A Breast Tumor Classification Method based on Ultrasound BI-RADS Data Mining

Jin Man Park [†*], Hyoungmin Park[*], Jong-Ha Lee[*], Yeong Kyeong Seong[*], Kyoung-Gu Woo[*], Kyuseok Shim[†]
[*]Samsung Electronics, Yong-In, Korea.
E-mail: { jinman77.park, hm94.park, jongha3.lee, yk.seong, kg.woo }@samsung.com
[†]Seoul National University, Seoul, Korea.
E-mail: shim@ee.snu.ac.kr

***Abstract*— In this paper, to reduce the response time of computer-aided diagnostic (CAD) systems, we proposed a feature selection algorithm that utilizes BI-RADS which is the standard clinical considerations for radiologists to illustrate the visual characteristics of breast tumors. We first apply the association rule mining technique to the medical database annotated with BI-RADS lexicons by doctors, to find out the interesting BI-RADS lexicon values. Then, we select the image processing algorithms which effectively represent the chosen BI-RADS lexicon values. Finally, the features obtained from the selected image processing algorithms are used to build our classifier using Support Vector Machine (SVM) to predict whether each tumor is benign or malignant. Our experimental result shows that our classifier is accurate with fast execution time.**

I. INTRODUCTION

According to American Cancer Society, it is reported [1] that breast cancer is the second leading cause for cancer death in women. It is also reported in [2,3] that early detection and treatment of breast cancers increase significantly the survival rates of breast cancer patients. The ultrasound diagnostic has been utilized widely by radiologists to detect breast cancers early. However, false positive interpretations by radiologists result in many unnecessary biopsies with benign outcomes. To reduce a high number of unnecessary breast biopsies, several computer-aided diagnostic (CAD) systems [12] have been recently developed to assist radiologists in displaying suspicious lesions on medical images by visual analysis and in predicting the malignancy of the lesions by classifying the characteristics of the lesions.

Classification of a breast tumor image in CAD systems is to predict the breast biopsy outcome of its patient, which is either benign or malignant. Such classification tasks consist of three steps [2-7] which are (1) the tumor segmentation, (2) the feature extraction and (3) the tumor classification. In the tumor segmentation step, tumor regions are selected from an original image. In the feature extraction step, feature sets are calculated from the segmented tumor regions by using many image processing algorithms. Finally, in the classification step, by using the extracted feature sets, a classifier predicts whether each tumor in the image is benign or malignant.

The response time of CAD systems is also important as well as their accuracy when radiologists examine patients interactively with CAD systems which requires to produce and process a lot of images that should be classified. The problem is that a lot of algorithms would be examined in the classification step. If we extract many features to improve the accuracy of classifiers in the step (2), many corresponding algorithms should be executed and the response time may suffer. Fortunately, some features are irrelevant or have redundant information to other features and thus hardly contribute to the performance of classifiers. Identifying and removing such features reduce the dimensionality of the data and may allow classifiers to operate faster and more effectively. Actually we found that the feature subset selection method called CFS in [9] improves the accuracy of classification. However, such a feature subset selection method does not consider the total execution times of processing extraction algorithms to generate the chosen features. In this paper, we focus on how to choose effective feature extraction algorithms to reduce the response time of CAD system.

Since radiologists analyze medical images based on the visual characteristics, the tumor classifiers of CAD systems should be modeled according to the clinical reasoning process by radiologists. The Breast Imaging-Reporting and Data System (BI-RADS)[16] provides the standard clinical considerations for radiologists to illustrate the visual characteristics of breast tumors. As it reduces errors in breast imaging interpretations, it is widely used by radiologists. In BI-RADS for ultrasounds images, there are standardized lexicons for describing the characteristics of a tumor. For example, the shape of a tumor can be represented by a "shape" BI-RADS lexicon and the value of the lexicon can be one of "irregular", "round", "oval" and "lobular". We provide a breast image table annotated with BI-RADS lexicons produced by doctors in Table I.

