International Journal of Soft Computing 10 (1): 65-75, 2015
ISSN: 1816-9503

A Novel Feature Selection and Discretization Algorithm to Support Medical Image Diagnosis with Efficiency

[1]J. Senthilkumar, [1]D. Manjula, [2]A. Kannan and [3]R. Krishnamoorthy
[1]Department of Computer Science and Engineering,
[2]Department of Information Science and Technology,
College of Engineering, Anna University, Chennai, Tamilnadu, India
[3]Department of Computer Science and Engineering,
Bharathidasan Institute of Technology, BIT Campus,
Anna University, Tiruchirapalli, Tamilnadu, India

Abstract: In this study, researchers propose a novel Automatic Supervised Feature Selection and Discretization algorithm to enhance the classification of medical images (mammograms). The proposed method consists of a new algorithm called, NANO for a filter based supervised feature selection and discretization. This algorithm solves two problems, viz., feature discretization and selection in a single step. An important contribution of the proposed algorithm is the reduction of irrelevant items to be mined. NANO selects the relevant features based on the average global inconsistency and average global cut point measures, speeding up the medical image diagnosis framework. Two set of experiments have been performed to validate the proposed method. Experiments are carried out to validate the performance of NANO algorithm in the task of feature selection and discretization. Performance evaluation was done for the first experiments using precision and recall metrics obtained from the query and retrieved images. The second set of experiments aim at validating the classification accuracy. From the experiments, it is observed that the proposed method shows high sensitivity (up to 98.64%) and high accuracy (up to 96.95%).

Key words: Data discretization, feature extraction, feature selection, support of image classification, query

INTRODUCTION

Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADx) Systems have been successfully introduced in many hospitals and specialized clinics to provide quick access to screening (Ganesan *et al*., 2013; Ribeiro *et al*., 2008). CADx System refers to techniques that diagnosis the test image based on the visual content automatically extracted from image and high level knowledges obtained from experts (Ribeiro *et al*., 2008). Content-Based Medical Image Retrieval (CBMIR) refers to techniques that retrieve images based on their visual content (Ribeiro *et al*., 2009a, b). In the medical domain, the objective of a CBMIR System is to aid the specialist in the medical diagnosis process, retrieving relevant past cases with images revealing proven pathology, along with the corresponding associated clinical diagnoses and other information (Muller *et al*., 2004; Ribeiro *et al*., 2009b). In the medical domain, the objective of a CADx System is to aid the specialist in the medical diagnosis process (Tang *et al*., 2009; Ribeiro *et al*., 2009a, b) and medical image retrieval (Muller *et al*., 2004; Ribeiro *et al*., 2009a, b) past cases with images revealing proven pathology, along with the corresponding associated clinical diagnoses and other information.

Recently, the CADx System development is based on training and test of the approaches (Ribeiro *et al*., 2008, 2009a, b). The classification of medical image content demands the automatic image segmentation and the extraction of the main image features automatically in the Region of Interests (ROIs) with a specific criterion (Mudigonda *et al*., 2001). Image processing algorithms are used to identify region of interests from the medical images and extract such relevant features from the ROIs images, organizing them in feature vectors. The feature vectors are then employed in place of the images to model as transactions which are then used in the classification or mining process. Features quantify intrinsic visual characteristics of the images such as color, shape and texture, leading to vectors with hundreds or even thousands of elements. Contrary to what one would think, having a large number of features can be a problem.

Corresponding Author: J. Senthilkumar, Department of Computer Science and Engineering, College of Engineering, Anna University, Chennai, Tamilnadu, India

Beyer *et al.* (1999) proved that an increasing number of features (and consequently the dimensionality of the data) leads to losing the significance of each feature. Thus, to avoid decreasing the discrimination accuracy, keeping the number of features as low as possible is important and this establishes a tradeoff between the discrimination power and the feature vector size.

In this study, a Novel Automatic Filter Based Supervised Feature selection and Discretization algorithm called, NANO is presented. NANO is employed to solve two problems, namely, feature discretization and selection in a single step. To measure the accuracy of the feature selection task, researchers use IDEA (Ribeiro *et al*., 2009) Method, C4.5 (Quinlan, 1993) and Naive Bayes (John and Langley, 1995) Classification algorithms. The proposed method shows high sensitivity (up to 98.64%) and high accuracy (up to 96.95%). The results testify that the proposed NANO algorithm is well suited for automatic data discretization and selection of most accurate features in real datasets.

