Application of J48 and Bagging for Classification of Vertebral Column Pathologies

Indriana Hidayah , Adhistya Erna P., Monica Agustami Kristy
Department of Electrical Engineering and Information Technology
Gadjah Mada University, Yogyakarta, Indonesia

***Abstract*— Disk hernia and spondylolisthesis are examples of pathologies on vertebral column. These traumas on vertebral column can affect spinal cord capability to send and receive messages from brain to the body systems that control sensor and motor. Therefore, accuracy and timeliness of diagnosis for these pathologies are critical. Hence, a classification system can assist radiologists to improve productivity and the quality of diagnosis. In general, Indonesia's public hospitals have many patients, thus, such classification system will be a great benefit. However, research about pathology of skeletal system classification in Indonesia is rare due to the unavailability of numerical database which quantitatively represents the disease. In this research, dataset of vertebral column from UCI Machine Learning was used to develop an optimum classification model. We ensemble decision tree (J48) and bagging as the classification model. Decision tree was chosen as the base learner due to its simplicity and interpretability. In addition, bagging was used to stable the prediction of new test instances. By applying 10-fold cross-validation we calculated true-positive rate (TP rate), false-positive (FP rate), accuracy parameters, and ROC AUC. The results showed that J48 and Bagging has better performance than J48 alone. The quantitative evaluation showed accuracy of J48 and Bagging is 85.1613%, whereas accuracy of J48 was 81.6129%.**

Keywords*— *vertebral column, disc hernia, spondylolisthesis, J48, bagging.

I. INTRODUCTION

In 2012, there are at least 270.000 people in United States suffer from spinal cord injury (SCI) [1]. One main cause of SCI is trauma in vertebral column [2]. This trauma can affect spinal cord capability to send and receive messages from brain to the body systems that control sensor and motor [3]. Therefore, accuracy and timeliness of diagnosis for these pathologies are critical. An automated diagnosis system or classification system can be used by radiologists with many patients to improve their productivity and diagnosis consistency.

Classification system of pathology or damage bones and joints of skeletal system is an implementation of data mining techniques in medical applications. However, research in this field in Indonesia is rare because there is no database with numerical attributes that is able to quantitatively represent the disease [4]. Therefore, we used dataset of vertebral column from UCI Machine Learning Repository to develop a classification system of vertebral pathologies. The dataset consisted of six attributes, including pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, and grade of spondylolisthesis, with 310 instances. The six attributes are used to classify data into three classes, which are hernia, spondylolisthesis, and normal.

The rest of the paper is organized as follows. In section 2, related works on pathology on vertebral column classifications and the use of ensemble methods are presented. In section 3, the dataset, J48 and Bagging algorithms, as well as the evaluation method are explained. The experimental results of the model proposed in this paper are presented in section 4. Finally, our work of this paper is concluded in the last section.

II. RELATED WORKS

A. Pathology of Vertebral Column Classifications

Main functions of vertebral column are to support human body, to protect medulla spinals and nerve center, as well as to facilitate body movement. Disc hernia and spondylolisthesis are example of pathology of vertebral column. Disc hernia occurs when rubbery cushions, which are disks constructed of soft center enclosed within a tougher exterior, between the vertebrae pushes out through a break in the tougher exterior. Thus, pushes surrounding nerves and result in pain. Furthermore, the herniated disc might be broken and causes serious pain even disability [1]. Whereas, spondylolisthesis is showed by the existence of slippage between vertebras and the bones start to press on nerves [1].

Several works on pathology on vertebral column classification have been done. Rocha et al. developed a framework to aid medical diagnosis of vertebral column pathologies. There were three subsystems, which are graphical interface, pathologies classification, and knowledge extraction. Pathologies classification sub-system used rejection method, which is called rejectSVM, to classify vertebral column disease into two classes, including normal and abnormal class. The experiment results showed that rejectSVM did not outperform the other classifiers; however, its simplicity and interpretability could assist medical experts [6]. Kotti and Diamantaras proposed an alternative to SVM which is called Mean Squared Slack (MSS) method. This method tried to decrease the energy of the loose variables. MSS was used to classify vertebral column dataset into two classes. The experiments showed that accuracy on adult dataset equivalent to 84.951% [7].

