Ensemble Systems for Automatic Fracture Detection

S. K. Mahendran and S. Santhosh Baboo

***Abstract*—Fracture detection based on image classification is an area of research which has proved to be challenging for the past several decades. This field has gained more attention due to the new challenges posed by voluminous image databases. In this research work, fusion-based classifiers are constructed, which extracts features from the images, use these features to train and test the classifiers for the purpose of detecting fractures in X-Ray images. The various features extracted are Contrast, Homogeneity, Energy, Entropy, Mean, Variance, Standard Deviation, Correlation, Gabor orientation (GO), Markov Random Field (MRF), and intensity gradient direction (IGD). Three classifiers, BPNN, SVM and NB classifiers are used. Using these features and classifiers, three single classifiers and four multiple classifiers were developed. All the classifiers were tested vigorously with the test dataset for evaluating the winner combination of classifiers and features that correctly identifies fractures in a bone image. The performance metrics used are sensitivity, specificity, positive predictive value, negative predictive value, accuracy and execution time. The experimental results showed that usage of fusion classifiers enhances the detection capacity and the combination SVM and BPNN produces the best result.**

***Index Terms*—X-ray image, gabor orientation (GO), markov random field (MRF), intensity gradient direction (IGD), BPNN, SVM and NB.**

I. Introduction

Medical imaging is a field that provides 'Quality healthcare' for the patients by using various automated techniques and procedures. Medical imaging is one of the top developments that 'changed the face of clinical medicine' during the last millennium. Today, imaging and radiation therapy are cornerstones of quality care. There is a growing interest during the last decades in finding diagnostic methods for skeletal system diseases [1]. Among these diseases, fractures detection and treatment, which affects people of all ages, is growing importance in modern society. Until recently, X-Ray images were maintained as hard film copy (like a photographic negative). Today, most images are digital files that are stored electronically. These stored images are easilyaccessible and are frequently compared to other X-Ray images for diagnosis and disease management. Now-a-days, X-Ray machines produce extremely high-quality images for radiologists to interpret.

X-Ray image classification is an area that has attracted researchers for the past few decades [2], [3], [4], [5]. Here classification is a pattern recognition problem where the primary goal is to separate a set of images into any one of the two predefined categories, namely, normal or fractured bone. Each of the two classes is represented by a set of features and the algorithm maps these feature vectors to a class using machine learning techniques. The ability to perform image classification as an automatic task using computers is increasingly becoming important in fracture detection domain. This is due to the huge volume of X-ray images available, which are proving to be difficult for manual analysis. The current market need is to have techniques which can detect fractures in X-ray images with minimum intervention from the operators in an efficient and effective manner.

A classifier is a systematic approach to building classification models from an input data set. Examples include Decision Tree Classifiers [6], Rule-Based Classifiers [7], Neural Networks [8], Support Vector Machines [9] and Naïve Bayes Classifiers [10]. Fusion classifiers or Multiple Classifier Systems (MCS) have received considerable attention in applied statistics [11], machine learning [12] and pattern recognition [13] for over a decade. Several studies demonstrate that the practice of combining several base classifier models into one aggregated classifier that leads to significant gains in classification performance over its constituent members [14].

The main focus of this paper is to design an automatic fracture detection system for detecting fractures in long bones from plain diagnostic X-rays using a series of sequential steps. The proposed Automatic Bone Fracture Detection System in Tibia Bones (ABFD-T) consists of three main steps. They are, (i) Preprocessing (ii) Segmentation and (iii) Fracture detection. In this paper the detection process is discussed using a multiple classifier approach. The rest of the paper is presented as follows. Section II discusses the steps in the proposed detection system and presents the various features and classifiers used. Section III discusses the various results obtained. Section IV concludes the work.

II. The Detection System

The fracture detection task is performed in two steps. The first step extracts features from the image and the second step use these features to classify the image as 'normal' or 'fracture'. Texture features and image fusion process are used during detection. Three classifiers namely, BPNN, SVM and NB are used during fusion. All the classifiers are modeled to work as a binary classifier, thus reporting whether a fracture is detected or not detected.

The following 11 texture features of the X-ray images are used to generate the feature vector in the present study. The texture features used are GLCM (Gray Level Co-Occurrence Matrix) features namely, Contrast, Homogeneity, Energy,

Manuscript received October 2, 2011, revised February 16, 2012.

S.K.Mahendran is with the Department of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India (e-mail: sk.mahendran@yahoo.co.in).

S.Santhosh Baboo is with the Postgraduate and Research department of Computer Science at Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai, Tamil Nadu. India (emaill:santhos2001@sify.com).

