# Developing new machine learning ensembles for quality spine diagnosis

Indrajit Mandal *

Computer Science and Engineering Department, Rajiv Gandhi Institute of Technology, Bangalore 560 032, India

A B S T R A C T

This research work adduces new hybrid machine learning ensembles for improving the performance of a computer aided diagnosis system integrated with multimethod assessment process and statistical process control, used for the spine diagnosis based on noninvasive panoramic radiographs. Novel methods are proposed for enhanced accurate classification. All the computations are performed considering steep error tolerance rate with statistical significance level of 5% as well as 1% and established the results with corrected $t$-tests. The kernel density estimator has been implemented to distinguish the affected patients against healthy ones. A new ensemble consisting of Bayesian network optimized by Tabu search algorithm as a classifier and Haar wavelets as the projection filter is used for relevant feature selection and attribute's ranking. The performance analysis of each method along with major findings is discussed using various evaluation metrics and concludes with propitious results. The results are compared to the existing SINPATCO platform that uses MLP, GRNN, and SVM. The optimization of machine learning algorithms is obtained using Design of Experiments scheme to achieve superior prediction accuracy. The highest classification accuracy obtained is 96.55% with sensitivity, specificity of 0.966 and 0.987 respectively. The objective is to enhance the software reliability and quality of spine disorder diagnosis using medical diagnostic system and reinforce the viability of precise treatment.

## 1. Introduction

A medical diagnostic system [1] generally consists of a knowledge base and some methods for solving an intended problem. On the basis of the query requested to the system, it outputs appropriate solutions and provides assistance to medical physicians in diagnosing patients. The knowledge base of such medical systems relies upon on inputs that spring up from the clinical experience of field experts [2]. Knowledge acquisition is the process to transform human expertise knowledge and skills acquired through clinical practice to software, besides being quite time consuming and labor intensive task. Common methods like Case Based Reasoning (CBR) solves the knowledge acquisition problem to some extent, in which the previous clinical cases consisting of patient's health symptoms in the form of the database is maintained with their possible remedies, past clinical decisions, preventive measures and expected diagnostic outcome measures. During patient diagnosis, the clinical database is explored for the most likely analogous past patient's record and provides the suitable diagnosis decisions for a new subject [3]. The opinion of physicians regarding the use of inference system is also appreciable in the literature [4]. The advantages of machine learning (ML) methods are that it uses mathematical models, heuristic learning and rules acquisition process for decision making and thus provides controllability, observability, stability and easily updateable by adding a new patient's record [5–7].

The application of machine learning models in the field of medicine for human disease diagnosis aids medical experts in the identification of diseases based on the symptoms at an early stage, though some diseases exhibit similar symptoms. Current advances in research for Spondylolisthesis and Disk Hernia treatment [8] since the past decades suggest that the related risk factors are identified. The state-of-the-art of technology is needed to be improved for significant accurate diagnostic decisions. The primary motivation of the research work is to design a reliable inference system and provide guidelines to caution against the selection of inaccurate inference system design for the clinical assessment of spine disorders, besides improving the quality of the diagnosis, prognosis, reduce delay in treatment and diagnosis cost [9].

## 2. Related works

This section provides a bird's eye view on the spine related advancements in the recent years. Kolta et al. [10] proposed a technique using the whole spine imaging based on standard DXA device

* Tel.: +91 8438366442.
  E-mail address: indrajit1729@gmail.com

(3D-XA) for 3D reconstruction of vertebral bodies in women. Masharawi [11] evaluated the shape characteristics of the Lumbar vertebral body (VB) and neural arch in Spondylolysis using a 3D digitizer and other parameters. Szeyko et al. [12] reported that vertebral infections caused by Coccidioides species need multidisciplinary management that always demands medical treatment and surgical intervention for stabilization. Dall'Ara et al. [13] reported that quantitative computed tomography based finite element modeling provide a quantitative and significant improvement in vertebral strength as compared to simulated dual energy X-ray absorptiometry. Yang et al. [14] introduced non-invasive fluoroscopy based image guided surgery that reduces radiation exposure due to vertebroplasty for vertebral compression fractures. Sapin-De et al. [15] gave subject-specific finite element models using low dose bi-planar X-ray devices for prediction of vertebral strength. Roberts et al. [16] achieved good accuracy for vertebral shape from lumbar radiograph using semi-automated active appearance models. Oh et al. [17] showed novel use of the liquid embolic agent Onyx injected within pseudoaneurysm, resulting in proximal and distal parent vessel closure. Wustrack et al. [18] introduced predictors of new vertebral fractures using data from the placebo arm of the HORIZON Pivotal Fractures. Roux et al. [19] showed that despite the use of PPI concomitant, there is reduced risk of new vertebral fractures during risedronate therapy as compared to placebo. Baum et al. [20] demonstrated that there is significant correlation between the volumes of vertebral bone marrow adipose tissue content to that of a volume of abdominal fat, spine bone mineral density using proton MR spectroscopy in women.

The major research contribution related to SINPATCO platform made by Rocha Neto et al. [21] in which they reported maximum accuracy of 85.9% with the implementation of SVM (linear and KMOD). In their second research article [22], they reported maximum accuracy of 90.32% by using Support Vector Machines (SVM), MLP (Multilayer Perceptron), GRNN (Generalized Regression Neural Network) and SOM (Self-Organizing Map). In this research paper, new ensembles are proposed and the results obtained over this problem using the same database are compared with the existing works. It reports higher precision, specificity and sensitivity values. Robust methods for treating vertebral column disorder including rotation forest ensemble comprising of bagging [23], Principal Components analysis (PCA) [24], decision trees [25], boosting methods [26], Haar wavelets [27], Naïve Bayes [28,29] and other machine learning algorithms are presented. A new ensemble for relevant feature selection is proposed that consists of rotation forest ensemble that employs Bayesian network [30] and Haar wavelets. The attribute ranking is procured using Ranker search algorithm. All the machine learning models are evaluated using statistical significance level of 5% and 1% along with corrected $t$-tests [31].

This research article is succinctly organized into various sections as follows. The radiograph pelvic spine data set with its various attributes and the proposed hybrid machine learning algorithms are presented in Section 3. The theory and calculations are briefly illustrated in Section 4. The design of the proposed inference system and the results obtained from the experiments are elucidated in Section 5. Finally, the Section 6 concludes with the interpretations of the accomplished results and their significance for further research.

## 3. Material and methods

### 3.1. Data

The Vertebral Column dataset consists of 310 instances and six attributes, having one class without any missing values. There are three different categories comprising of Spondylolisthesis, Disk Hernia and normal cases. The two categories Disk Hernia and Spondylolisthesis are combined into one category labeled as "abnormal". Each patient data was collected from a medical residence period in spine surgery at The Group of Applied Research in Orthopaedics (GARO), The Centre medico-Chirurgical de Réadaptation des Massues, Lyon, France by Dr. Henry Mota (Orthopedic Surgeons' Hospital Monte Klinikum). The database is obtained from sagittal panoramic radiographs of the spine of format $30 \times 90$ cm. The data are obtained from radiographs of patients operated for herniated discs (60 patients) and Spondylolisthesis (150 subjects) and the rest 100 subjects are volunteers who do not have any conditions in the column, called normal category. Each patient is described by six biomechanical attributes which are derived from the shape and orientation of the pelvis and lumbar spine: pelvic tilt, grade of spondylolisthesis, sacral slope, lumbar lordosis angle, pelvic incidence, and pelvic radius. All these biomechanical parameters of the vertebral column are discussed elaborately in [21,22].

The radiograph's measured features along with their corresponding values are presented using parallel coordinates plot as shown in Fig. 1 after preprocessing for its visualization of high-dimensional geometry and analyzing the multivariate pelvic spine data. For showing a set of distinct points in multidimensional space, a backdrop is projected that consists of $n$ parallel lines, projecting vertically and equally spaced. A point in the multidimensional space is reflected by a polyline with vertices lying on the parallel axes in such a way that the position of vertices on the $k$th axis corresponds to the $k$th coordinate of that point.

The machine learning ensembles have been trained with 80% of the entire clinical radiograph dataset and subsequently the degree of learning of these ensembles is tested on the remaining dataset, such that the testing is not performed on the same instances as it leads to biased evaluation. The data set is standardized in such a way that the overall mean and standard deviation is equal to 0 and 1 respectively. Then data cleaning methods are applied for dimension reduction as well as outlier's [32] removal. A new ensemble is developed for the appropriate feature selection and their ranking for the model's input and also the outliers from the dataset are removed based on the quantile information obtained statistically beyond 5% and 95%.

The research work primarily concentrates on the problem of finding optimal features needed for training and subsequent extirpate the frivolous features from the clinical dataset and additionally eliminates the latent outliers by applying felicitous techniques to develop a superior classifier for accurate prediction. In the literature, feature reduction techniques are commonly branched as extraction and selection techniques. The former converts the original features into a new feature space; the latter retains only minimum and best features from the available original ones. Both these techniques are used widely in data engineering applications for achieving computational complexity advantages while training the ensemble models. The Spearman's correlation matrix is shown in Table 1 and it reflects the degree of association among the variables of pelvic spine.

Here multiple threads are used for the generation of CHAID decision trees as it gives a simple and better interpretable method for multivariate analysis of pelvic spine data, shown in Fig. 3. The major advantages are ease of coding with modern programming languages, threads share information more efficiently, enables parallel execution and saves execution time. It's highly suitable for analyzing large medical data sets. The computation for goodness of split of the candidate attribute is optimized using a metric. The proposed pseudo code for splitting a node in the multithreaded CHAID decision tree is shown in Fig. 4. The tree is constructed using training data and then tested to obtain the LIFT-ROC curve
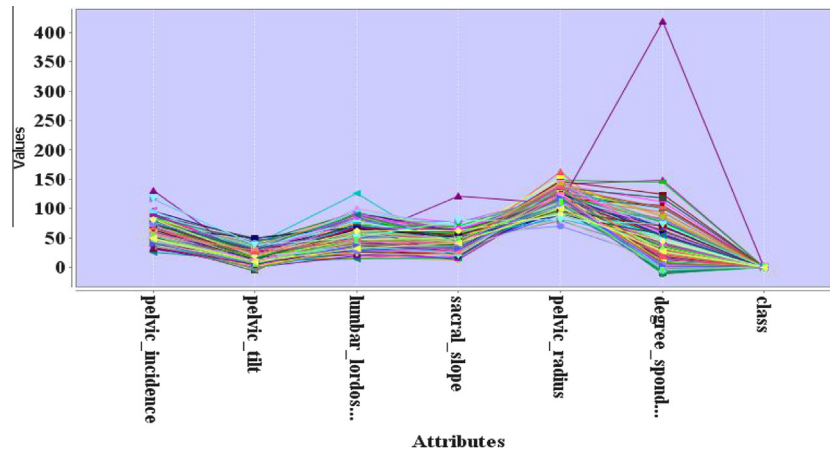
**Fig. 1.** Parallel coordinates plot for multi dimensional vertebral column radiograph dataset with *X*-axis as the measured parameters and *Y*-axis as the range of values.