978-1-4799-5423-0/14/$31.00 ©2014 IEEE

Reddy et al. used Naïve Bayes classifier to two different datasets of vertebral column. One dataset consists of two class labels, i.e. normal and abnormal. Whereas, the other dataset has three class labels. The experiments showed that Naïve Bayes yielded 83.7419% of accuracy [8].

B. *Application of some Ensemble Methods*

Many research on classification in medical field exploited decision tree and ensemble methods. One common goal to those works was to achieve higher accuracy. In their research, Tu et al used Bagging and decision tree (J48) to classify whether a patient suffers from heart disease or not [9]. Whereas, Ya-Qin et.al. ensemble bagging with decision tree (C5) to predict the living expectancy of breast cancer patients [10]. Asha and Natarajan compared the use of three different ensemble methods, including Bagging, AdaBoost, and Random Forest. The three ensemble methods were used for binary classification of tuberculosis (TB), i.e. pulmonary tuberculosis (PTB) and retroviral PTB [11]. The experiment showed that Random Forest was the weakest method while Bagging was the strongest method in classification of tuberculosis (TB).

C. *Computer-Aided Diagnosis (CAD)*

In radiology, the use of CAD can be defined as the construction of a diagnosis by computation of quantitative analysis of radiological images. Practically, the diagnosis from CAD was used as a second opinion for the radiologist to decide the final diagnosis. By applying CAD, the quality and productivity of radiologist can be improved in terms of accuracy, consistency, and computation time. Common approach of CAD was to detect abnormality and to find possible pathologies that might happen. Furthermore, CAD also can be used for skeletal imaging. Therefore, CAD can be very useful for physicians to assist their work [13].

To evaluate the performance of machine learning algorithm that was implemented by CAD, some parameters were used, among others accuracy, TP (true positive) rate, FP (false positive) rate. Especially in medical diagnosis other parameters were commonly used, such as ROC [14].

Based on previous findings and the need of auxiliary diagnosis system for vertebral column pathologies, in this paper we proposed a vertebral column pathologies classification model by exploiting an ensemble method, which is a combination of bagging and decision tree (J48) as the base learner. Decision tree was chosen because it is a method that is easy to represent and understood. Explanation facility in medical diagnosis systems is very important to aid medical experts. Performance of the classification model was evaluated by 10-fold cross-validation. Several parameters are used, including TP rate, FP rate, accuracy, and ROC AUC. The parameters will be explained in next section.

III. METHODOLOGY

A. *Vertebral column dataset*

Vertebral column dataset used in this paper was taken from UCI Machine Learning Repository. The dataset was in ARFF (Attribute-Relation File Format) format, with 310 instances. This dataset was collected by Dr. Henrique da Mota from Group of Applied Research in Orthopaedics (GARO) in Médico-Chirurgical de Réadaptation des Massues, Lyon, France. There were six numerical attributes that were related to spino-pelvic, including pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, and grade of spondylolisthesis. The spino-pelvic system can be shown in Fig 1.

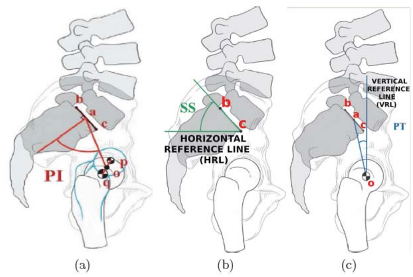

Fig. 1. Spino-pelvic system [6].

Fig 1a shows pelvic radius (PR) that was the distance of line $\overline{ao}$ and pelvic incidence (PI). Where, pelvic incidence (PI) was the angle between line $\overline{oa}$ and line which was drawn from middle of femoral head to the midpoint and perpendicular to sacral. Also in Fig 1a, sacral endplate was shown by line $\overline{bc}$. Lordosis angle was the sagittal angle between sacrum superior plate and lumbar vertebra superior plate or the border of thorax. Sacral slope (SS) in Fig 1b was the angle between sacral endplate ($\overline{bc}$) and horizontal reference line (HRL). In Fig 1c, pelvic tilt (PT) was the angle between vertical reference line (VRL) and a line that connect middle of sacral endplate and femoral heads axis. The last attribute, slipping, was the percentage of slipping between inferior plate on fifth lumbar vertebra and sacrum.