TABLE I. A BREAST IMAGE TABLE ANNOTATED WITH BI-RADS LEXICONS

Image Id	Shape	Margin	Orientation	Diagnostic
1	Irregular	Indistinct	Not Parallel	Malignant
2	Irregular	Angular	Not Parallel	Malignant
3	Round	Circumscribed	Parallel	Benign
4	Round	Circumscribed	Parallel	Benign

Radiologists actually classify the type of a tumor in an ultrasound image as malignant if its shape is irregular. The first and second records in Table I show examples of such diagnoses. In real world, radiologists diagnose the malignancy of each tumor based on not only a single value of BI-RADS

lexicon but also the combinations of several values of BI-RADS lexicons. Thus, in order to build a classifier with good accuracy for breast cancer diagnosis, it would be beneficial to build a classifier using important combinations of BI-RADS lexicons values. Unfortunately, the inputs of CAD systems are not BI-RAD lexicon values but ultrasound images. Thus we should extract suitable features from ultrasound images which describe BI-RAD lexicon values well.

We first apply the association rule mining technique [8] to the medical database annotated with BI-RADS lexicons by doctors, to find out the subset of interesting BI-RADS lexicon values and their combination reflecting the important clinical considerations by doctors for deciding malignancy of breast tumors. We next remove redundant information to get a compact set of BI-RADS lexicon values to be used. Then we select some image processing (feature extraction) algorithms which effectively represent the compact BI-RADS lexicon values from a given set of image processing algorithms. Finally, the features obtained from the selected image processing algorithms are used to build our classifier using Support Vector Machine (SVM) [11] to predict whether each tumor is benign or malignant.

II. PRELIMINARY

We introduce the association rule mining technique [8] which will be used to find out the critical BI-RAD lexicons reflecting the clinical considerations by doctors. Consider an example of a breast image table annotated with BI-RADS lexicons produced by doctors in Table I. Let $I=\{i_1, i_2,\ldots, i_m\}$ be the set of items where an item is a pair of a BI-RAD lexicon and its value. For instance, (Shape: Irregular) is an item where "Shape" is a BI-RAD lexicon and "Irregular" is its value. Let D be a set of records where each record R is a set of items such that $R \subset I$. We say that a record R contains a set of items X if $X \subset R$.

An association rule has the form of $X \rightarrow Y$, where $X \subset I$, $Y \subset I$ and $X \cap Y = \Phi$. In a rule $X \rightarrow Y$, X is called its condition set and Y is the consequence. The rule $X \rightarrow Y$ holds in D with the *confidence* c if the rule $X \rightarrow Y$ holds in the c % of the records containing X in D. The rule $X \rightarrow Y$ has the *support* s in D if the rule $X \rightarrow Y$ holds in the s % of the records in D. We refer to the support and confidence of the rule $X \rightarrow Y$ as support ($X \rightarrow Y$) and confidence ($X \rightarrow Y$) respectively. For example, support(R_1) and confidence(R_1) of the following rule R_1 from Table I are 50% and 100% respectively.

R_1: (Shape:Irregular)&(Diagnostic:Malignant)$\rightarrow$(Orientation:Not Parallel)

Given a set of records D, the problem of mining association rules is to produce all association rules whose supports and confidences are at least the user-specified minimum support and minimum confidence respectively. The two classical association rule mining algorithms are the Apriori [8] and FP-growth [14].

Among the association rules of the form $X \rightarrow Y$, we refer to the association rules with Y = (Diagnostic: Malignant) or Y = (Diagnostic: Benign) as *diagnostic* rules. The rule R_2 shown below is an example of a *diagnostic* rule.

R_2:(Shape:Irregular)&(Orientation:Not Parallel)$\rightarrow$ (Diagnostic:Malignant)

Similarly, the problem of mining *diagnostic* rules is to produce all *diagnostic* rules satisfying minimum support and minimum confidence thresholds.

III. GENERATION OF A CLASSIFIER FOR TUMOR DIAGNOSIS

A. Selection of Interesting BI-RAD Lexicon Values

There are many image processing algorithms which generate the features to be used as input for classifiers of breast cancer diagnosis. However, we want to provide only a small number of important features to our classifier as input since invoking image processing algorithms to generate all possible features is very expensive. Thus, we decided to select the features related to the BI-RADS lexicon values which capture the important clinical knowledge used by radiologists. However, for classification purposes, some of BI-RADS lexicon values may not be useful in the diagnosis of breast cancer. Thus, we next investigate how to find out the critical BI-RADS lexicon values appearing in the *diagnostic* rules.