LITERATURE REVIEW

CADx techniques use intrinsic visual features of color, shape and/or texture to represent the images which are generated with intelligent rules and compared. Textures are one of the important properties of image regions are often used in computer aided diagnosis (Ganesan *et al*., 2013; Tang *et al*., 2009; Ribeiro *et al*., 2008) and content based medical image retrieval (Ribeiro *et al*., 2009a, b; Kinoshita *et al*., 2007) systems. Texture is one of the fundamental pattern elements used in interpreting pictorial information. There is no known method that is able to consistently and accurately diagnosis medical images. Fusion texture features improve the overall quality of image segmentation (Clausi and Deng, 2005), classification (Li and Shawe-Taylor, 2005) and retrieval (Krishnamoorthy and Sathiya Devi, 2008).

An important drawback of feature fusion design is the "curse of dimensionality" in the medical image analysis. The feature dimension concept recognizes that there is a need for reduction of feature space dimension to produce maximum accuracy, given a finite number of feature vectors for a particular class (Dash and Liu, 1997). When the number of features exceeds a certain limit, the classification accuracy begins to decrease (Huang *et al*., 2009). This is an important problem in the medical image analysis and researchers motivate the feature reduction techniques (Liu and Yu, 2005).

Feature Selection (FS) algorithm aims at choosing a reduced number of features that preserves the most relevant information of the dataset (Liu and Yu, 2005).

Feature Selection (FS) is one of the important and frequently used techniques in data preprocessing for data mining (Ribeiro *et al*., 2008, 2009a; Hall and Holmes, 2003). FS is an active area in statistical pattern recognition (Jain and Zongker, 1997), machine learning (Hall, 2000) and data mining (Kim *et al*., 2000), since 1970. FS Method reduces the number of features, removes irrelevant, redundant or noisy data and brings more confident for Data Mining and Learning algorithms. The feature selection approach supports Data Mining and Learning algorithms to improve the speed, predictive accuracy and result comprehensibility. Feature Selection algorithm has also been widely applied to many fields such as medical image classification (Mudigonda *et al*., 2001), medical image diagnosis (Ribeiro *et al*., 2008, 2009a, b), medical image retrieval (Ribeiro *et al*., 2009a, b; Oliveira *et al*., 2007), medical image restoration (Seng *et al*., 2010) and gene expression analysis (Peng *et al*., 2005). Since, selection of an optimal subset is always very difficult, many research related to feature selection have expressed difficulties to select a best subset.

A typical FS process consists of four basic steps, namely, subset generation, subset evaluation, stopping criterion and result validation (Liu and Yu, 2005). Based on these four steps with different evaluation criteria the FS algorithms can be designed into three categories, such as, the filter model (Ribeiro *et al*., 2008, 2009a; Kira and Rendell, 1992), the Wrapper (Quah and Quek, 2007; Hsu *et al*., 2002; Kohavi and John, 1997) and the Hybrid Model (Liu and Yu, 2005; Das, 2001). Filter models evaluate the "properness" of the feature subset by using intrinsic characteristics of the data (Kira and Rendell, 1992). Popular criteria employed in this direction are distance, information, consistency and dependency measurements (Robnik-Sikonja and Kononenko, 2003). Consistency measurements aim at finding a minimum number of features that separate classes as consistently as the full set of features does. Inconsistency occurs when two instances with the same feature values have different class labels. Filter methods are computationally not expensive, since they do not involve Mining (induction) algorithms (Senthilkumar *et al*., 2009; Ribeiro *et al*., 2009a, b, 2008). However, they can select subsets of features that may not perform well with the algorithm embedded in the user application.

Wrapper algorithms require one Predetermined Mining algorithm (classification, clustering and others) itself and use to evaluate the candidate feature subsets. Wrapper methods generally select features which are more suited to the mining algorithm aiming to improve mining performance than the filter methods but they generally present a higher computational cost (Quah and

Quek, 2007; Hsu *et al*., 2002; Kohavi and John, 1997) and hence these methods are not suitable for medical big data application.

