is best, as SAS execute graphical user interface (GUI) which risky to robotize the tests. On WEKA creating experimentation are simple, which make the researchers possible of executing different variations by means of operating multiple modes. The proposed architecture along with way of processing task is clearly described below;

3.2 Architecture

In the research field, it is proven that J48 is one of the popular decision tree-based classification technique. Initially, the preprocessing is computed on the dataset for removing the noisy details. The preprocessing section is carried out with selected dataset from the UCI repository. Preprocessing removes the noisy from the sample and applies future selection for extracting the relevant features such as its attributes, instances, and sum of weights. According to the attributes the images are classified and forms tree structure respectively. The decision tree J48 is the extension of ID3 which is perfect for detecting the hidden pixels in the images. These classified images are arranged on leaf structure and get pruned. Then the classification process is executed by implementing SPRINT algorithm. Where the SPRINT undergoes training sets and pixels are labeled. According to the indexing labels, pixels are grouped, on every pixel, the extracted information's are under goes tested. From the resultant pixel, the best one is chosen which proving these classifiers are the perfect one for handling both discrete and continuous values. The proposed SPRINT algorithm is expressed below;

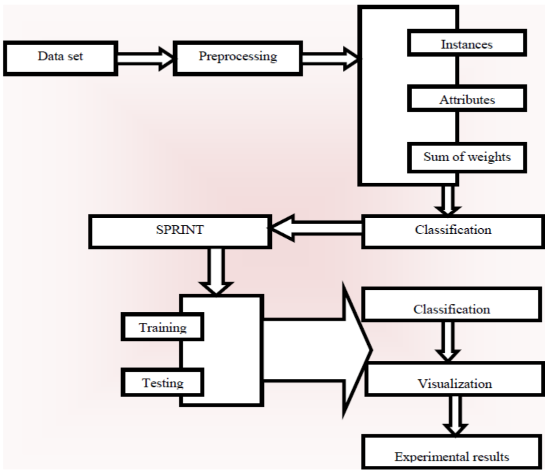

Fig-2 Architecture of SPRINT classification

Proposed SPRINT Algorithm

1. *Define nodes (n) #;*
2. *If (n) → find the termination condition;*
3. *Nodes to (q) → nodes into queue level as labeled at parent node;*
4. *Return ();*
5. *}*
6. *For (for each attribute A)*
7. *{*
8. *Update the histogram level ();*
9. *Calculate and performance evaluate for index segmentation ();*
10. *Each attributes segmentation points;*
11. *and find the best segmentation point ();*
12. *for two-part N1 and N2*
13. *make parent node N1;*
14. *make parent node N2;*
15. *}*
16. *}*
17. *End*

Definition

1. Check for the above base cases.

2. For each attribute A, find the normalized information gain ratio from splitting on a.

3. Let a_best be the attribute with the highest normalized information gain.

4. Create a decision node that splits on a_best.

5. Recur on the sublists obtained by splitting on a best, and add those nodes as children of the node.

Confusion Matrix

```
  a   b   <-- classified as
202   8 |   a = Abnormal
 19  81 |   b = Normal
```

Table 1 Performance Parameters

Correctly Classified Instances	283	91.2903 %
Incorrectly Classified Instances	27	8.7097 %
Kappa statistic	0.7948	
Mean absolute error	0.1381	
Root mean squared error	0.2627	
Relative absolute error	31.5629 %	
Root relative squared error	56.2066 %	
Total Number of Instances	310	

Table 2: Accuracy obtained by three classifiers As seen in the above table 2, it is clear that comparing to K-NN (81), REP-Tree (86), and SPRINT achieves the highest accuracy of about 91%. The experimental is carried out with UCI dataset on these three classifiers in a single classification process. It is executed on WEKA tool, an open source data mining tool. The process is executed at total training dataset at 0.2 seconds. The main factors taken for considering the performances are Correctly Classified Instances, Incorrectly Classified Instances, Kappa statistic, mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error and Total Number of Instances. The detailed values of these factors are clearly mentioned in table1. Table 3 shows the classification between normal and abnormal by means of the proposed algorithm.