Table 1 showed several instances in the dataset. All of the attributes were used to classify the instances into one of three classes, i.e. disc hernia, spondylolisthesis, or normal.

TABLE I. SEVERAL INSTANCES IN THE VERTEBRAL COLUMN DATASET FROM UCI MACHINE LEARNING REPOSITORY

	1	**2**	**3**	**4**
Pelvic_ incidence	63.028	39.057	74.378	38.505
Pelvic_tilt	22.553	10.061	32.053	16.964
Lumbar_ lordosis_ angel	39.609	25.015	78.772	35.113
Sacral_slope	40.475	28.996	42.325	21.541
Pelvic_radius	98.673	114.405	143.561	127.633
Degree_ Spondylolis-thesis	-0.254	4.564	56.126	7.987
Class Attribute	*Hernia*	*Hernia*	*Spondyl olisthes is*	*Normal*

B. *J48 Classification on the dataset*

Divide-and-conquer strategy is one approach for developing a decision tree which is implemented in algorithms such as C4.5. J48 is an implementation of this approach that is provided by computation software Weka. In this work, we used Weka 3.6.9 to test the proposed model.

Because the evaluation of the classifier was done by using 10-fold cross validation technique, dataset was divided into 10 subsets. Every subset had equal proportion of each class examples. There were 10 iterations, i.e. classifier development. In each iteration, 1 subset was used for testing and the rest of the subsets were used for training. Finally, we take the mean of all of errors from each iteration as the error estimation of the classifier. Post pruning, i.e. pruning that was done after the complete tree have been constructed, was applied to avoid over fitting to the training data

C. Bagging Classification on the dataset

The main function of bagging or bootstrap aggregation is to strengthen a classifier by generating some classifiers then defined the best classifier by majority voting. Therefore, bagging can be used to handle instability of a learning method, such as J48. Fig. 2 showed the working process of bagging.

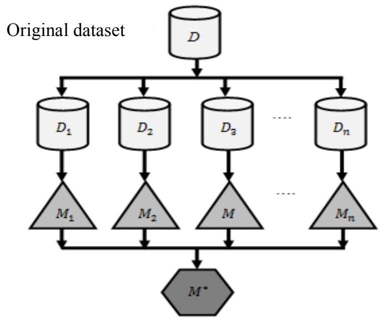

Fig. 2. Bagging process

Each classifier is extracted from a training set that is constructed by randomly delete some instances of the original dataset and replicate some other instances. This method is called sampling with replacement [12].

D. Evaluation on classification result

Evaluation on the classification result was done by applying 10-fold cross-validation as we have explained in section 3.2. Several parameters are chosen to measure the goodness of the classification, including accuracy, TP rate, FP rate, accuracy, and ROC AUC.

Accuracy is the most common parameter to show the predictive capability of a classifier that is represented as percentage [12]. Accuracy of classifier can be computed by (1).

$$Accuracy = \frac{correctly\ classified}{total\ instance} \tag{1}$$

TP rate (true positive rate) shows the percentage of positive examples which are correctly classified as positive. While, FP rate (false positive rate) shows the percentage of negative examples which are misclassified to positive. Therefore, one can evaluate the risks and gains of a classifier by computing TP rate and FP rate [12]. The value of TP rate and FP rate can be calculated by (2) and (3), respectively.

$$TP\ Rate = \frac{TP}{Total\ positive\ instance} = \frac{TP}{TP+FN} \tag{2}$$

$$FP\ rate = \frac{FP}{Total\ negative\ instance} = \frac{TN}{FP+TN} \tag{3}$$

On the other hand, ROC (Receiver Operating Characteristic) curve was a visualization to show trade-off between TP rate and FP rate [12]. Area under ROC curve, also called as area under the curve (ROC AUC), can be described as the probability of TP rate per FP rate [12]. Fig 3 shows an example of ROC curve as in [12].