We will generate a lot of *diagnostic* rules by association rule mining and select a small number of *diagnostic* rules which can capture the important clinical knowledge used by radiologists. When the condition set of a rule r_i is a subset of the condition set of another rule r_j, we want to keep the rule r_j only if the rule r_j has a significantly better confidence since it captures more important clinical knowledge than r_i.

Given a minimum improvement threshold δ ($\delta>0$), we say that a rule r_i *dominates* a rule r_j if (1) the condition set of r_i is a subset of the condition set of r_j, and (2) confidence(r_j) - confidence(r_i) $< \delta$ holds. Given a set of *diagnostic* rules S_D, a *compact diagnostic* rule set S_C with δ is a maximal subset of S_D where (1) every rule in Sc is not dominated by all other *diagnostic* rules in S_C and (2) every rule in ($S_D - S_C$) is dominated by a rule in S_C. The goal of generating *compact diagnostic* rules with δ is to find distinctive and complementary rules enough to capture doctors' clinical knowledge.

To compute the *compact diagnostic* rule set with δ from a set of *diagnostic* rules S_D, a brute-force algorithm may enumerate every possible subset S_C of S_D and check whether S_C is a compact diagnosis rule set. However, it takes exponential time and is thus unpractical to be used. Thus, we next develop the greedy algorithm GEN-CDR(S_D, δ) which is shown in Figure I.

The GEN-CDR works as follows: A min-heap is built from all *diagnostic* rules, which satisfy the minimum support and confidence thresholds, using the number of items in the condition set of each *diagnostic* rule as the key. A min-heap is a complete binary tree [17] such that the rule contained in each node has a smaller number of items in its condition set than the rules in its child nodes. The rule with the minimum number of items in the condition set is placed at the root node in the min-heap. If a pair of rules has the same number of items in their condition sets, the rule with a larger confidence has higher priority in our min-heap.

```
Procedure GEN-CDR(SD, δ)
    Input: SD is a set of Diagnostic rules
    Output: Compact diagnosis rules
    δ: The minimum improvement threshold
    PQ: A min-heap for Diagnostic rules.
1: Insert all Diagnostic rules into the PQ
2: RESULT = { }
3: while PQ is not empty do
4:    R←extract-min (PQ)
5:    RESULT = RESULT ∪ {R}
6:    for each rules S ∈ PQ do
7:       if (R dominates S) then
8:          Remove S from PQ
9: return RESULT
```

Figure 1. GEN-CDR(SD, δ)

After the min-heap is built, a rule R with the highest priority is extracted from the min-heap (line4) by invoking the procedure *extract-min* and is added to RESULT (line5), while the min-heap is not empty. Note that the rule R has the minimum number of items in the condition set and is not dominated by any other rule in the min-heap. We next check whether the rules in the min-heap, which contain the condition set of R, are dominated by the rule R or not. The rules dominated by the rule R are removed from the priority queue (line 8). We guarantee that the rule to be extracted in the next iteration will not be dominated by any other rule in the current RESULT. When the min-heap becomes empty, we return RESULT as a *compact diagnostic* rule set.

To select the interesting BI-RAD lexicon values from the computed *compact* diagnostic rules, we first rank *compact* diagnostic rules based on their confidences and supports. A *compact diagnostic* rule CDR_1 is ranked higher than another *compact diagnostic* rule CDR_2, iff (1) confidence(CDR_1) > confidence(CDR_2) holds or (2) confidence(CDR_1) = confidence(CDR_2) and support(CDR_1) > support(CDR_2) holds.

The top-J *compact diagnostic* rules, denoted by $TopCDR_J$, are the top-J highest ranking *compact diagnostic* rules. An example of $TopCDR_3$ of *compact diagnostic rules* generated by GEN-CDR in Section III.A is shown in Table II.