**Table 1**
Shows Spearman's correlation coefficients between different measurable parameters obtained from the radiographs of patients involved in the clinical study ($p < 0.001$).

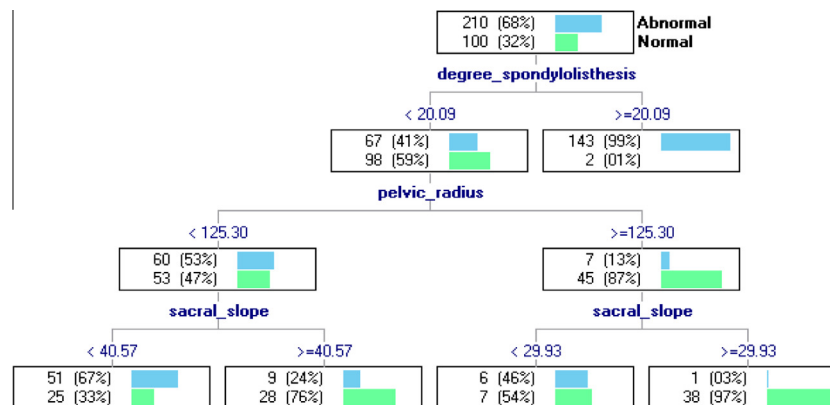| Attributes | pelvic_incidence | Pelvic_tilt | Lumbar_lordosis_Angle | Sacral_slope | Pelvic _radius |
|---|---|---|---|---|---|
| pelvic_incidence | 1 | | | | |
| Pelvic_tilt | 0.629 | 1 | | | |
| Lumbar_lordosis_angle | 0.717 | 0.432 | 1 | | |
| Sacral_slope | 0.814 | 0.623 | 0.598 | 1 | |
| Pelvic _radius | -0.247 | 0.032 | -0.080 | -0.342 | 1 |
| Degree_spondylolisthesis | 0.638 | 0.397 | 0.533 | 0.523 | -0.026 |



**Fig. 2.** Multithreaded CHAID decision tree for the radiographs of pelvic spine data: max depth = 4.
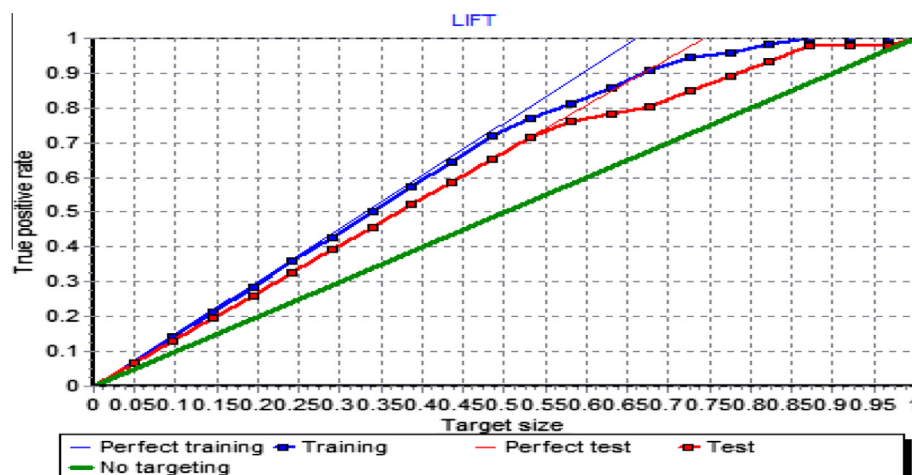


**Fig. 3.** Shows LIFT-ROC curve in which 80 percent of data is used for training and rest is used for evaluation. The graph shows training and testing curves.

```
Func_Split(Attribute_set) : chosen_attribute
High= - ∞
Chosen_attribute = NULL
For(i = attribute_1; i <= Attribute_set ; i++  )
    {
      Relevent_attribute = goodness_of_split(attribute)     // optimization of metric
      If( Relevent_attribute > High) Then
      High = Relevent_attribute ;
      chosen_attribute = attribute;
    }
Return(chosen_attribute)
```

**Fig. 4.** The proposed pseudo code for splitting a node in the multithreaded CHAID decision tree.

**Table 2**
Attribute ranking by Rotation forest ensemble comprising of Bayes network with Haar wavelets as projection filter optimized by Tabu search algorithm and Markov blanket using 10-fold cross validation. The ranks are generated using Ranker search method.

| Average merit | Average rank | Attribute |
|---|---|---|
| 77.061 ± 1.21 | 1 ± 0 | degree_spondylolisthesis |
| 58.566 ± 1.39 | 3 ± 0.89 | lumbar_lordosis_angle |
| 59.068 ± 0.782 | 3.1 ± 0.94 | pelvic_incidence |
| 58.566 ± 2.908 | 3.6 ± 1.11 | pelvic_radius |
| 55.233 ± 3.98 | 4.5 ± 1.2 | pelvic_tilt |
| 50.502 ± 0.65 | 5.8 ± 0.4 | sacral_slope |

for the abnormal class as shown in Fig. 2. It performs multi level splits and the probabilities for merging and splitting are 0.05 and 0.001 respectively.

### 3.2. Feature selection

One of the important problems in multivariate techniques is to select relevant features from the available set of attributes. The common feature selection techniques [33,34] include wrapper subset evaluation, filtering and embedded models. In case of embedded models, classifiers are used in constructing ensemble models; whereas wrapper subset evaluation method provides ranks to features based on their importance using base algorithms and filter methods rank the features based on statistical measurements. Combining wrapper and filter techniques provides superior computational results as compared to embedded methods, due to its having independent nature of classification models. Also embedded feature selection models provide better accurate results as it takes classification methods for enhancing the accuracy of the feature selection process.

In this work, the relevant feature selection is done by employing rotation forest algorithm that evaluates the worth of the features using Bayesian network as base classifier and Haar wavelets as the projection filter for the transformation of the data. Bayesian network learning uses search algorithm and quality measures. Simple estimator is used for the estimation of conditional probability tables of the network after the structure has learnt. It estimates the probability directly from the dataset by using alpha as a parameter for estimating the probability tables and can be interpreted as the initial count on each value. Here alpha has been set as 0.50. The Bayes network learning algorithm uses Tabu search [35] approach for finding a well scoring Bayes network structure. The Tabu search method is a hill climbing method that searches till an optimum is reached. The following step is the worst possible step. The last steps are kept in a list and none of the steps in this so called Tabu list is considered in taking the next step. The best optimized network structure formed during this process is returned. Entropy [36] is used as scoring metric to determine the measure for judging the quality of a network structure. After the network structure is learnt, Markov Blanket correction is applied to the network structure. This ensures that all the nodes in the network are part of the Markov blanket of the classifier node. In the Bayes leaning, Active Directory (AD) tree is used, so that the learning time of classifier goes down typically. Also AD trees are memory intensive and run with less memory. The ranks of the individual attributes are obtained by Ranker search with leave-one-out 10-fold cross validation. This work focuses on the problem of optimizing the feature selection and thereby filtering out the irrelevant variable from the dataset and removes the outliers by using appropriate methods to obtain higher accurate classifier. The ranks of all the spine dataset attributes with their description are provided in Table 2.

The methodology of the research work is grouped into four stages: (1) attribute extraction; (2) preprocessing and feature reduction; (3) optimization of classifiers by fine tuning their parameters; and (4) design of the proposed inference system. An important matter in any multivariate classification problem is handling suitable features in the sense that those input features combinations produce the best output. Even though lots of care has been taken during the data acquisition, it's practically impossible to get noise free data, while recording the feature's values using different automated software and tools. Also, all variables in the radiograph dataset may not be useful in the analysis. The working of machine learning methods and model are mentioned in this section with the parameters involved in each of them. This work addresses the advancement of diagnostic methods for pelvic spine diagnosis. We propose robust methods of treating vertebral column disorder, including rotation forest ensemble comprising random tree and Principal Component analysis, decision trees, boosting methods, Haar wavelets, Naïve Bayes and other machine learning algorithms.

### 3.3. Rotation forest – random tree PCA ensemble

Rotation forest (RF) [37] is an ensemble construction technique that consists of two components. The first is the heuristic component based on feature extraction. It is used to build a feature subset. This method improves the classification accuracy by improving the training space. The second is the classification component. A base classifier algorithm is used for constructing a classifier model. In this study, two flavors of RF are presented. In the first case, the base classifier is Random tree and Principal component analysis is used for feature extraction. The latter consist of decision table as classifier and haar wavelets as the feature extraction technique. The proposed ensemble is detailed below:-

Let $\mathbf{x} = [x_1, \ldots, x_n]^T$ be an instance given by $\mathbf{n}$ variables and $\mathbf{X}$ be the training sample in a form of $\mathbf{N} \times \mathbf{n}$ matrix. Let vector $\mathbf{Y} = [y_1, \ldots, y_N]$ be class labels, where $\mathbf{y_j}$ takes a value from the set. Let $\mathbf{D1}, \ldots, \mathbf{DL}$ be the classifiers in ensemble and $\mathbf{F}$ is feature set.

In ensemble learning, choosing $\mathbf{L}$ in advance and training classifiers in parallel is necessary.