Entropy, Mean, Variance, Standard Deviation, Correlation. Apart from this, Gabor Orientation (GO), Markov Random Field (MRF), and Intensity Gradient Direction (IGD) features [158] are also used. The features thus collected are stored as a feature vector having 12 columns and 'n' rows, where n is the number of images in the dataset. Each column represents a feature. The data structure used to store the feature vector is a 2-dimensional matrix array as given in Fig. 1.

Feature 1	Feature 2	Feature 3	Feature 4	…	Feature 11	Image Label
Real Value	Real Value	Real Value	Real Value	…	Real Value	0 / 1

Fig. 1. Feature Vector Data Structure.

A. Techniques used during Fusion Classification

Having decided on the number of classifiers and type of classifiers to combine the features extracted from the image, the next step is to decide the specific methods that are to be used for partitioning dataset and aggregation of results.

1) Partitioning(evaluation) method - 'Holdout' Method

Given a data set Z of size N x n, containing n-dimensional feature vectors describing N images, it is desirable to use as much.

As possible of the data to build the classifier (training) and also as much as possible unseen data to test its performance (testing). However, using the same data for training and testing, results in "over-training" of the classifier. In such a situation, the classifier perfectly learns the available data, but fails with unseen data. Thus, it becomes important to have a separate data set to train and test a classifier and make the best use of Z. Several methods exist, like, Re-substitution (R-Method), Hold-Out method (H-Method), Bootstrap method and Cross-validation method.

The proposed fusion classifier uses the hold-out method for splitting the dataset into training and testing samples. The holdout method randomly partitions the dataset into two independent sets, training and testing. Generally, two-thirds of the data are allocated to be the training set and remaining one-third is allocated as test set. The method is pessimistic because only a portion of the initial data is used to derive the model.

2) Aggregation Method

While using multiple classifiers, a method that combines the results of the various classifiers is needed. Several techniques exists, namely, majority voting, maximum, sum, min, average, product, Bayes, decision template and behavior knowledge space. This research work uses the majority voting scheme to combine the outputs of classifiers. Majority vote scheme is one of the oldest strategies for decision making. Its roots are traced back to the era of ancient Greek city states and the Roman Senate. This technique is chosen because of its simplicity and speed. The method is explained below.

Let the decision of the i[th] classifier be defined as $d_{t,j} \in \{0, 1\}$, $t = 1, \ldots, T$ and $j = 1, \ldots, C$, where T is the number of classifiers and C is the number of classes. If the i[th] classifier chooses class ω_j, then $d_{t,j} = 1$ and 0, otherwise. In majority voting scheme, a class ω_J is chosen, if

$$\sum_{t=1}^{T} d_{t,J} = \max_{j=1}^{c} \sum_{t=1}^{T} d_{t,j} \quad (1)$$

The majority voting is an optimal combination rule under the minor assumptions of

- An odd number of classifiers for a two class problem
- The probability of each classifier choosing the correct class is p for any instance x; and
- The classifier outputs are independent

Then, with majority voting, the fusion classifier makes the correct decision if at least $\lfloor T/2 \rfloor + 1$ classifiers choose the correct label, where the floor function $\lfloor . \rfloor$ returns the largest integer less than or equal to its argument. The accuracy of the fusion classifier can be represented by the binomial distribution as the total probability of choosing $k \geq \lfloor T/2 \rfloor + 1$ successful ones out T of classifiers, where each classifier has the success rate of p. Hence, P_{ens}, the probability of fusion classification success is

$$P_{ens} = \sum_{k=(T/1)+1}^{T} \binom{T}{k} p^k (1-p)^{T-k} \quad (2)$$

B. Type of Training

There are various methods used while training a multiple classifier system.

- Training of the individual classifiers and applying aggregation that does not require further training (e.g., aggregation techniques like average, minimum, product, maximum, etc.)
- Training of the individual classifiers followed by training the aggregation
- Simultaneous training of the whole scheme.

The present scheme uses the first method where after training the individual classifier, further classification is not required. This method is selected because the fusion classification depends on the result of the individual classifier.

Using the various features described and the three selected classifiers, seven classifiers (3 single and 4 fusion classifiers) were built. The details are listed below.

a. Texture features with BPNN (T1)
b. Texture features with SVM (T2)
c. Texture features with NB (T3)
d. Texture features with BPNN and SVM (T12)
e. Texture features with BPNN and NB (T13)
f. Texture features with SVM and NB (T23)
g. Texture features with BPNN, SVM and NB (T123)

All the proposed models work in a three-step procedure.

1. Train the classifiers with the training feature vector
2. Use the selected classifiers to classify the test features vector to an output label
3. Perform aggregation to combine the results and make the final decision.