However, an implementation of the Wrapper Feature Selection algorithm is MCES (Quah and Quek, 2007). The main drawback of MCES algorithm is that it does not work with continuous data, requiring data discretization a process of splitting continuous values into discrete intervals. Discretization is a process of splitting continuous values into discrete intervals (Kerber, 1992). The chai-square algorithm proposed by Liu and Setiono (1997) is an improved version of ChiMerge (Kerber, 1992) that s upports both data discretization and feature selection. The algorithm chai-square uses the chai-square statistics to merge consecutive intervals, fusing the consecutive intervals that lead to the smallest chi-square value at each step. The feature selection process is done by removing from the set of features the ones that generate only one interval which are the features that tend to be class-independent.

Hybrid Model attempts to take advantage of the other two models and avoids the pre-specification of a stopping criterion. A typical Hybrid algorithm makes use of both an independent measure and a Mining algorithm to evaluate feature subsets. It uses independent measure to decide the best subsets for a given cardinality and uses the Mining algorithm to select the final best subset among the best subsets across different cardinalities. The quality of results from a Mining algorithm provides a natural stopping criterion in the Hybrid Model but they take higher computational cost (Bacauskienea *et al*., 2009; Liu and Yu, 2005; Das, 2001). On other hand, the Hybrid Model is suitable for few medical applications where time is not a critical issue. But in, in the case of medical diagnosis, time is critical and hence Hybrid algorithms are not suitable.

Another approach for FS is the use of statistical association rules which led to the development of the StARMiner algorithm (Ribeiro *et al*., 2009a, b). The goal of StARMiner is to implement statistical association rule mining to find features that best discriminate images into categorical classes. It is also employed as a baseline for comparison with the method. Due to their potential, the proposed method employ filter based approach in this research to perform feature selection specific for the CADx domain.

In this study, a new feature selection algorithm is proposed to solve two problems in a single step: feature discretization and selection. An important reduction of irrelevant items to be mined is achieved with the proposed NANO algorithm to select the relevant features based on the Average Global Inconsistency (AGI) and Average Global Cut Point (AGCP) measures, speeding up the CADx framework. The analysis built over Precision and Recall (P&R) curves shows that the proposed method, based on AGI and AGCP measures, outperforms all the other methods. Note that this study is an extended version of the preliminary work presented in the conference study (Senthilkumar *et al*., 2009).

THE NANO ALGORITHM

In this study, a novel automatic supervised algorithm called, NANO is presented. NANO algorithm performs simultaneously data discretization and feature selection of the continuous values of the features in a single step. A measure of inconsistence is employed to determine the final number of intervals and to select features. As the number of intervals gets smaller, the number of inconsistence increases. NANO aims at keeping the minimal number of intervals with minimal inconsistency, establishing a tradeoff between these measures. The following descriptions are necessary before detailing the NANO algorithm.

Definition 1: The class is the most important keyword of a diagnosis given by a specialist.

Definition 2: An interval is also called bin and the most frequent class in a bin is called majority class of the bin.

NANO processes each feature separately. Let D be the set of training input transactions. Let f be a feature of the input feature vector F. Let f_i be the value of the feature f in a transaction i. NANO uses a data structure that links f_i to the class C_i, for all, $i^* \in D$ where C_i is the class of transaction i. Let T_r be the interval determined by the NANO algorithm. Let H_{p_i} be the cut point in the bin H and is determined by the NANO algorithm. Each line in the data structure is called an instance.

Definition 3: An instance I_i belongs to an interval T_r if its value f_i is between two consecutive cut points h_p and h_{p+1}, that is, $f_i \in T_r = [h_p, h_{p+1}]$. Figure 1 shows the input of the continuous feature used by the NANO algorithm to perform the following steps:

In step 1, NANO first sorts the continuous value of the features f_i with class C_i from D. To perform effective feature selection and discretization the proposed NANO algorithm uses Quick Sort algorithm proposed by Hoare (1962). Figure 2 shows the input and sorted output of the continuous feature for the processes of the NANO algorithm.

In step 2, NANO defines the initial cut points. Let M_r be the majority class of interval T_r. Let $|M_r|$ be the number of occurrences of M_r in the interval T_r. When determining the data intervals, the algorithm NANO defines a cut points h_p based on the following condition 1.