Follow the steps to prepare the training sample for classifier $\mathbf{Di}$:

1. Divide $\mathbf{F}$ randomly into $\mathbf{K}$ disjoint or intersecting subsets. To maximize degree of diversity, disjoint subsets are chosen.
2. Let $\mathbf{F_{i,j}}$ be jth subset of features to train set of classifier Di.

Initialize Random tree with $\mathbf{F_{i,j}}$ subset of features such that randomly k attributes are chosen at each node.

Draw a bootstrap sample of instances of size 75% by selecting randomly subset of classes for every such subset. Run PCA to extract features for only $\mathbf{M}$ features in $\mathbf{F_{i,j}}$ and the selected subset of X. Store the coefficients of the computed principal components, $\mathbf{a_{ij[1]}}, \ldots, \mathbf{a_{i,j[Mj]}}$, each of size $\mathbf{M} \times \mathbf{1}$.

3. Arrange obtained vectors using coefficients in a sparse "rotation" matrix **Ri** having dimensionality **n** × $\sum$**Mj**. Compute the training sample for classifier **Di** by rearranging the columns of **Ri**. Represent the rearranged rotation matrix **Rᵢᵃ** (size N x n). So the training sample for classifier **Di** is **X Rᵢᵃ**.

4. In the classification phase, for a given input x, compute the confidence of each class by **m(x)** = $\frac{1}{L}\sum_{i=1}^{L}di$, $l$ **(xRᵢᵃ)**, $j$ = 1, 2, ..., $c$. Assign x to the class with largest confidence value.

### 3.4. Bagging-C4.5

Bagging is a procedure to increase the prediction accuracy of an algorithm by creating its multiple versions by aggregating and predicts a class based on plurality voting scheme. Here Gibbs classifier voting is used for aggregation first time and it provides better results.

The proposed method of bagging C4.5 is as follows:

1. Given training data set **D**, **{(Cₙ,dₙ), n = 1, 2, ..., N}** where **C** is class labels, **d** is data points. **D** is randomly divided into test set (say **T** with 20%) and learning set (say **L** with 80%).

2. Let **P(d,Lₖ)** be predictor model where **Lₖ** is specific k sequence learning set. Replace **P(d,Lₖ)** by its average as **Pₐ(d) = EₗP(d,L)** where **Eₗ** is expectations over **L**, **Pₐ** is aggregation.

3. The aggregation is done using Gibbs classifier using rule: **ȳ = sign(wᵀx) = f_GC(x,L)** as it draws a **w** from the distribution **p(w)**. Bootstrapping dataset **{L⁽ᴮ⁾}** and form **{P(d,Lᴮ)}**.

4. Form **C5** classification tree is constructed by 10-fold cross validation. If **C** is the class label, then **{P(d,Lᴮ)}** predict **P_B(d)** with the highest probability of $\int$**maxⱼP′(C|d)** * **P′ₓ(d)**. Pruning of based on confidence factor = 0.25.

5. Above steps are repeated till specified iterations and reports the optimized learned tree. Then the test data is run over the tree.

### 3.5. Improved NNGE

Nearest-Neighbor Generalized Exemplars (NNGE) [38] learns incrementally by classifying and followed by generalizing each new instance using Minkowski distance function that manages hyper-rectangles, feature weights, symbolic features and exemplar. Minkowski distance is given by Eq. (1)

$$\sqrt[p]{\sum_{i=1}^{n}|X_i - Y_i|^p} \tag{1}$$

Numeric values are standardized by dividing with the range of observed values. It uses dynamic looping to modify exemplar and feature weights after each new instance to be classified. While classifying a data point, several hyper-rectangles may be subset of that new example, but may belong to wrong class. It prunes these hyper rectangles so that the new instance is no more a member. After classification, the new instance is generalized by merging with nearest exemplar belonging to same class. It resolves affected area of feature space that conflict with the proposed new hyper rectangle. NNGE classifies an input instance by computing nearest neighbor in the created exemplar set using Minkowski distance function. It chooses a class having least distance.

The exemplar weight *WH* is *WH* = $(p + n)/p$. The feature weight *Wi* is initialized to 1 and increased or decreased according to the following rules:

- For correct classifications, raise *Wi* if feature in new exemplar not matched to that of chosen exemplar, otherwise reduce it.
- For incorrect classifications, raise *Wi* if the feature matched, otherwise reduce it.

### 3.6. Improved RIPPER

There are shortcomings with the original Repeated Incremental Pruning to Produce Error Reduction algorithm (RIPPER) with respect to optimization of rule set size and accuracy. The improved RIPPER is optimized with a different prune metric as described below:

Initially, the rule set is empty.
Grow ()
{
Grow the rule set by adding one rule
Based on highest information gain: $p(\log\left(\frac{p}{t}\right) - \log\left(\frac{P}{T}\right))$
}
Prune ()
{
Incrementally prune each rule
Based on metric: $(p + 1)/(p + n + 2)$
}
Optimization ()
{
Optimize the initial rule set to obtain minimum specified description length.
Repeat Grow () and Prune () based on prune metric $(TP + TN)/(P + N)$.
Delete rules that are irrelevant from the rule set.
}

### 3.7. AdaBoost AD tree

In Adaboosting method [39], the alternating decision tree as shown in Fig. 5 grow iteratively by adding one base rule from each iterations corresponding to a sub-tree with a decision node as root. This new sub-tree is appended as a child to a predictor node. The advantage of this method is that a decision node can be added to any location of tree and splitting criteria is based on weighted error of added rules. The algorithm is described below:

1. Provide a training sample to the initial AD tree classifier.
2. Denote $r(x)$ as a set of base rules r that are associated with instance $x$.
3. Define the set of preconditions ($P_t$) and rules ($R_t$) on boosting iteration.

```
|   (1)degree_spondylolisthesis < 20.085: 0.575
|   |   (3)sacral_slope < 28.019: -0.956
|   |   (3)sacral_slope >= 28.019: 0.331
|   (1)degree_spondylolisthesis >= 20.085: -2.266
|   (2)pelvic_radius < 125.371: -0.367
|   |   (5)sacral_slope < 40.566: -0.537
|   |   |   (9)pelvic_tilt < 10.486: 0.461
|   |   |   (9)pelvic_tilt >= 10.486: -0.789
|   |   (5)sacral_slope >= 40.566: 0.433
|   |   |   (7)degree_spondylolisthesis < 9.334: 0.671
|   |   |   (7)degree_spondylolisthesis >= 9.334: -1.11
|   (2)pelvic_radius >= 125.371: 0.712
|   (4)degree_spondylolisthesis < 5.202: 0.279
|   |   (8)pelvic_radius < 117.436: -0.736
|   |   (8)pelvic_radius >= 117.436: 0.476
|   |   (10)lumbar_lordosis_angle < 31.531: -0.675
|   |   (10)lumbar_lordosis_angle >= 31.531: 0.496
|   (4)degree_spondylolisthesis >= 5.202: -0.747
|   (6)pelvic_tilt < 18.99: 0.406
|   (6)pelvic_tilt >= 18.99: -0.664
Legend: -ve = Abnormal, +ve = Normal
Tree size (total number of nodes): 31
Leaves (number of predictor nodes): 21
Weight: 3.24
```

**Fig. 5.** Shows an alternating decision tree with weights.

4. Each iteration associate a positive weight with each training instance.
5. Determine total weight $W$ of the training sample that satisfy the predicate $c$. Then label the instances of the test set by as either class based on $W + c$, $W - c$ respectively.

### 3.8. Real AdaBoost DTNB

The real AdaBoost method [39] exploits class probability estimates $P_m(x)$ to construct real valued contributions $f_m(x)$. Here the base classifier is decision table/Naive Bayes hybrid classifier (DTNB). The detailed method of hybridization is given below:

Initialize weights $W_i$ = 1/log $N$, $i$ = 1, 2, …, $N$.
For($i$ = 0; $i < M$; $i$+ +)
{
Adjust the DTNB classifier to obtain a class probability
  estimate $P_m(x) = P_w(y = 1|x)$ $\varepsilon$ [0, 1], using Wi.
Assign $f_m(x) = \frac{1}{2}$ log $P_m(x)/(1 - P_m(x))$
Assign Wi = Wi exp[$-y_i f_m(x_i)$], $i$ = 1, 2, …, $N$ and renormalize
  $\sum_{m=1}^{M} Wi = 1$.
}
Classifier sign: $\sum_{m=1}^{M} f_m(x)$

## 4. Theory/calculation

Communalities [40] reflect vital information regarding the amount of variance involved in each variable that is used in the analysis. The initial communalities are measures of the variance in each feature accounted for by all factors. The extraction communalities are measures of variance in each feature accounted for by other factors in the factor solution as shown in Table 4. Variables not fitting well in the factor solution should be deleted from the analysis. Raw communalities are computed on the basis of covariance and therefore they are on the same scale as the item variances. Rescaled communalities indicate the adjustment in proportions of variance for item variances and its values ranges 0 (completely unrelated)–1 (completely determined).

For a better analysis of the machine learning algorithms some of the metrics used in this work are described as follows. The abbreviations used to describe the metrics are as follows. TP is true positive cases, TN is true negative cases, FP is false positive cases and FN is false negative cases.

Accuracy is the percentage of correct classification of the test samples. $P$ and $N$ are positive and negative instances respectively, and is given by Eq. (2)

$$\text{Accuracy} = \frac{TP + TN}{P + N} \tag{2}$$

Area Under Curve (AUC) is the area calculated under ROC curve plotted using TP rate vs. FP rate and is given by the Eq. (3). Its value ranges from 0 to 1.

$$\text{AUC} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{3}$$

Cohen's Kappa Statistics ($K$) value is the probability measure of classification that occurs due to chance and ranges from −1 to 1. Higher the value of $K$ indicates greater confidence in classification and is measured by the Eq. (4)

$$K = (P_0 - P_c)/(1 - P_c) \tag{4}$$

where $P_0$ = Total agreement probability and $P_c$ = agreement probability due to chance.