This section presented the various methods and techniques used by the proposed fusion classification system. Classification systems are proposed that combines multiple classifiers and multiple features. Several experiments were conducted to analyze the classifiers and to ascertain which combination produced best results for fracture detection. The results are tabulated and discussed in the next section.

III. Experimental Results

During experiments, a 10-fold cross-validation method is used. The average results were taken as the final outcome of the classifier. Further, a standard three-layered back-propagation network with the tangent-sigmoid transfer function is considered. The weights and biases of the neural networks are initialized randomly, and the number of neurons in the hidden node is determined heuristically as inputs + outputs. A small value of the learning rate (0.15) and a large value of the momentum rate (0.8) are chosen to avoid local minima. The number of training epochs was 500. To implement the principles of SVMs, the LIB-SVM is used. The two most important steps in implementation of SVM is scaling and kernel selection; for scaling, the values of all features were linearly scaled to the range [1, +1] to prevent the cases that features great numeric ranges dominating those in smaller numeric ranges. Among many available kernel functions linear kernel was used. The experimental results of ensemble classification with different base classifiers are presented in the following sections. The experimentation was conducted with the objective of analyzing the resultant data so as to estimate the following two performance behaviors.

1. Which combination of the classifier gives the best result?
2. Which combination of features gives the best result?

The performance metrics used during evaluation are fracture detection rate (DR), False Alarm Rate (FAR) and classification Accuracy (A). Table I shows the performance of the seven classifiers with respect to these performance measures.

TABLE I: Performance of the Classifiers

Classifier	DR	FAR	A
T1	81.10	4.61	86.39
T2	81.23	4.34	87.94
T3	80.69	4.79	82.58
T12	89.05	1.59	97.97
T13	87.00	1.74	96.58
T23	87.77	1.73	97.52
T123	85.05	2.61	93.57

From the table data, it is clear that the performance of the classifiers that combines SVM with texture features produce the best results with respect to fracture detection rate, false alarm rate and accuracy. This is followed by BPNN-based classifiers. While comparing TS1, TS2 and TS3, the TS2 classifier's performance is enhanced. While comparing multiple classifiers, it is clear that the 2-classifer fusion systems perform better than the 3-classifier fusion systems. Among the 2-classifier fusion systems, the combination that fuses the results of BPNN and SVM produces high quality results, followed by SVM and NB combination. Surprisingly, the combination that combines all the three classifiers produces a degraded performance when compared with 2-classifer fusion systems.

Speed of classification is the time taken for the classifiers to make to decide whether a fracture is present in the bone X-ray image or not. The results are presented in Table II.

TABLE II: Time Efficiency (Seconds)

Classifiers	Training Time	Testing Time	Total Time
T1	42.84	31.46	74.30
T2	21.31	18.36	39.67
T3	32.11	24.39	56.50
T12	63.49	43.79	107.28
T13	55.25	41.33	96.58
T23	34.02	36.52	70.54
T123	67.81	46.95	114.76

With regard to time efficiency, the trend obtained is the obvious result. The single classifiers are the fastest, followed by multiple classifiers. However, owing to the high accuracy obtained by 2-classifier systems, this time inefficiency can take a second place while making the decision for winning classifier.

The results clearly indicate that the 2-classification fusion algorithm that combines BPNN and SVM is better than all the other proposed models. This combination shows high performance in terms of accuracy, false alarm rate and detection rate. But when time is considered important, single classifier schemes works better.

IV. Conclusion

The main focus of the present research work is to automatically detect fractures in long bones from plain diagnostic X-Rays using a series of sequential steps. Three classifiers, namely, Back Propagation Neural Networks, Support Vector Machine and Naïve Bayes were considered. Two feature categories, texture and shape, were collected from the X-Ray image. Totally 11 features were extracted from the image which are used to detect the fracture bones through training and testing of classifiers. From these three base classifiers, four fusion classifiers were proposed. Experimental results proved that the fusion of classifier is efficient for fracture detection and achieved maximum accuracy. The time complexity of the algorithms was also on par with the industry requirements. One difficulty encountered with fusion classification is the detection of a classifier which produces the best result. This process could be automated in future and the computer aided diagnosis program can intelligently identify the best combination of classifier and feature to produce highest performance. The present research work considers only simple fractures and experimental results showed that the performance degrades with fractures parallel to the bone edge are not detected as well as those perpendicular to the bone edges. Future research can consider these challenges.