0.98	0.85	0.65	0.68	0.99	0.70	0.92	0.71	0.72	0.94	0.73	0.74	0.78	0.91	0.75	0.80	0.83	0.81
No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes

Fig. 1: The input for the continuous feature process of the NANO algorithm

0.65	0.68	0.70	0.71	0.72	0.73	0.74	0.75	0.78	0.80	0.81	0.83	0.85	0.91	0.92	0.94	0.98	0.99
No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	No	No

Fig. 2: The input and output process of the NANO algorithm in step 1

1		2		3			4	5			6	7	8		
0.65	0.68 0.70 0.71	0.72	0.73	0.74	0.75	0.78	0.80	0.81	0.83	0.85	0.91	0.92	0.94	0.98	0.99
No	Yes Yes Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	No	No

Fig. 3: The output determined by the NANO algorithm in condition 1

1		2	3		4				6		7		
0.65	0.68 0.70 0.71	0.72	0.73	0.74 0.75 0.78	0.80	0.81	0.83	0.85	0.91	0.92	0.94	0.98	0.99
No	Yes Yes Yes	No	No	Yes Yes Yes	No	Yes	Yes	Yes	No	No	Yes	No	No

Fig. 4: The output process of the NANO algorithm in condition 2

	2	3	4			6		7	
0.65 0.68 0.70 0.71	0.72	0.73	0.74 0.75 0.78	0.80	0.81	0.83	0.85	0.91	0.92	0.94	0.98	0.99
No Yes Yes Yes	No	No	Yes Yes Yes	No	Yes	Yes	Yes	No	No	Yes	No	No

Fig. 5: The final cut point determined by the relation (1)

Condition 1: The class label of the current instance I_i and i should be ≥1. The current instance I_i is different from the class label of the previous instance, i.e., $c_i \neq c_{i-1}$.

Condition 1 may generate too many cut points, especially when working with noisy data. The large number of cut points leads to higher number of intervals. Each interval represents an item in the process of mining patterns, like decision tree classification, association rule mining or pattern clustering. The use of many items potentially generates a huge number of irrelevant rules with low confidence. Hence, it is important to keep the small number of cut points and as a result, generates a small number of items. In this step, NANO produces pure bins which are the bins of lowest possible entropy (zero). This step produces intervals that minimize the inconsistencies present with the discretization process. This procedure defines 8 cut points for the continuous feature. Figure 3 shows the cut point determined by condition1 as an output for the processes of the proposed NANO algorithm.

In step 3, NANO restricts the minimum frequency that a bin must present, so as to avoid a huge number of cut points. NANO restricts the minimal number of occurrences of the majority class allowed in an interval; the algorithm NANO initiates an input parameter value MRPerInt = 2(Minimum Range PerInterval) for the following condition 2.

Condition 2: The number of occurrences of the majority class in an interval T_r must be greater than or equal to the MRPerInt parameter, i.e., $|M_r| \geq$MRPerInt. If the input parameter MRPerInt is not satisfied by the interval $T_r = [h_p, h_{p+1}]$, the right cut point h_{p+1} of the interval T_r is removed. NANO produces fewer bins, for higher values of the MRPerInt parameter. However, some adjustment should be taken before setting up input parameter MRPerInt as higher the MRPerInt, the higher is the inconsistencies generated by the discretization process. Figure 4 shows the fused output determined by condition 2 with the input parameter MRPerInt = 2 of the NANO algorithm. This procedure fused 3 cut points for the continuous feature.

In step 4, NANO fuses consecutive intervals, using the measure of inconsistence rate to determine which

intervals should be merged. Let, M_r be the majority class of an interval T_r. NANO fuses consecutive intervals T_r and T_{r+1}, i.e., those have the same majority class ($M_r = M_{r+1}$) and also have inconsistence rates ς_{T_r} below or equal to an input parameter $\varsigma_{T_{min}} = (0 \le \varsigma_{T_{min}} \le 1)$. This procedure fuses remaining 2 cut points for the continuous feature. The inconsistence rate ς_{T_r} of an interval T_r is given by the relation (1). Figure 5 shows the final cut point determined by the relation (Eq. 1).

$$M_r = M_{r+1}\ \varsigma_{T_r} = \frac{|T_r| - |M_{T_r}|}{|T_r|} \tag{1}$$

In step 5, NANO algorithm calculates Global Inconsistence (GI) measure. Let,T_r be the total number of intervals $(T_1, T_2, ..., T_n)$ in which the features are discretized. For each feature f, the algorithm NANO computes the GI measure ς_{T_r} for with the relation:

$$\xi_{T_r} = \frac{\sum_{T_i \in T}\left(|T_{f_i}| - |M_{T_{f_i}}|\right)}{\sum_{T_i \in T}\left(|T_{f_i}|\right)} \tag{2}$$