Sensitivity (TPR) measures the ability to correctly classify healthy patients as positive and is calculated using the Eq. (5)

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

Specificity (TNR) measures the ability of the classifier to correctly classify ill patients and is given by Eq. (6)

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{6}$$

False Negative Rate (FNR) is the number of individuals identified as healthy, but actually they are abnormal. It is calculated by Eq. (7)

$$\text{FNR} = \frac{FN}{TP + FN} \tag{7}$$

Positive Predicted Value (PPV) is the number of individuals with positive test results and is diagnosed correctly and is given by Eq. (8)

$$\text{PPV} = \frac{FN}{TP + FN} \tag{8}$$

*F*-value measures the performance of a test which is the harmonic mean of recall and precision and is given by Eq. (9)

$$F = 2x\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

False Positive Rate (FPR) is defined as the number of individuals identified as abnormal, but the patient is healthy and is given by Eq. (10)

$$\text{FPR} = \frac{FP}{FP + TN} \tag{10}$$

Root Mean Square Error (RMSE) is the square root of the average squared error in the classification of the test dataset and is given by Eq. (11)

$$\text{RMSE} = \sqrt[2]{\frac{1}{N}\sum_{J=1}^{N}(Ei - Ei')^2} \tag{11}$$

The computation of reliability statistics alpha is based on the number of members in the analysis ($K$) and the ratio of average inter-item covariance to average member variance. It is mathematically defined as follows:

$$\alpha = \frac{K}{K - 1}\left(1 - \frac{\sum_{i=1}^{K}\sigma_{Y_i}^2}{\sigma_X^2}\right) \tag{12}$$

where $K$ = number of members, $\sigma_X^2$ = observed variance and $\sigma_{Y_i}^2$ = variance of component $i$ for each member. The obtained value of alpha = 0.710 suggests that the analysis performed on the radiographs pelvic spine data is acceptable and when it is extended for even larger data sets, the results will hold true or may vary to a negligible extent. The research work carried out signifies the important information about the validity of the work.

In this paper, the proposed method of visual distinction using Kernel Density Estimator (KDE) [6] of bivariate random vector for the pelvic spine diagnosis integrated with the inference system is presented (as shown in Fig. 7). Here KDE is a metric for comparing the severity of illness based on the correlation coefficient between the features of the illness. If the correlation coefficient (cc, ranging between −1 and 1) is higher in magnitude then the severity of illness is very high else it's low. The nature of plot shows the difference between the healthy as well as abnormal patients. The inference system evaluates the newly arrived patient data using the following method. The input variables are the measured parameters from the radiographs features obtained from the pelvic spine of the patient. Then the machine learning (ML) algorithms are run over the input data and the output from each ML model are obtained along with a correlation coefficient (cc). If the cc $\geqslant$ 0.9 then the abnormality is higher else it's lower. Since the

|  | Component 1 (Abnormal) | Component 1 (Normal) |
|---|---|---|
| pelvic_incidence | .118 | .406 |
| pelvic_tilt | .021 | .064 |
| lumbar_lordosis_angle | .123 | .366 |
| sacral_slope | .068 | .225 |
| pelvic_radius | .011 | -.117 |
| degree_spondylolisthesis | .796 | .028 |
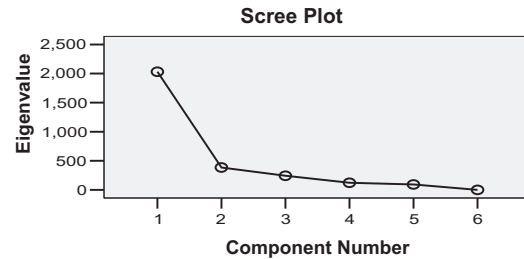


**Fig. 6.** The component score coefficient matrix and scree plot for class abnormal and normal category extracted using Principal Component Analysis with varimax rotation with Kaiser Normalization.
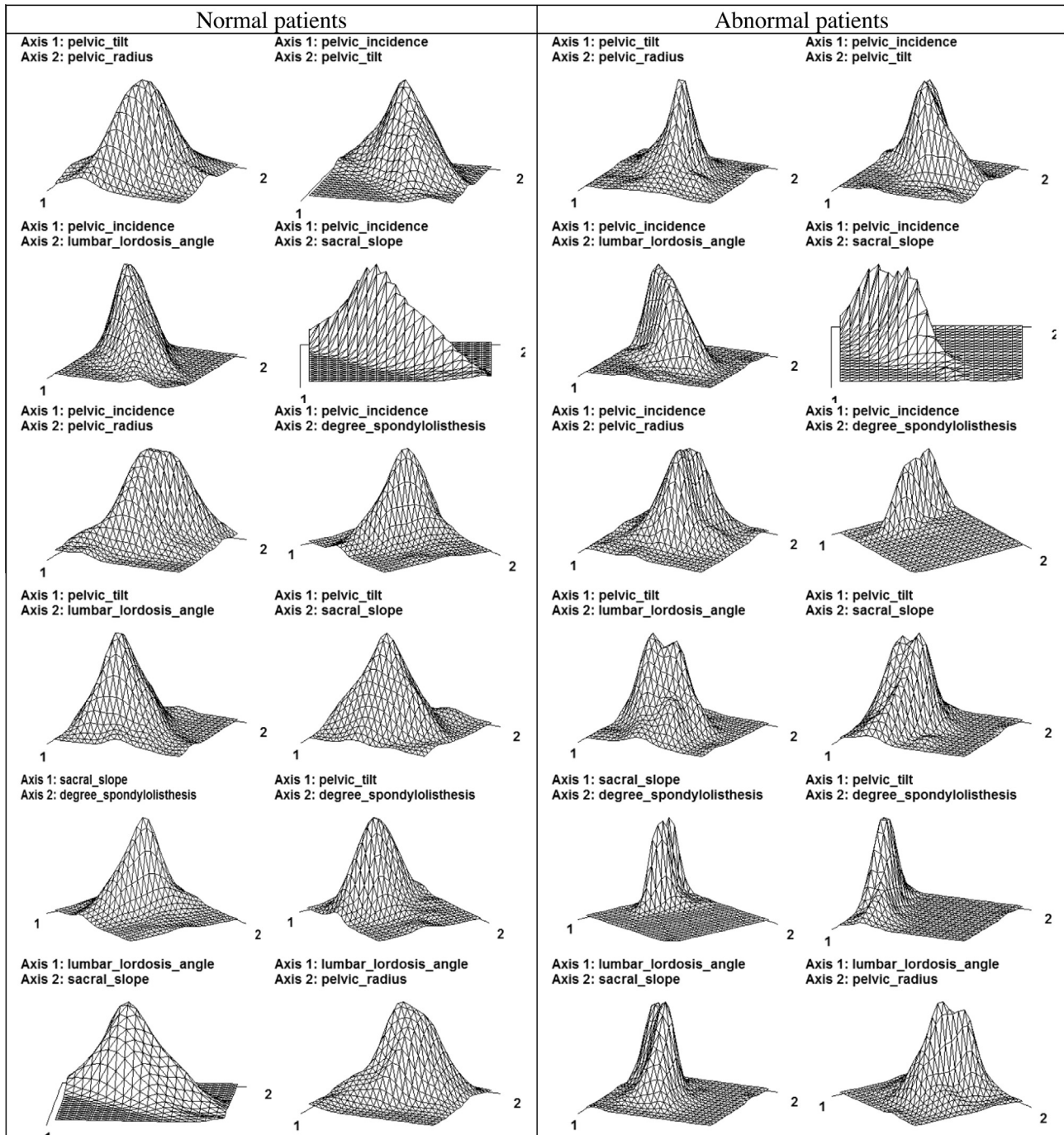


**Fig. 7.** The various characteristics curves between normal and abnormal patients computed on Kernel density estimation of bivariate random vector after preprocessing and standardization of the vertebral column dataset. It clearly indicates difference between the two classes of patients.

ML models are designed in parallel and independently. The estimation of reliability as discussed in [1] of the decision from the inference system is computed using the Eq. (13)

$$R_s = 1 - \prod_{i=1}^{n}(P/Q) * P(E) \qquad (13)$$

where $P$ = # failures occurred in $n$ runs. $Q$ = #runs sampled in input sub domains $E$. $P(E)$ = probability of data $E$ is accessed in the operational environment of inference system.

The proposed inference system consists of highly reliable machine learning algorithms for accurate diagnosis of Disc hernia and Spondylolisthesis. The system is trained with the past clinical records of patients having the spine disorders. The machine learning algorithms are trained over the clinical data after preprocessing of raw data and their optimization along with fine tuning is achieved after repeated experimentations. Once the system is learnt, then it is ready for diagnosis. The final decision is provided by a Gibbs classifier (GC) that takes the output of all machine learning algorithms. The GC draws w from distribution **q(w)** and produces output based on the rule **ȳ = sign(wᵀx) ≡ f_GC(x, w)**. The probability that the final output is a correct decision computed using Bayesian model given by

$$P(X) = \sum_n P(X \cap Yn) = \sum_n P(X|Yn)P(Yn) \qquad (14)$$

where $X$ is the correct decision from GC classifier and $Y_n$ is the correct decision from machine learning algorithms in the inference system. The radiographs data are repeatedly run and tested the inference system for 100 times by bootstrapping the dataset and find that the average accuracy of the inference system achieved is 98.9999%.

## 5. Experiments and results

Sensitivity reflects the ratio of abnormal patients who are correctly identified. Specificity gives the ratio of identifying a healthy patient as healthy. Therefore the target is to design an inference system with a maximized degree of specificity and sensitivity. In this work corrected $t$-test is employed because it provides better results than $t$-tests as it accounts for the reduction of dependency

of involved variables. The experiments are conducted using mathematical models and new machine learning ensemble algorithms after optimization of the metrics involved for correctly identifying disc hernia and spondylolisthesis patients at early stage based on the radiograph measured parameters with confidence level of 95% and 99% as presented in Tables 9 and 10.

The component score coefficient matrix in rotated space using Principal Component Analysis (PCA) using varimax with Kaiser Normalization and scree plot with Eigen value in Y-axis and component number in the X-axis for pelvic spine data classes is shown in Fig. 6. The result reflects the coordinates of the variables plotted on each dimension. Table 6 shows the component matrix of the factor loadings for each variable on the unrotated factors for abnormal and normal class obtained from factor analysis based on covariance of the radiograph data. The Kendall's W coefficient of concordance is a non parametric statistic that gives mean ranks between the involved parameters and the normality test values are presented in Table 3.