Table 2: Accuracy obtained by three classifiers

Correctly Classified Instances (K-NN)	81%
Correctly Classified Instances (REP-tree)	86%
Correctly Classified Instances (SPRINT)	91%

The fig-3 obtained results are graphed to finalize the accuracy among the three classifiers in a clear manner. It is clearly showing that the proposed SPRINT is far better than K-NN and REP-Tree algorithm on the basis of classification accuracy.

Table3: Detailed Accuracy

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC	PRC	Class
0.962	0.190	0.914	0.962	0.937	0.798	0.953	0.963	Abnormal
0.810	0.038	0.910	0.810	0.857	0.798	0.953	0.887	Normal
0.913	0.141	0.913	0.913	0.911	0.798	0.953	0.939	

IV. RESULT AND DISCUSSION

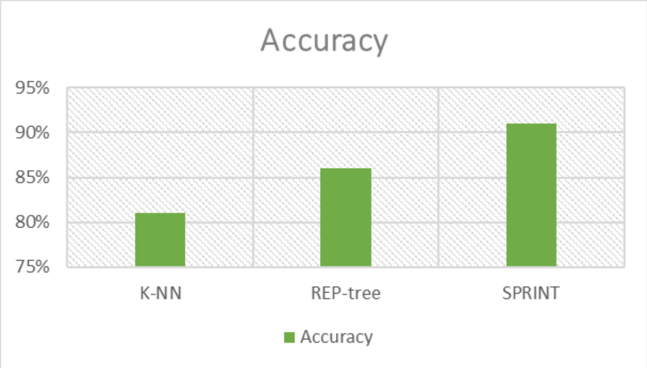

Fig: 3 Accuracy between the three classifiers

Fig-4 Detailed accuracy of three classifiers The above fig2 states the Detailed Accuracy among the class and determines the abnormal and normal cases. For this, the major parameters taken for estimation are TP rate, FP rate, Precision, Recall, F-Measure, MCC and ROC. All the obtained results are tabulated on table2 and a graphical view of observation is shown in fig2. From the figure, it is perfect to conclude the estimation of abnormal from normal by means of SPRINT rather than the K-NN and REP-Tree.

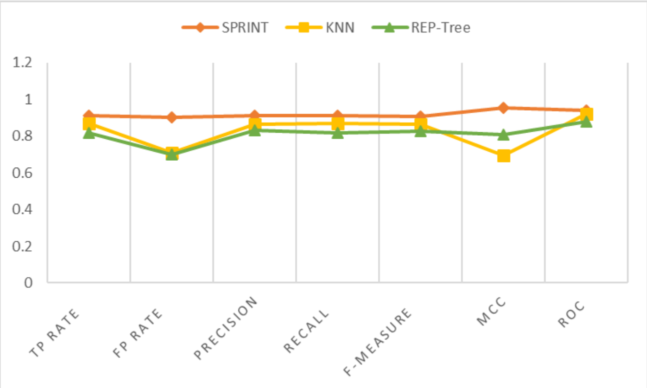

Fig-4 Detailed accuracy of three classifiers

V. CONCLUSION

In this study, the important spinal issue Lower back pain is analyzed on the aspect of data mining. This work is carried out to be effective in data mining. The main intention of this work is detecting unhealthy spines. It can be done by means of considering lumbar and sacral parameters. For this, we propose a SPRINT algorithm which works on the concept of Decision Tree. To prove the efficiency of the proposed algorithm is compared with K-NN and Rep-tree. The experimental results are carried out on WEKA tool using dataset. The performance is carried out by means of major three parameters such as accuracy, sensitivity, and specificity. All the obtained results are tabulated and graphed for considerations. It is clear from the observations that the proposed SPRINT algorithm is most prominent in detecting the abnormal spines from the normal spine compared to the other two methodologies. As Spinal in the medical field is a vast area, a deep study on the basis of automatic accuracy detection is a necessary one. The future work on the varied aspect of data mining can be carried out by improving several factors of our SPRINT algorithm. The extracting of details from every labeled pixel are concentrated on the future scope in order to attain more accuracy which saves a lot of time and helps the doctor in perfect diagnosis.. Although not everything need be disclosed, a paper must contain new, useable, and fully described information. For example, a specimen's chemical composition need not be reported if the main purpose of a paper is to introduce a new measurement technique. Authors should expect to be challenged by reviewers if the results are not supported by adequate data and critical details.