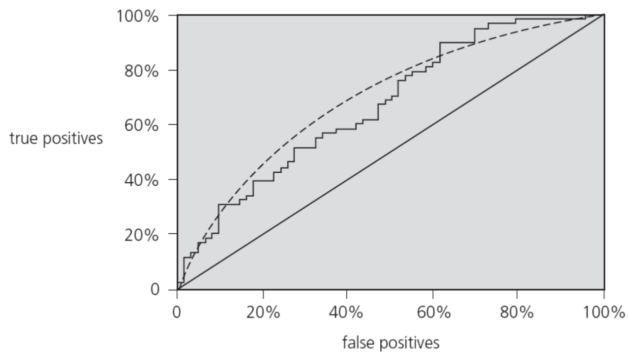

Fig. 3. An example of ROC curve [12]

IV. EXPERIMENTAL RESULTS

The experiments were done by using a machine learning software Weka 3.6.9. Experimental results of the classification by using J48 and Bagging with J48 as a base learner are shown in Fig. 4. Fig. 4a showed the classification result by using J48 algorithm in Weka. Meanwhile, Fig. 4b showed the classification result by using J48 combined with bagging. Based on the values in the confusion matrixes, TP rate, FP rate, and ROC AUC were calculated and shown in Table 2.

The value of TP rate at Hernia class was 0.633 or 63.3%. It showed that 38 instances of total instances in Hernia class, which was 60 instances, successfully classified as the patient of disc hernia. In addition, the value of FP rate was 0.092. FP rate value showed that suppose hernia was considered as the positive class, whereas, Spondylolisthesis and Normal class were considered as negative class, then there were 23 instances of total instances in class spondylolisthesis and normal (250 instance) were classified as Hernia. By having TP rate equaled 0.633 and FP rate equaled 0.092, then the ROC AUC value was 0.835.

Based on the values in Table 2, bagging method yielded better result than only J48. Bagging was able to improve the accuracy, TP rate, and ROC AUC, as well as to decrease FP rate. In medical diagnosis, accuracy of a diagnosis or prediction is very important.

```
Correctly Classified Instances         253               81.6129 %
Incorrectly Classified Instances        57               18.3871 %
Kappa statistic                          0.7045
Mean absolute error                      0.1349
Root mean squared error                  0.3176
Relative absolute error                 32.3769 %
Root relative squared error             69.6203 %
Total Number of Instances              310

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.633     0.092      0.623     0.633     0.628      0.835    Hernia
                 0.967     0.063      0.935     0.967     0.951      0.951    Spondylolisthesis
                 0.7       0.114      0.745     0.7       0.722      0.847    Normal
Weighted Avg.    0.816     0.085      0.813     0.816     0.814      0.895

=== Confusion Matrix ===

   a   b   c   <-- classified as
  38   2  20 |   a = Hernia
   1 145   4 |   b = Spondylolisthesis
  22   8  70 |   c = Normal
```

(a)

```
Correctly Classified Instances         264               85.1613 %
Incorrectly Classified Instances        46               14.8387 %
Kappa statistic                          0.7625
Mean absolute error                      0.1276
Root mean squared error                  0.2636
Relative absolute error                 30.6401 %
Root relative squared error             57.7737 %
Total Number of Instances              310

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.683     0.076      0.683     0.683     0.683      0.942    Hernia
                 0.967     0.025      0.973     0.967     0.97       0.988    Spondylolisthesis
                 0.78      0.11       0.772     0.78      0.776      0.927    Normal
Weighted Avg.    0.852     0.062      0.852     0.852     0.852      0.96

=== Confusion Matrix ===

   a   b   c   <-- classified as
  41   1  18 |   a = Hernia
   0 145   5 |   b = Spondylolisthesis
  19   3  78 |   c = Normal
```

(b)