The vital information extracted from the factor analysis is the nature of the behavior of radiograph features regarding their dependency and independence among each other. The newly generated rotation axes give highly vital information about the features. While working with the machine learning ensembles for feature selection, the curse of dimensionality is given due care to be avoided, that is, too less input feature variables shall potentially lead to a simpler model increasing predictive accuracy. But the modeling of real life complex phenomena or problems cannot be mapped using simpler models as it is more prone to fragile system or model.

For each factor solution Eigen values, variance explained and cumulative variance explained are shown in Table 5. The first column shows individual values depending on initial Eigen values. In case of the initial factor solution, the number of factors equals the number of variables. 'Total" shows the degree of variance in the observed variables for factor. "% of Variance" indicates the percent of variance by each specific component, with respect to the total variance. "Cumulative%" shows the percent of variance by all components added up to as well as the current one. There will be a few factors that indicate the lion's share of variance and remaining factors indicate little variance in a good factor analysis. Sum of squared loadings provide information about extracted com-

**Table 3**
Shows the results from Normality tests and Kendall's Concordance mean ranking of the features. (All computations are performed with Spearman's Rho is 0.98152 and p < 0.001).

| | Shapiro–Wilk | Lilliefors D | Anderson–Darling | d'Agostino | Kendall's W Concordance = 0.82146 | Brown Forsythe's | Fisher's Test | Bartlett's's Test | Levene's Test |
|---|---|---|---|---|---|---|---|---|---|
| pelvic_incidence | 0.971123 | 0.0708 | 1.945889 | 13.7549 | 4.7032 | 18.0798 | 2.0393 | 15.386 | 17.659 |
| pelvic_tilt | 0.966386 | 0.0829 | 3.424959 | 24.8099 | 1.5774 | 15.1731 | 2.4067 | 22.764 | 15.850 |
| lumbr_lrdsis_angle | 0.971812 | 0.0687 | 2.360988 | 17.1629 | 3.871 | 27.2899 | 2.5319 | 25.261 | 26.925 |
| sacral_slope | 0.96398 | 0.0578 | 0.947005 | 53.2972 | 3.0452 | 15.5191 | 2.2749 | 20.119 | 15.102 |
| pelvic_radius | 0.988722 | 0.0593 | 1.232183 | 0.00122 | 5.971 | 10.7979 | 2.4434 | 23.499 | 10.954 |
| degree_spondylolisthesis | 0.696977 | 0.1726 | 17.83157 | 326.926 | 1.8323 | 45.3901 | 41.6301 | 252.22 | 50.079 |

**Table 4**
Shows the communalities of each feature obtained in the factor analysis for class Abnormal and normal.

| Communalities (a) | Class = Abnormal | | | | Class = Normal | | | |
|---|---|---|---|---|---|---|---|---|
| Attributes | Raw | | Rescaled | | Raw | | Rescaled | |
| | Initial | Extraction | Initial | Extraction | Initial | Extraction | Initial | Extraction |
| pelvic_incidence | 311.951 | 184.754 | 1.000 | .592 | 152.971 | 139.632 | 1.000 | .913 |
| pelvic_tilt | 110.584 | 16.884 | 1.000 | .153 | 45.948 | 11.633 | 1.000 | .253 |
| lumbar_lordosis_angle | 386.888 | 160.499 | 1.000 | .415 | 152.804 | 113.383 | 1.000 | .742 |
| sacral_slope | 210.701 | 89.935 | 1.000 | .427 | 92.621 | 70.659 | 1.000 | .763 |
| pelvic_radius | 198.545 | 2.498 | 1.000 | .013 | 81.257 | 21.760 | 1.000 | .268 |
| degree_spondylolisthesis | 1656.225 | 1576.984 | 1.000 | .952 | 39.784 | 2.533 | 1.000 | .064 |

**Table 5**
Presents eigen values, variance explained, and cumulative variance explained for factor solution.

| Component | Total Variance Explained[b] | | | | | |
|---|---|---|---|---|---|---|
| | Initial Eigen values[a] | | | Extraction Sums of Squared Loadings | | |
| | Total | % of Variance | Cumulative% | Total | % of Variance | Cumulative% |
| *Raw* | | | | | | |
| 1 | 2031.554 | 70.665 | 70.665 | 2031.554 | 70.665 | 70.665 |
| 2 | 385.165 | 13.398 | 84.063 | | | |
| 3 | 242.926 | 8.450 | 92.513 | | | |
| 4 | 122.143 | 4.249 | 96.761 | | | |
| 5 | 93.105 | 3.239 | 100.000 | | | |
| 6 | 3.66E−014 | 1.27E−015 | 100.000 | | | |
| *Rescaled* | | | | | | |
| 1 | 2031.554 | 70.665 | 70.665 | 2.551 | 42.523 | 42.523 |
| 2 | 385.165 | 13.398 | 84.063 | | | |
| 3 | 242.926 | 8.450 | 92.513 | | | |
| 4 | 122.143 | 4.249 | 96.761 | | | |
| 5 | 93.105 | 3.239 | 100.000 | | | |
| 6 | 3.66E−014 | 1.27E−015 | 100.000 | | | |
| *Raw* | | | | | | |
| 1 | 359.602 | 63.603 | 63.603 | 359.602 | 63.603 | 63.603 |
| 2 | 80.338 | 14.209 | 77.812 | | | |
| 3 | 51.435 | 9.097 | 86.910 | | | |
| 4 | 41.085 | 7.267 | 94.176 | | | |
| 5 | 32.927 | 5.824 | 100.000 | | | |
| 6 | 1.29E−013 | 2.29E−014 | 100.000 | | | |
| *Rescaled* | | | | | | |
| 1 | 359.602 | 63.603 | 63.603 | 3.002 | 50.039 | 50.039 |
| 2 | 80.338 | 14.209 | 77.812 | | | |
| 3 | 51.435 | 9.097 | 86.910 | | | |
| 4 | 41.085 | 7.267 | 94.176 | | | |
| 5 | 32.927 | 5.824 | 100.000 | | | |
| 6 | 1.29E−013 | 2.29E−014 | 100.000 | | | |

Extraction Method: Principal Component Analysis.
[a] When analyzing a covariance matrix, the initial eigen values are the same across the raw and rescaled solution.
[b] Only cases for which class = Normal are used in the analysis phase.

ponents. The values are equal that are reported under Initial Eigen-values for principal components, whereas in other extraction methods, the values may be lesser than initial values for errors in measurements. The rescaling affects extraction and rotation values so raw and rescaled values for initial Eigen values remains unchanged. All the rescaled values are standardized for accounting variances in initial Eigen values.

Table 6 shows the factor loadings for each variable on the unrotated components or factors. Each value represents the correlation between item and unrotated factor. These correlations give an

**Table 6**
Shows the component matrix of the factor loadings for each variable on the unrotated factors for abnormal and normal class obtained from factor analysis.

| | Raw Component 1 | Rescaled Component 1 |
|---|---|---|
| *Class = Abnormal* | | |
| pelvic_incidence | 13.592 | .770 |
| pelvic_tilt | 4.109 | .391 |
| lumbar_lordosis_angle | 12.669 | .644 |
| sacral_slope | 9.483 | .653 |
| pelvic_radius | 1.581 | .112 |
| degree_spondylolisthesis | 39.711 | .976 |
| *Class = Normal* | | |
| pelvic_incidence | 11.817 | .955 |
| pelvic_tilt | 3.411 | .503 |
| lumbar_lordosis_angle | 10.648 | .861 |
| sacral_slope | 8.406 | .873 |
| pelvic_radius | −4.665 | −.517 |
| degree_spondylolisthesis | 1.592 | .252 |

important interpretation of the factors. It's done by searching for a common thread among the features that have large loadings for a specific factor. The raw values are in the scale of the original variables. The rescaled values are standardized such that all values are measured on the same scale. It is observed that items with large loadings values of the unrotated factors, which can make interpretation difficult, therefore in such a case, it is worth examining a rotated solution.

Table 7 gives reproduced covariance and residual for the factor analysis solution. The solution is a good one, if the reproduced covariance will be close to the observed values. Residuals show the difference between the predicted and observed values. For a good factor analysis solution, most of these values will be small. The scree plot as shown in Fig. 6 shows the Eigen values for initial components. It is used to help determine the optimal number of factors or components to retain in the solution. For a good factor analysis, this chart will look roughly like the intersection of two lines. Generally, the factors to be kept are the ones on the steep slope. The ones on the shallow slope contribute relatively little to the solution, and can be excluded. The factor score coefficient matrix is shown in Fig. 6 values used to compute factor scores for each case. For each case, the factor score is computed by multiplying variable values by factor score coefficients. For principal component models, these give exact component scores. For other extraction methods, exact scores cannot be computed. The coefficients are used to generate estimates of the true factor scores. M-estimators as shown in Table 8 are robust measures of central tendency as they are not affected by deviations from normality. Vital information regarding the magnitudes of each variable is obtained by observing the magnitude of values.