REFERENCES

[1] Materka, A. and Cichy, P. *Texture analysis of x-ray images for detection of changes in bone mass and structure*, M.K. Pietikillinen, editor, Texture Analysis in machine Vision, world Scientific, 2000

[2] Chai, H.Y., Wee, L.K. Swee, T.T., Salleh, S., Ariff, A.K. and Kamarulafizam, R. "Gray-Level Co-occurrence Matrix Bone Fracture Detection", *American Journal of Applied Sciences*, Vol. 8, Issue 1, Pp. 26-32., 2011

[3] Lim, V.L.F., Leow, L.W., Chen, T., Howe, T.S. and Png, M.A. Combining classifiers for bone fracture detection in X-ray images, IEEE International Conference On Image Processing, ICIP 2005, Vol. 1, Pp.I - 1149-1152

[4] Yap, D.W.H., Chen, Y., Leow, W.K., Howe, T.S. and Png, M.A. Detecting femur fractures by texture analysis of trabeculae*, Proc,t. Conf. on Pattern Recognition*, 2004

[5] He, J.C., Leow, W.K. and Howe, T.S. Hierarchical classifiers for detection of fractures in x-ray images, *Proceedings of 12the International Conference on Computer Analysis of Images and Patterns*, Springer-Verlag Berlin, Heidelberg, 2007

[6] Safavian, S.R. and Landgrebe, D. *A survey of decision tree classifier methodology, Systems,* IEEE Transactions on Man and Cybernetics, Vol. 21, Issue 3, Pp. 660-674,1991

[7] Qin, B., Xia, Y., Praphakar, S. and Tu, *Y.* A Rule-Based Classification Algorithm for Uncertain Data, *IEEE Internationa Conference on Data Engineering,* Pp.1633-1640,2009

[8] Zhou, Z. H., Wu, J. X., and Tang, W. Ensembling neural networks: Many could be better than all. Artificial Intelligence, Vol. 137, Issues 1-2, Pp. 239-63,2002

[9] Lee, S., Lee, J.W., Jeong, J.W., Yoo, D.S and Kim, S. A preliminary study on discrimination of osteoporotic fractured group from nonfractured group using support vector machine, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008, Vol. 3, pp.3261-3263.

[10] Wei, D. and Yang, L.X. Weighted Naive Bayesian Classifier Model Based on Information Gain, *International Conference on Intelligent System Design and Engineering Application (ISDEA),* Pp.819-822.,2010

[11] Hastie, T., Tibshirani, R., and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, New York, 2001

[12] Dietterich, T. G. Ensemble methods in machine learning, Kittler, J. and Roli, F. (Eds.), *Proc. of 1st International Workshop on Multiple Classifier Systems (MCS 2001),* Springer-Verlag, Berlin / Heidelberg, 2000

[13] Kuncheva, L.I. *Combining pattern classifiers: methods and algorithms.* John Wiley & Sons, Hoboken, New Jersey, 2004

[14] Eric Bauer and Ron Kohav, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants", *Machine Learning, Kluwer Academic Publishers. Manufactured in The Netherlands*.Vol. 36, Issues 1-2, Pp. 105-39.1999.

[15] Lim, S.E., Xing, Y., Chen, Y., Leow, W.K., Howe, T.S. and Png, M.A. M.A.:Detection of femur and radius fractures in x-ray images, *Proc. 2nd Int. Conf. on Advances in Medical Signal and Information Processing,* Pp.1-8., 2004

S.K.Mahendran is a doctoral students in the department of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India. He is working as HOD in Master of Computer Application Department, Sankara College of Science and Commerce, Saravanampatti, Coimbatore 641035.. His Research interests lies in the area of Medical Image Processing. He has authored three different technical books.

S.Santhosh Baboo has around Seventeen years of postgraduate teaching experience in Computer Science, which includes Six years of administrative experience. He is a member, board of studies, in several autonomous colleges, and designs the curriculum of undergraduate and postgraduate programmes. He is a consultant for starting new courses, setting up computer labs, and recruiting lecturers for many colleges. Equipped with a Masters degree in Computer Science and a Doctorate in Computer Science, he is a visiting faculty to IT companies. It is customary to see him at several national/international conferences and training programmes, both as a participant and as a resource person. He has been keenly involved in organizing training programmes for students and faculty members. His good rapport with the IT companies has been instrumental in on/off campus interviews, and has helped the post graduate students to get real time projects. He has also guided many such live projects. Lt.Dr. Santhosh Baboo has authored a commendable number of research papers in international/national Conference/journals and also guides research scholars in Computer Science. Currently he is Reader in the Postgraduate and Research department of Computer Science at Dwaraka Doss Goverdhan Doss Vaishnav College (accredited at 'A' grade by NAAC), one of the premier institutions in Chennai.