TABLE II. AN EXAMPE OF $TopCDR_3$

Rule	Support	Confidence
Shape:irregular → Malignant	51.86%	99%
EchoPattern:homogeneous & Margin:circumscribe → Benign	28.26%	99%
Margin:speculated →Malignant	16.43%	99%

After $TopCDR_J$ are selected, we compute a set of important BI-RADS lexicon values that is the union of the condition sets of CDRs in $TopCDR_J$. We refer to the important BI-RADS lexicons and their values as BDV_J.

$$BDV_J = \cup_{CDR \in TopCDRJ} \{v \mid v \in \text{condition set of CDR}\}$$

For example, the BDV_3 of the $TopCDR_3$ in Table II is {Shape:irregular, Margin:circumscribe, EchoPattern:homogeneous, Margin:speculated}

B. *Selection of Effective Image Processing Algorithms*

Intuitively, the features related to BDV_J will improve the accuracy of tumor malignancy classification. Thus, we choose the image processing algorithms with descriptive features for BDV_J.

To measure the describability of an image processing algorithm for BDV_J, we exploit one of classification techniques such as SVM [11]. Let IPA be an image processing algorithm. We first extract the feature vectors from a given set of images D and represent each image with its extracted feature vector. Then for each BI-RADS lexicon value in BDV_J, we train and test a classifier on D with the existence/absence of the value as the class label. For a classifier C, the describability of IPA to a value v in BDV_J , denoted by $desc_C$(IPA, v), is defined as

$desc_C$(IPA, v) = The ratio of images that are correctly classified as v by the classifier C

Then, the describability of IPA to BDV_J is defined as the average describability of every value in BDV_J.

$$desc_C(\text{IPA}, BDV_J) = \text{avg}_{v \in BDV\_J}\ (desc_C(\text{IPA}, v))$$

To use a smaller number of expensive image processing algorithms, we decided to use the above describability. Furthermore, we remove the BI-RADS lexicon values from BDV_J which do not satisfy a user-defined accuracy threshold for the all IPAs. Then, we simply choose IPAs with K highest describabilities from a set of IPAs and call it the K-effective IPAs represented by $EIPA_k$.

The experimental evaluation show that the classifier trained by the outcomes of $EIPA_k$ has faster execution time and comparable accuracy to a classifier trained by every features created by twenty IPAs[7] and selected by traditional feature selection algorithms such as CFS[9].

IV. EXPERIMENTAL EVALUATION

To show the effectiveness of our proposed algorithms, we conducted a performance study. All experiments reported in this section were performed on the machine with Intel(R) Xeon(R) CPU 2.93GHz and 16GB of main memory running 64 bit Windows operating systems. For our experiments, we use a data set which contains 5,252 images of tumors with biopsy results. In the dataset, 2,745 cases are benign results and 2,507 cases are malignant results. For each tumor region, all BI-RADS lexicons are annotated by professional medical doctors and a contour line of the region is automatically segmented by our image processing tool.

We utilized WEKA [10,13] to generate *diagnostic* rules from the BI-RADS lexicons of a data set. We set the minimum support and minimum confidence as 10% and 90% respectively. We also set δ used by GEN-CDR as 2%. For the extraction of features from the automatically segmented tumor images, twenty IPAs[7] are implemented. The $EIPA_5$ generated from $TopCDR_3$ in Table II is shown in Table III in the non-increasing order of IPAs' describability.

TABLE III. $EIPA_5$

ID	IPA for feature extraction
IPA_1	Spatial gray-level dependence matrix (SGLD) [2]
IPA_2	Fourier with centroid distance (Magnitude) [7]
IPA_3	The average gray changes between tissue area and mass area
IPA_4	The number of depressions [2]
IPA_5	Fourier with shape context [7]

TABLE IV. COMBINATIONS OF $EIPA_5$

ID	Combination of IPAs	ID	Combination of IPAs
G1	All IPAs	G5	IPA_1, IPA_2 , IPA_3
G2	Apply the CFS to G1	G6	IPA_1, IPA_2 , IPA_3, IPA_4
G3	IPA_1	G7	IPA_1,IPA_2 ,IPA_3,IPA_4,IPA_5
G4	IPA_1, IPA_2		

We conducted our performance study for each combination of $EIPA_5$ in Table IV using the implementation of SVM [11] in the library of LIBSVM [15] with RBF kernel (γ = 1/the number of features). When we perform our evaluation, we test the classifier with 10-fold cross-validation. The classification accuracy for each combination of $EIPA_5$ in Table IV is shown in Figure 2. The execution time of each combination of $EIPA_5$ is shown in Figure 3.