**Table 7**
The reproduced covariance and residuals for the classes abnormal and normal. Residual is computed between observed and reproduced covariance.

| | | pelvic_incidence | pelvic_tilt | lumbar_lordosis_angle | sacral_slope | pelvic_radius | degree_spondylolisthesis |
|---|---|---|---|---|---|---|---|
| *Class = Abnormal* | | | | | | | |
| Reproduced Covariance | pelvic_incidence | 184.754(b) | 55.851 | 172.200 | 128.902 | 21.484 | 539.772 |
| | pelvic_tilt | 55.851 | 16.884(b) | 52.056 | 38.967 | 6.495 | 163.173 |
| | lumbar_lordosis_angle | 172.200 | 52.056 | 160.499(b) | 120.144 | 20.024 | 503.094 |
| | sacral_slope | 128.902 | 38.967 | 120.144 | 89.935(b) | 14.989 | 376.598 |
| | pelvic_radius | 21.484 | 6.495 | 20.024 | 14.989 | 2.498(b) | 62.767 |
| | degree_spondylolisthesis | 539.772 | 163.173 | 503.094 | 376.598 | 62.767 | 1576.984(b) |
| Residual(a) | pelvic_incidence | | 50.065 | 63.967 | 77.132 | −42.419 | −85.856 |
| | pelvic_tilt | 50.065 | | 26.770 | −43.634 | 29.883 | −26.141 |
| | lumbar_lordosis_angle | 63.967 | 26.770 | | 37.197 | −.947 | −105.734 |
| | sacral_slope | 77.132 | −43.634 | 37.197 | | −72.302 | −59.715 |
| | pelvic_radius | −42.419 | 29.883 | −.947 | −72.302 | | 21.192 |
| | degree_spondylolisthesis | −85.856 | −26.141 | −105.734 | −59.715 | 21.192 | |
| *Class = Normal* | | | | | | | |
| Reproduced Covariance | pelvic_incidence | 139.632(b) | 40.303 | 125.825 | 99.329 | −55.122 | 18.808 |
| | pelvic_tilt | 40.303 | 11.633(b) | 36.318 | 28.670 | −15.910 | 5.429 |
| | lumbar_lordosis_angle | 125.825 | 36.318 | 113.383(b) | 89.507 | −49.672 | 16.948 |
| | sacral_slope | 99.329 | 28.670 | 89.507 | 70.659(b) | −39.212 | 13.379 |
| | pelvic_radius | −55.122 | −15.910 | −49.672 | −39.212 | 21.760(b) | −7.425 |
| | degree_spondylolisthesis | 18.808 | 5.429 | 16.948 | 13.379 | −7.425 | 2.533(b) |
| Residual(a) | pelvic_incidence | | 12.846 | −18.630 | .493 | .458 | −3.186 |
| | pelvic_tilt | 12.846 | | −10.956 | −21.469 | −5.267 | 2.345 |
| | lumbar_lordosis_angle | −18.630 | −10.956 | | −7.673 | 22.108 | 3.386 |
| | sacral_slope | .493 | −21.469 | −7.673 | | 5.725 | −5.531 |
| | pelvic_radius | .458 | −5.267 | 22.108 | 5.725 | | 4.117 |
| | degree_spondylolisthesis | −3.186 | 2.345 | 3.386 | −5.531 | 4.117 | |

(b) Reproduced communalities.

**Table 8**
The robust measures of central tendency for all the variables in the analysis of the vertebral column diagnosis.

| | Huber's M-Estimator(a) | Tukey's Biweight(b) | Hampel's M-Estimator(c) | Andrews' Wave(d) |
|---|---|---|---|---|
| pelvic_incidence | 59.40327717 | 59.27306435 | 59.70228994 | 59.26788915 |
| pelvic_tilt | 16.46980798 | 15.88088155 | 16.46242944 | 15.85723831 |
| lumbar_lordosis_angle | 50.12615486 | 49.75186267 | 50.47460421 | 49.75498957 |
| sacral_slope | 42.41765039 | 42.25285693 | 42.41770038 | 42.24425421 |
| pelvic_radius | 118.33056286 | 118.45181729 | 118.30428030 | 118.45296439 |
| degree_spondylolisthesis | 17.37936992 | 14.11153427 | 17.84915911 | 14.01182345 |

The corresponding weighting constant: a = 1.339, b = 4.685, c = 1.700, 3.400, 8.500, d = 1.340∗pi.

The reason for achieving statistically improved results is mainly due to preprocessing stage (data mining) of the raw clinical data, relevant feature selection method, outlier's identification and removal, development of new machine learning ensemble algorithms and design of robust inference system that included statistical quality control evaluation scheme. Also the new ensemble algorithms are optimized based on the metrics involved in achieving higher classification accuracy. The paper proposes the design of a reliable inference system that consists of a collection of tested robust algorithms machine learning algorithms to reduce the probability of overall misclassifications. The final decision of the inference system is not dependent on one model instead powerful and statistically heuristic learning models are employed. The kernel density estimator graph compared to the correlation coefficient provides vital information about the patient's health condition. As per our knowledge, this method is used first time for pathologic diagnosis.

In this research study, the best prediction accuracy achieved is 96.55% for AdaBoost and improved RIPPER algorithm. The corrected *t*-test is employed as it provides statistically confident results than *t*-tests due to reduction of dependency among variables. The experiments are conducted based on several reliable

**Table 9**
The experimental results of the machine learning models with 95% confidence level after preprocessing and standardization of the data set. All models are trained with 80% of dataset and tested over rest of it.

| | Sensitivity-TP | FP | F value | ROC | Kappa | RMSE | PPV Precision | False negative | Specificity TN rate | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| RF-Random Tree-PCA Ensemble | 0.948 | 0.137 | 0.947 | 0.982 | 0.837 | 0.217 | 0.947 | 0.052 | 0.863 | 94.82 |
| **RF-decision table-wavelets Ensemble** | 0.914 | 0.207 | 0.912 | 0.966 | 0.729 | 0.273 | 0.912 | 0.086 | 0.793 | 91.38 |
| Bagging-C4.5 | 0.948 | 0.013 | 0.95 | 0.98 | 0.856 | 0.218 | 0.959 | 0.052 | 0.987 | 94.83 |
| **Improved NNge** | 0.9310 | 0.08 | 0.933 | 0.926 | 0.802 | 0.262 | 0.938 | 0.069 | 0.920 | 93.10 |
| Improved RIPPER | 0.966 | 0.132 | 0.964 | 0.917 | 0.888 | 0.190 | 0.967 | 0.034 | 0.868 | 96.55 |
| **AdaBoostM1 AD Tree** | 0.966 | 0.071 | 0.966 | 0.998 | 0.895 | 0.141 | 0.966 | 0.034 | 0.929 | 96.55 |
| Real AdaBoost DTNB | 0.948 | 0.137 | 0.947 | 0.966 | 0.837 | 0.244 | 0.947 | 0.052 | 0.863 | 94.82 |

**Table 10**
The experimental results of the machine learning models with 99% confidence level after preprocessing and standardization of the data set with corrected T-test with respect to RF-Random Tree-PCA classifier. All models are trained with 80% of dataset and tested over rest of it.

| | Sensitivity TP | FP | F value | ROC | Kappa | RMSE | PPV Precision | False negative | Specificity = TN | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Random Tree- PCA Ensemble | 0.98 | 0.17 | 0.97 | 0.98 | 0.84 | 0.22 | 0.96 | 0.02 | 0.83 | 94.83 |
| **Decision table-wavelets Ensemble** | 0.93 | 0.17 | 0.95 | 0.96 | 0.75 | 0.28 | 0.96 | 0.07 | 0.83 | 91.38 |
| Bagging-C4.5 | 0.93 | 0.00 | 0.97 | 0.98 | 0.86 | 0.22 | 1.00 | 0.07 | 1.00 | 94.83 |
| **Improved NNge** | 0.93 | 0.08 | 0.96 | 0.93 | 0.80 | 0.26 | 0.98 | 0.07 | 0.92 | 93.10 |
| Improved RIPPER | 1 | 0.17 | 0.98 | 0.92 | 0.89 | 0.19 | 0.96 | 0.00 | 0.83 | 96.55 |
| **AdaBoostM1 AD Tree** | 0.98 | 0.08 | 0.98 | 1.00 | 0.89 | 0.14 | 0.98 | 0.02 | 0.92 | 96.55 |
| Real AdaBoost DTNB | 0.98 | 0.17 | 0.97 | 0.97 | 0.84 | 0.24 | 0.96 | 0.02 | 0.83 | 94.83 |
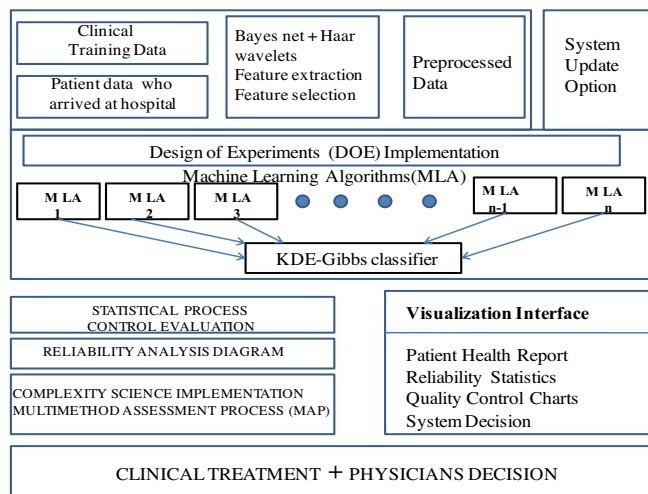


**Fig. 8.** The proposed model for pelvic spine diagnosis inference model that includes Statistical Process Control evaluation (SPC) integrated with machine learning algorithms and Multimethod Assessment Process (MAP).

machine learning algorithms developed for distinguishing spondylolisthesis affected patients against healthy subjects based on non invasive radiograph measured parameters. Also the results obtained from corrected *t*-test over the new ensembles with

respect to the Rotation forest ensemble prove the robustness and reliability of the models for predictive accuracy as shown in Tables 9 and 10.

In the proposed inference system as shown in Fig. 8, Design of Experiment (DOE) [6] is implemented to discover the optimized values of each parameter that affect the classification accuracy of the machine learning algorithms (MLA). The fractional factorial design technique is computationally applied to get the optimum values. The architecture presents an independent interaction of the machine learning algorithms and therefore parallel computation is carried out by feeding the preprocessed clinical data to each MLA. The training data are the clinical data acquired during medical examination of patients and is undergone through the data cleaning process. Feature selection and extraction is performed for identifying the minimum and sufficient number of attributes. The Reliability diagram as shown in Fig. 9 provides graphical visualization and evaluation of probabilistic forecasts. The mean score of the group and the observed relative frequency of positive class value are plotted on the $X$ and $Y$ axis respectively.