In step 6, NANO algorithm calculates Global Cut Point (GCP) measure. Let, H_{P_r} be the total number of cut points (h_p, h_{p+1}, ..., h_{p+r}) in which the feature selection and discretization is performed by the NANO algorithm. For each feature f, NANO computes the GCP measure ξ_{T_r} from given the relation:

$$\xi_{T_r} = \sum_{T_i \in T}\left(|H_{P_r}|\right) \tag{3}$$

where, $|H_{P_r}|$ is the total number of cut points in the feature f for an interval T_r. Finally, the algorithm NANO introduces two measures, called, the Average Global Inconsistency (AGI) and Average Global Cut Point (AGCP) measures. The algorithm NANO uses AGI measure $\bar{\xi}_{T_r}$ and AGCP measure $\bar{\xi}_{T_r}$ based on the following relation (Eq. 4 and 5) for the processes of automatic feature selection and discretization:

$$\bar{\zeta}_{T_r} = \frac{1}{k}\sum_{i=1}^{k}\zeta_{T_i} \tag{4}$$

and the average global cut point as:

$$\bar{\xi}_{T_r} = \frac{1}{k}\sum_{i=1}^{k}\xi_{T_i} \tag{5}$$

where, k is the total number of global inconsistencies and global cut points. The feature selection criterion employed by NANO algorithm removes from the set of attributes every attribute with average global inconsistence rate $\bar{\xi}_{T_r}$ below or equal to an input parameter $\bar{\zeta}_{T_{r\,min}} = (0 \le \bar{\zeta}_{T_{r\,min}} \le 1)$ and average global cut point rate measure less than the input parameter $\bar{\xi}_{T_{r\,min}} = (1 \le \bar{\xi}_{T_{r\,min}} \le k)$.

The algorithm NANO solves two problems in a single step: feature discretization and selection. An important reduction of irrelevant items to be mined is achieved with the proposed NANO algorithm. NANO selects the relevant features based on the AGI and AGCP measures, speeding up the Mining and Classification algorithms. Since, the number of inconsistencies of the features is the factor that contributes more to disturb the Learning algorithm, discarding the most inconsistent attributes contribute to improve accuracy and speed up the learning algorithm. Algorithm 1 summarizes the steps involved in the proposed NANO. It is proved in the experimental section that NANO is well-suited in feature selection and discretization of numeric and ordinal attributes.

Algorithm 1 (The NANO algorithm):

Input: F(Feature vectors), C(Image classes), Input parameters $(\text{MRPerInt}, \varsigma_{T_{r\,min}}, \bar{\zeta}_{T_{r\,min}}, \bar{\xi}_{T_{r\,min}})$.

Output: V(Processed discrete feature vectors).

1. for each feature f∈Fdo.
2. Sort feature vectors with classes are available in f.
3. For each transaction i, create an instance I_i of the form c_i, f_i where $c_i \in C$.
4. To create cut points h_p for the feature vector f, if the class label of the current instance I_i, $i \ge 1$ is different from the class label of the previous instance, i.e., $c_i \ne c_{i-1}$.
5. end for.
6. for each $h_p \in H_{P_r}$ do.
7. Remove h_p according to the number of occurrences of the majority class in an interval T_r must be equal or greater than the MRPerInt parameter, i.e., $|M_r| \ge$ MRPerInt.
8. Remove h_p according to NANO fuses consecutive intervals T_r and T_{r+1} that has the same majority class ($M_r = M_{r+1}$) and inconsistence rates $(\varsigma_{T_r}, \varsigma_{T_{r+1}})$ below or equal an input parameters $\varsigma_{T_{r\,min}} = (0 \le \varsigma_{T_{r\,min}} \le 1)$ according to the Eq. 1.
9. Calculate the global inconsistence measure ξ_{T_r} for each feature f according to the Eq. 2.
10. Calculate the global cut point measure ξ_{T_r} for each feature f according to the Eq. 3.
11. end for.
12. for each $\zeta_{T_r} \in \zeta_{T_r}$ and $\xi_{T_{r_i}} \in \xi_{T_r}$ do.
13. Calculate average global inconsistence measure $\bar{\xi}_{T_r}$ according to the Eq. 4.
14. Calculate average global cut point measure $\bar{\xi}_{T_r}$ according to the Eq. 5.
15. Select the consistent features based on the average global cut point and the average global inconsistence measures.
16. end for.
17. Write the selected features discretized in V.
18. Return V.
19. Stop.