Fig. 4. Classification result in Weka

TABLE II. CLASSIFICATION RESULT

Class	Parameter	Method	
		J48	Bagging
Hernia	TP rate	0.633	0.683
	FP rate	0.092	0.076
	ROC AUC	0.835	0.942
Spondylolis-thesis	TP rate	0.967	0.967
	FP rate	0.063	0.025
	ROC AUC	0.951	0.988
Normal	TP rate	0.7	0.78
	FP rate	0.114	0.11
	ROC AUC	0.847	0.927
	Accuracy	81.6129%	85.1613%

V. CONCLUSION AND FUTURE WORK

In this paper, two data mining methods were evaluated, i.e. J48 and Bagging by using J48 as the base learner in classifying pathologies of vertebral column dataset. Based on the experimental results we concluded that Bagging was able to improve the accuracy of J48 on vertebral column dataset from 81.6129% to 85.1613%. Therefore, Bagging with J48 as the base learner can be used as an aid for the radiologist to classify pathologies of vertebral column.

For future work the overall performance of the classifier will be further increased. Then, the classifier will be applied to vertebral column dataset in Indonesia. However, as the dataset are not available at this time, a data gathering effort should be incorporated in the future study.

REFERENCES

[1] National Spinal Cord Injury. 2012. Spinal Cord Injury Facts and Figures at Glance. Diakses dari https://www.nscisc.uab.edu pada 2 Juni 2013.

[2] Vogel, L. C., Hickey, K. J., Klaas, S. J., & Anderson, C. J. 2004. Unique Issues in Pediatric Spinal Cord Injury. Orthopaedic Nursing, 23(5), 300-308.

[3] American Association of Neurological Surgeons. Spinal Cord Injury. Diakses dari http://www.aans.org pada 2 Juni 2013.

[4] Neto, A. R., & Barreto, G. A. 2009. WCI 04 On the Application of Ensembles of Classifiers to the Diagnosis of Pathologies of the Vertebral Column: A Comparative Analysis. Latin America Transactions, IEEE (Revista IEEE America Latina), 7(4), 487-496.

[5] Witten, I. H., Frank, E., & Hall, M. A. 2011. Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques. Elsevier.

[6] da Rocha Neto, A. R., Sousa, R., Barreto, G. D. A., & Cardoso, J. S. 2011. Diagnostic of Pathology on The Vertebral Column with Embedded Reject Option. Pattern Recognition and Image Analysis (pp. 588-595). Springer Berlin Heidelberg.

[7] Kotti, M., & Diamantaras, K. I. 2012. Towards Minimizing The Energy of Slack Variables for Binary Classification. Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European (pp. 644-648). IEEE.

[8] Reddy, S.K., SandhyaR.K., and Jaya L.G. 2012 Classification of Vertebral Column using Naïve Bayes Technique. International Journal of Computer Applications (0975 – 8887). Volume 58– No.7, November 2012

[9] Tu, M. C., Shin, D., & Shin, D. 2009. Effective Diagnosis of Heart Disease through Bagging Approach. Biomedical Engineering and Informatics, 2009. BMEI'09. 2nd International Conference on (pp. 1-4). IEEE.

[10] Ya-Qin, L., Cheng, W., & Lu, Z. (2009). Decision Tree Based Predictive Models for Breast Cancer Survivability on Imbalanced Data. Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009. 3rd International Conference on (pp. 1-4). IEEE.

[11] Asha, T., Natarajan, S., & Murthy, K. N. B. 2010. Diagnosis of Tuberculosis Using Ensemble Methods. Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on (Vol. 8, pp. 409-412). IEEE.

[12] Han, Jiawei and Michelin Kamber & Pei, J. 2011. Data Mining: Concepts and Techniques. Morgan kaufmann.

[13] Doi, K. 2007. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. Computerized medical imaging and graphics, 31(4-5), 198-211.

[14] Lasko, T. A., Bhagwat, J. G., Zou, K. H., & Ohno-Machado, L. 2005. The use of receiver operating characteristic curves in biomedical informatics. Journal of biomedical informatics, 38(5), 404-415.