Statistical process control [41] technique is a valuable applied statistical method to monitor and control a process that operates to produce conforming results. Here we have used control charts that results in continuous quality improvement as shown in Fig. 10. In primary health care service quality improvement processes have mostly met with limited success due to its inherent complexity. In such a situation, complexity science provides better edge to understand primary care services by using multimethod
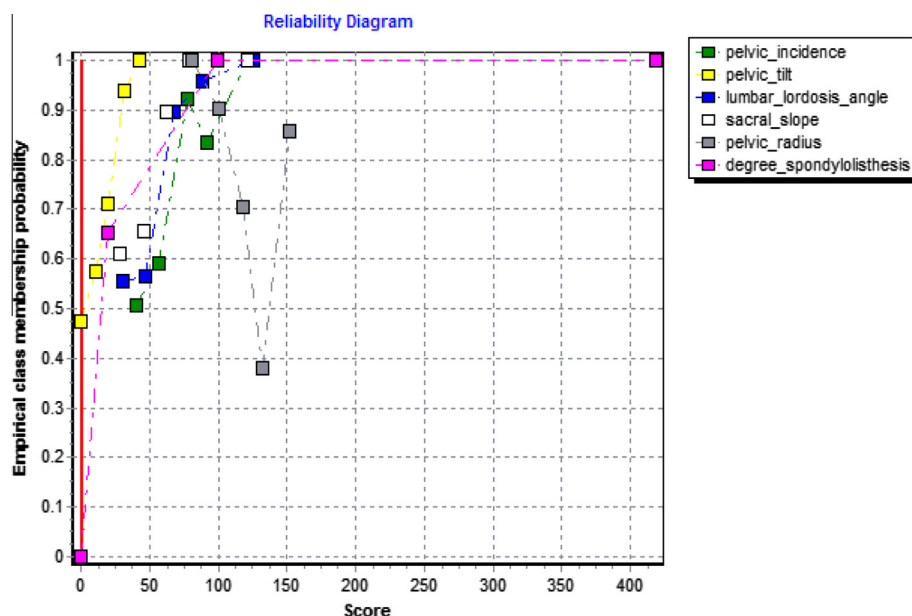


**Fig. 9.** The reliability diagram of scores of each vertebral radiograph features on X-axis to empirical class membership probability of abnormal patients.

**Control Chart: pelvic_incidence**
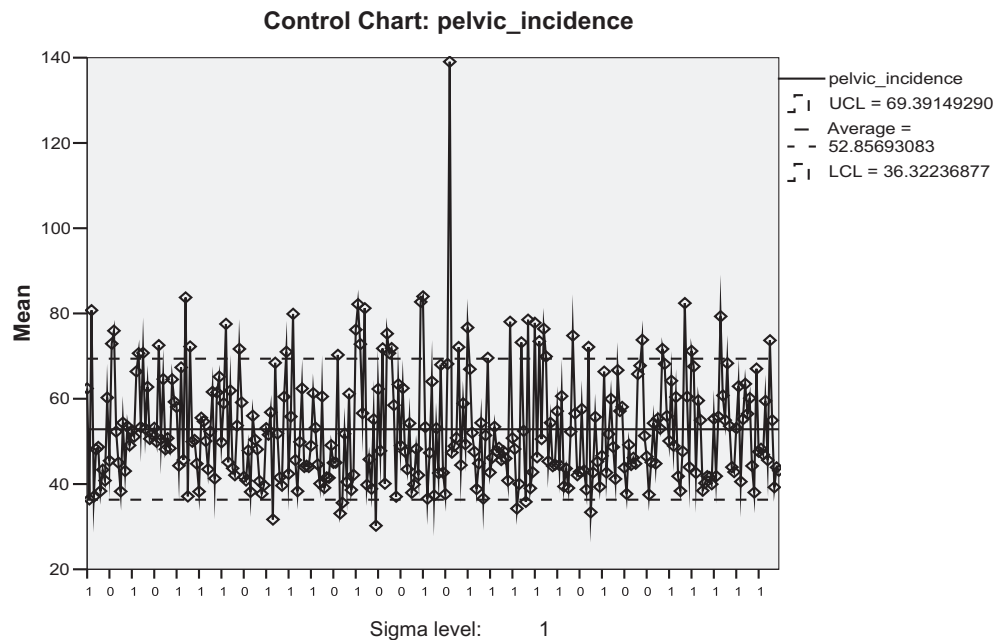


**Fig. 10.** The statistical quality control charts for the class with respect to the mean. The upper control limit (UCL) and lower control limit (LCL) and the average is shown for the pelvic_incidence feature.

assessment process (MAP) [42]. Its principles include in-depth understanding vision and mission, learning and reflections that help organizations foster adaptability and insights for constructive change. Thus, complex adaptive systems (CAS) [43] demand technological innovations like electronic clinical records and study interactions among different components. Continuous quality improvement stimulates sustained organizational change that uses interdisciplinary teams to achieve measurable outcomes. The main reason for limited success is that existing approaches are based on enhancing the quality of individual factors rather than complex interactions of many interrelated parts. CAS dynamically evolves over time and exhibit self organization, emergence and co-evolution. Since the relationships are dynamic and non-linear, therefore, more attention is paid to the relationship between agents than to the quality of the individual agents. The concern is more for the development of agent's learning capabilities, focus on co-evolution and encourage diversity among agents. MAP is a change process reflected from complexity science for guiding and informing its methods to foster sense-making, learning, improvisation and adaptation towards an impact of the change. It provides qualitative and quantitative data collection using multi methods such as observation, collection of practice documents, patients' survey and pathways, review, key-informant and in-depth interviews, clinician and staff questionnaires and practice genograms. Reflective Adaptive Process is initiated by the outcome of MAP to identify priority improvement opportunities and potential solutions. The system update option enables the user to update the system from time to time and the new data resources can be added to the database.

The major findings are that the ensemble Rotation forest consisting of Random tree as a base classifier and PCA as a projection filter transformation that leads to improvement in the classification and predictive analysis of the inference system. The most appropriate machine learning ensemble methods and models that could be used to advance the existing technology for spine diagnosis is presented. The integration of statistical quality control evaluation with the inference system and the experimental results obtained in this work provides a new dimension towards quality spine diagnosis. The superior performance of proposed algorithms

developed in this research work exhibits excellent results as compared to the current works accomplished by other researchers.

## 6. Discussion and conclusions

The major contributions in this work are towards development of new hybrid machine learning ensembles, computation of software reliability and quality associated with medical diagnosis systems, integrating statistical quality control evaluation strategy to have accurate and robust decisions useful for assisting physicians. The discovery of statistical relationships among the features of pelvic spine using Principal component analysis and the design of a reliable inference system for making fast decisions and with greater degree of certainty to improve the spine treatment. Also, another major contribution lies in the relevant feature selection based on an ensemble of machine learning algorithms. Here MAP is implemented as a practical technique using complexity science principles as a theoretical platform for informing practice improvement and further it could be easily adapted to any organizations other than primary health care.

The primary aim of the work is to upgrade the software reliability of the spine diagnosis and reduce the rate of misdiagnosis and the obtained test and experimental results indicate that the goal is reached. But still there is lots of room at the bottom in this direction as the diagnosis is possible by several means. Also, these machine learning models need to be tested over large clinical data sets. To provide quality treatment, reduce costs and prevent misdiagnosis of patients, leading to prolonged suffering are the prime motivation of medical diagnostic systems. The work described here provides new direction towards telemonitoring applications. Recognition of disease in a patient accurately is a great challenge in medical field. Huge amount of capital investment is spent for research on medical equipments and advancing primary health care devices as they are largely considered as critical systems.

Also the use of such medical system solves the problem of availability of clinical data in the electronic form for further research as against the use of paper based systems. Some of the major

problems confronted while developing an expert diagnosis system such as medical experts are less interested to share acquired skill knowledge with others. The knowledge gained over experience (called common sense) is practically impossible to be separated. And designing an ideal expert system for medical application is difficult as each system is suitable for a specific health disorders due to the unique behavior of algorithms in solving specific problems. The similar diagnostic system can be developed for other chronic diseases.

## References

[1] I. Mandal, N. Sairam, Accurate prediction of coronary artery disease using reliable diagnosis system, J. Med. Syst. 36 (2012) 3353–3373, http://dx.doi.org/10.1007/s10916-012-9828-0.

[2] D. Pal, K.M. Mandana, S. Pal, D. Sarkar, C. Chakraborty, Fuzzy expert system approach for coronary artery disease screening using clinical parameters, Knowl.-Based Syst. 36 (2012) 162–174, http://dx.doi.org/10.1016/j.knosys.2012.06.013.

[3] I. Mandal, N. Sairam, New machine-learning algorithms for prediction of Parkinson's disease, Int. J. Syst. Sci. 45 (3) (2014) 647–666, http://dx.doi.org/10.1080/00207721.2012.724114.

[4] S.-Y. Hung, Y.-C. Ku, J.-C. Chien, Understanding physicians' acceptance of the Medline system for practicing evidence-based medicine: A decomposed TPB model, Int. J. Med. Informatics 81 (2) (2012) 130–142, http://dx.doi.org/10.1016/j.ijmedinf.2011.09.009.

[5] S.L. Ting, S.K. Kwok, A.H.C. Tsang, W.B. Lee, A hybrid knowledge-based approach to supporting the medical prescription for general practitioners: Real case in a Hong Kong medical center, Knowl.-Based Syst. 24 (3) (2011) 444–456, http://dx.doi.org/10.1016/j.knosys.2010.12.011.

[6] I. Mandal, N. Sairam, Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system, Int. J. Med. Informatics 82 (5) (2013) 359–377, http://dx.doi.org/10.1016/j.ijmedinf.2012.10.006.

[7] Z. Jiang, S. Zhang, J. Zeng, A hybrid generative/discriminative method for semi-supervised classification, Knowl.-Based Syst. 37 (2013) 137–145, http://dx.doi.org/10.1016/j.knosys.2012.07.020.

[8] D.G. Anderson, W. Limthongkul, A. Sayadipour, C.K. Kepler, J.S. Harrop, M. Maltenfort, A.R. Vaccaro, A. Hilibrand, J.A. Rihn, T.J. Albert, A radiographic analysis of degenerative spondylolisthesis at the L4-5 level: clinical article, J. Neurosurgery: Spine 16 (2) (2012) 130–134, http://dx.doi.org/10.3171/2011.10.SPINE11140.