In Figure 2, the accuracy is the ratio of correctly classified the images out of all test images. The combination of IPA_1 and IPA_2, represented by G4, shows comparable accuracy to the combination of all IPAs denoted by G1.

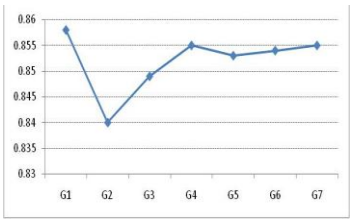

Figure 2. The classification accuracy for each combination of EIPA5 of Table IV

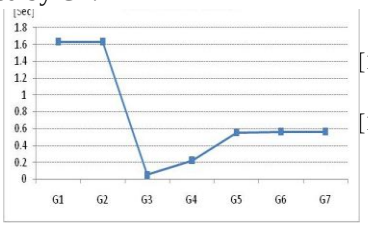

Figure 3. The execution time for each combination of EIPA5 of Table IV

In Figure 3, the execution time shown is the time spent by all IPAs in each combination. We found that G3 is the fastest one and G4 is the second best one. Note that G4 suggested by our technique is a lot faster than G1 with comparable accuracy for the ultrasound CAD systems.

V. CONCLUSION

We proposed a feature selection algorithm that utilizes the *compact diagnostic* rule set and developed a classifier for breast cancer diagnosis. Our experimental result shows that our classifier is accurate with fast execution time.

VI. REFERENCES

[1] A Jemal, R Siegel, E Ward , et al. Cancer statistics. CA CancerJ Clin 2006; 56:106–130

[2] WC Shen. "Breast Ultrasound Computer-Aided Diagnostic Using BI-RADS Features". Academic Radiology Vol.14, Issue 8, 2007, 928–939.

[3] KG Kim, JH Kim, BG Min. Classification of malignant and benign tumors using boundary characteristics in breast ultrasonograms. J Digit Imaging 2002; 15(suppl 1):224–227

[4] RF Chang, Wu WJ, Moon WK, et al. Support vector machines for diagnostic of breast tumors on US images. Acad Radiol 2003; 189–197.

[5] Chen, Breast Lesions on Sonograms: Computer-aided Diagnostic with Nearly Setting-Independent Features and Artificial Neural Networks, Radiology 2003

[6] K Drukker, ML Giger, CJ Vyborny, et al. Computerized detection and classification of cancer on breast ultrasound. Acad Radiol 2004; 526–535.

[7] JH Lee. Shape Context Based Frequency Descriptor for Computer-Aided B-Mode Ultrasound Diagnostic of Breast Cancer, EMBC 2012

[8] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. VLDB, pages 487-499, Santiago, Chile, September 1994R.

[9] M. A. Hall, "Correlation-based Feature Subset Selection for Machine Learning". (1998). Hamilton ,New Zealand

[10] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In KDD'98, New York, NY, Aug.1998

[11] V.N. Vapnik, The Nature of Statistical Learning Theory. Springer: New York, 1995

[12] JJ Fenton, SH Taplin, PA Carney, L Abraham, EA Sickles, C D'Orsi et al. Influence of computer-aided detection on performance of screening mammography. N Engl J Med 2007 April 5;356(14):1399-409

[13] Machine Learning Group at University of Waikato, http://www.cs.waikato.ac.nz/ml/weka/.

[14] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In SIGMOD'00, Dallas, TX, May 2000

[15] CC Chang, CJ Lin. http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[16] American College of Radiology. http://www.acr.org/Quality-Safety/Resources/BIRADS/Ultrasound

[17] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein: Introduction to Algorithms (3. ed.). MIT Press 2009: I-XIX, 1-1292