REFERENCES

1. Gadow HF. The Evolution Of The Vertebral Column, Cambridge University Press, 2014.
2. Netter FH. Atlas Of Human Anatomy, Professional Edition: Including Netter Elsevier Health Sciences, 2014.
3. Mcgill SM. Low Back Disorders, 3E. Human Kinetics, 2015.
4. Barrey C, Pinheiro-Franco JL, Le-Huec JC, Perrin G, Roussouly P. Compensatory Mechanisms Contributing To The Maintenance Of Sagittal Balance In Degenerative Diseases Of The Lumbar Spine In Advanced Concepts In Lumbar Degenerative Disk Disease, Springer Berlin Heidelberg, 2016.
5. Aebi M. Is Low Back Pain After Disc Herniation With Modic Type 1 Changes A Low-Grade Infection European Spine Journal, 2013; 22: 689-689.
6. Joglekar SB, Schwender JD. Minimally Invasive Spine Surgery In Lumbar Spondylolisthesis In Minimally Invasive Spine Surgery, Springer New York, 2014.
7. Unal Y, Kocer HE. Diagnosis Of Pathology On The Vertebral Column With Back Propagation And Naïve Bayes Classifier, In Technological Advances In Electrical, Electronics And Computer Engineering (TAEECE), International Conference On IEEE 276: 2792013
8. Kelm BM, Wels M, Zhou SK, Seifert S, Suehling M, Zheng Y, Comaniciu D. Spine Detection In CT And MR Using Iterated Marginal Space Learning, Medical Image Analysis 2013; 17: 1283-1292.
9. Canale ST, Beaty JH. Campbell's Operative Orthopaedics: Adult Spine Surgery, 2012;
10. Elsevier Health Sciences. 10) Been E, Kalichman L. Lumbar Lordosis The Spine Journal 2014; 14: 87-97.
11. Unal Y, Polat K, Kocer HE. Pairwise FCM Based Feature Weighting For Improved Classification Of Vertebral Column Disorders, Computers In Biology And Medicine 2014; 46: 61-70.
12. Hidayah I, Adhistya EP, Kristy MA. Application Of J48 And Bagging For Classification Of Vertebral Column Pathologies, In Information Technology And Multimedia (ICIMU), International Conference On IEEE 2014; 314-317. 13)

13. Abdrabou E. A Hybrid Intelligent Classifier For The Diagnosis Of Pathology On The Vertebral Column, Proc Of Intelligent Information And Engineering Systems 2012; INFOS.
14. Ansari S, Sajjad F, Naveed N, Shafi I. Diagnosis Of Vertebral Column Disorders Using Machine Learning Classifiers, In Information Science And Applications (Icisa), International Conference On Ieee 2013; 1-6
15. . 15) Roy Ag, Chatterjee A, Hossain Mk, Perrizo W. Fast Attribute-Based Table Clustering Using Predicate-Trees: A Vertical Data Mining Approach, Journal Of Computational Methods In Sciences And Engineering 2012; 12: 139-146.
16. Keles A. Expert Doctor Verdis: Integrated Medical Expert System, Turkish Journal Of Electrical Engineering & Computer Sciences, 2014; 22: 1 17)
17. 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 18)
18. 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
19. Jan Egger, Christopher Nimsky, And Xiaojun Chen. Vertebral Body Segmentation With Growcut: Initial Experience, Work Ow And Practical Application. Sage Open Medicine, 5:2050312117740984, 2017.
20. 20) 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
21. 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