[9] A.A. Edidin, K.L. Ong, E. Lau, J.K. Schmier, J.E. Kemner, S.M. Kurtz, Cost-effectiveness analysis of treatments for vertebral compression fractures, Appl. Health Econom. Health Policy 10 (4) (2012) 273–284, http://dx.doi.org/10.2165/11633220-000000000-00000.

[10] S. Kolta, S. Kerkeni, C. Travert, W. Skalli, R. Eastell, C.C. Glüer, C. Roux, Variations in vertebral body dimensions in women measured by 3D-XA: a longitudinal in vivo study, Bone 50 (3) (2012) 777–783, http://dx.doi.org/10.1016/j.bone.2011.12.005.

[11] Y. Masharawi, Lumbar shape characterization of the neural arch and vertebral body in spondylolysis: a comparative skeletal study, Clin. Anat. 25 (2) (2012) 224–230, http://dx.doi.org/10.1002/ca.21203.

[12] L.A. Szeyko, M.S. Taljanovic, R.B. Dzioba, J.L. Rapiejko, R.D. Adam, Vertebral coccidioidomycosis: presentation and multidisciplinary management, Am. J. Med. 125 (3) (2012) 304–314, http://dx.doi.org/10.1016/j.amjmed.2011.10.018.

[13] E. Dall'Ara, D. Pahr, P. Varga, F. Kainberger, P. Zysset, QCT-based finite element models predict human vertebral strength in vitro significantly better than simulated DEXA, Osteoporos. Int. 23 (2) (2012) 563–572, http://dx.doi.org/10.1007/s00198-011-1568-3.

[14] C.-D. Yang, Y.-W. Chen, C.-S. Tseng, H.-J. Ho, C.-C. Wu, K.-W. Wang, Non-invasive, fluoroscopy-based, image-guided surgery reduces radiation exposure for vertebral compression fractures: a preliminary survey, Formosan J. Surgery 45 (1) (2012) 12–19, http://dx.doi.org/10.1016/j.fjs.2011.12.003.

[15] E. Sapin-De Brosses, E. Jolivet, C. Travert, D. Mitton, W. Skalli, Prediction of the vertebral strength using a finite element model derived from low-dose biplanar imaging: benefits of subject-specific material properties, Spine 37 (3) (2012) E156–E162, http://dx.doi.org/10.1097/BRS.0b013e3182293628.

[16] M.G. Roberts, T. Oh, E.M.B. Pacheco, R. Mohankumar, T.F. Cootes, J.E. Adams, Semi-automatic determination of detailed vertebral shape from lumbar radiographs using active appearance models, Osteoporos. Int. 23 (2) (2012) 655–664, http://dx.doi.org/10.1007/s00198-011-1604-3.

[17] D.C. Oh, J.A. Hirsch, A.J. Yoo, Novel use of Onyx for treatment of intracranial vertebral artery dissection, J. NeuroInterventional Surgery 4 (1) (2012) 31–33, http://dx.doi.org/10.1136/jnis.2011.004952.

[18] R. Wustrack, E. Seeman, C. Bucci-Rechtweg, S. Burch, L. Palermo, D.M. Black, Predictors of new and severe vertebral fractures: results from the HORIZON pivotal fracture trial, Osteoporos. Int. 23 (1) (2012) 53–58, http://dx.doi.org/10.1007/s00198-011-1664-4.

[19] C. Roux, J.L. Goldstein, X. Zhou, A. Klemes, R. Lindsay, Vertebral fracture efficacy during risedronate therapy in patients using proton pump inhibitors,

[20] Osteoporos. Int. 23 (1) (2012) 277–284, http://dx.doi.org/10.1007/s00198-011-1574-5.

[20] T. Baum, S.P. Yap, D.C. Karampinos, L. Nardo, D. Kuo, A.J. Burghardt, U.B. Masharani, A.V. Schwartz, X. Li, T.M. Link, Does vertebral bone marrow fat content correlate with abdominal adipose tissue, lumbar spine bone mineral density, and blood biomarkers in women with type 2 diabetes mellitus?, J. Magn. Reson. Imaging 35 (1) (2012) 117–124, http://dx.doi.org/10.1002/jmri.22757.

[21] A.R. Da Rocha Neto, G.D.A. Barreto, On the application of ensembles of classifiers to the diagnosis of pathologies of the vertebral column: a comparative analysis, IEEE Latin Am. Trans. 7 (4) (2009) 487–496, http://dx.doi.org/10.1109/TLA.2009.5349049. art. no. 5349049.

[22] E. Berthonnaud, J. Dimnet, P. Roussouly, H. Labelle, Analysis of the sagittal balance of the spine and pelvis using shape and orientation parameters, J. Spinal Disorders Techniques 18 (1) (2005) 40–47, http://dx.doi.org/10.1097/01.bsd.0000117542.88865.77.

[23] P.-C. Zou, J. Wang, S. Chen, H. Chen, Bagging-like metric learning for support vector regression, Knowl.-Based Syst. 65 (2014) 21–30, http://dx.doi.org/10.1016/j.knosys.2014.04.002.

[24] J. Lee, C.-H. Jun, PCA-based high-dimensional noisy data clustering via control of decision errors, Knowl.-Based Syst. 37 (2013) 338–345, http://dx.doi.org/10.1016/j.knosys.2012.08.013.

[25] Y. Zhang, S. Wang, P. Phillips, G. Ji, Binary PSO with mutation operator for feature selection using decision tree applied to spam detection, Knowl.-Based Syst. 64 (2014) 22–31, http://dx.doi.org/10.1016/j.knosys.2014.03.015.

[26] L.V. Utkin, Y.A. Zhuk, Robust boosting classification models with local sets of probability distributions, Knowl.-Based Syst. 61 (2014) 59–75, http://dx.doi.org/10.1016/j.knosys.2014.02.007.

[27] E. Yildizer, A.M. Balci, T.N. Jarada, R. Alhajj, Integrating wavelets with clustering and indexing for effective content-based image retrieval, Knowl.-Based Syst. 31 (2012) 55–66, http://dx.doi.org/10.1016/j.knosys.2012.01.013.

[28] P. Bermejo, J.A. Gámez, J.M. Puerta, Speeding up incremental wrapper feature subset selection with Naive Bayes classifier, Knowl.-Based Syst. 55 (2014) 140–147, http://dx.doi.org/10.1016/j.knosys.2013.10.016.

[29] I. Mandal, N. Sairam, Enhanced classification performance using computational intelligence, Commun. Comput. Informat. Sci. 204 (2011) 384–391, http://dx.doi.org/10.1007/978-3-642-24043-0_39. CCIS.

[30] B. Yet, D.W.R. Marsh, Compatible and incompatible abstractions in Bayesian networks, Knowl.-Based Syst. 62 (2014) 84–97, http://dx.doi.org/10.1016/j.knosys.2014.02.020.

[31] M.W. Fagerland, T-tests, non-parametric tests, and large studies a paradox of statistical practice?, BMC Med Res. Methodol. 12 (78) (2012), http://dx.doi.org/10.1186/1471-2288-12-78.

[32] L. Bai, Z. Wang, Y.-H. Shao, N.-Y. Deng, A novel feature selection method for twin support vector machine, Knowl.-Based Syst. 59 (2014) 1–8, http://dx.doi.org/10.1016/j.knosys.2014.01.025.

[33] I. Mandal, A novel approach for accurate identification of splice junctions based on hybrid algorithms, J. Biomol. Struct. Dyn. (2014) 1–10, http://dx.doi.org/10.1080/07391102.2014.944218. PMID: 25203504.

[34] X. Sun, Y. Liu, M. Xu, H. Chen, J. Han, K. Wang, Feature selection using dynamic weights for classification, Knowl.-Based Syst. 37 (2013) 541–549, http://dx.doi.org/10.1016/j.knosys.2012.10.001.

[35] J.-F. Cordeau, M. Maischberger, A parallel iterated tabu search heuristic for vehicle routing problems, Comput. Oper. Res. 39 (9) (2012) 2033–2050, http://dx.doi.org/10.1016/j.cor.2011.09.021.

[36] Y. Li, D. Li, S. Wang, Y. Zhai, Incremental entropy-based clustering on categorical data streams with concept drift, Knowl.-Based Syst. 59 (2014) 33–47, http://dx.doi.org/10.1016/j.knosys.2014.02.004.

[37] A. Ozcift, A. Gulten, Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms, Comput. Methods Programs Biomed. 104 (3) (2011) 443–451, http://dx.doi.org/10.1016/j.cmpb.2011.03.018.

[38] S.B. Kotsiantis, An incremental ensemble of classifiers, Artif. Intell. Rev. 36 (4) (2011) 249–266, http://dx.doi.org/10.1007/s10462-011-9211-4.

[39] S. Kotsiantis, Combining bagging, boosting, rotation forest and random subspace methods, Artificial Intell. Rev. 35 (3) (2011) 223–240, http://dx.doi.org/10.1007/s10462-010-9192-8.

[40] E. D'Haenens, J. van Damme, P. Onghena, Constructing measures for school process variables: the potential of multilevel confirmatory factor analysis, Quality Quantity 46 (1) (2012) 155–188, http://dx.doi.org/10.1007/s11135-010-9346-4.

[41] L. Minne, S. Eslami, N. de Keizer, E. de Jonge, S.E. de Rooij, A. Abu-Hanna, Statistical process control for validating a classification tree model for predicting mortality – a novel approach towards temporal validation, J. Biomed. Inform. 45 (1) (2012) 37–44, http://dx.doi.org/10.1016/j.jbi.2011.08.015.

[42] J.A. Kairys, J. Orzano, P. Gregory, C. Stroebel, B. DiCicco-Bloom, B. Roemheld-Hamm, F.A. Kobylarz, J.G. Scott, L. Coppola, B.F. Crabtree, Assessing diversity and quality in primary care through the multimethod assessment process (MAP), Qual. Manag. Health Care 10 (4) (2002) 1–14.

[43] C.K. Stroebel, R.R. McDaniel Jr., B.F. Crabtree, W.L. Miller, P.A. Nutting, K.C. Stange, How complexity science can inform a reflective process for improvement in primary care practices, Joint Comm. J. Quality Patient Safety/Joint Commiss. Resourc. 31 (8) (2005) 438–446.