The time complexity measure: The time depends on number of distinct values, their combination and probabilistic distribution of measures. Researchers now analyze time complexity for NANO algorithm to compute

both average global inconsistent and average global cut point measures for the k training input samples with n features. For an n-feature system, the computational complexity of the proposed NANO algorithm, in terms of number of times that the average global inconsistency and average global cut point functions are called of O(n). However, the implementation of some simple optimization in the NANO algorithm uses Quick Sort algorithm during preprocessing phase with the complexity of $O(\log_2^n)$. One source of increased speed is to be more aggressive about merging when constructing the initial list of intervals. In addition, several pairs of intervals, not near each other and it could be merged with only one iteration. However, the lower bound of NANO is $O(\log_2^n)$ the complexity of sorting is introduced in the initial step of the algorithm unless some means can be devised to eliminate this step.

EXPERIMENTAL RESULTS

Performance measures: There are two set of experiments, performed to validate the proposed method. The first set of experiments aims at validating the performance of the NANO algorithm in the task of feature selection and discretizations of continuous features. The second set of experiments aims at evaluating the classification accuracy for the selected features. To validate the proposed method, researchers compare the proposed method (considering only the diagnosis of calcification (benign and malignant), masses (benign and malignant) and normal) with two well-known classifiers.

To measure the effectiveness of feature selection task performed by the NANO algorithms, the proposed method employs an approach based on the well-known Precision and Recall (P&R) graphs. The measures of precision and recall are defined as:

$$\text{Precision} = \frac{\text{TRS}}{\text{TS}} \tag{6}$$

$$\text{Recall} = \frac{\text{TRS}}{\text{TR}} \tag{7}$$

Where:

TR = The number of relevant images in the dataset
TRS = The number of relevant images in the query result
TS = The number of images in the query result

Considering the size of the dataset, K-Nearest Neighbor (K-NN) queries are applied to the image dataset. A K-NN query is a similarity search performed by comparing the feature vectors using a distance function to quantify how close (or similar) each pair of vectors is. An example of a K-nearest neighbor query is: "given the mammogram of Jane Doe (query center), find the five images most similar to it." As a rule of the thumb for analyzing P&R curves, the closer the curve is to the top of the graph, the better the retrieval technique. In the first experiment, researchers have asked K-NN queries, varying from one up to the dataset size, taking the query centers randomly. For each executed query, the values of precision and recall are calculated. The average of the precision values is obtained for each interval and 10% of the recall values are taken to plot the graph. Researchers employ the Euclidean distance as the similarity measure between two feature vectors.

Researchers also measure the accuracy of the selected feature by the proposed NANO algorithm by comparing them with the diagnoses given by specialists and the results of tissue biopsy. To evaluate the classification performance of the proposed method researchers use the measures of accuracy, sensitivity, specificity, false positive rate and false negative rate in suggesting the diagnosis of medical images. The performance measures are described in the following relations:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{8}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{9}$$

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

$$\text{False positive rate} = \frac{\text{FP}}{\text{FP+TN}} \tag{11}$$

$$\text{False nagative rate} = \frac{\text{FN}}{\text{FN+TP}} \tag{12}$$

Where:

TP = The number of true positive cases (abnormal cases correctly classified)
TN = The number of true negatives (normal cases correctly classified)
FP = The number of false positives (normal cases classified as abnormal by the method)
FN = The number of false negatives (abnormal cases classified as normal by the method)

Accuracy is the proportion of the correctly diagnosed cases to the total number of cases. Accuracy measures the ability of the method to correctly classify the cases. Sensitivity measures the ability of the method to identify abnormal (positive) cases. Sensitivity is the measure of how reliable a system is at making positive

identifications or in other words, correctly identifying that which is inspected as being specifically that which is sought. A highly sensitive system will recognize what it is looking for most of the time and rarely produces a false negative. Specificity, measures the ability of the method to identify normal (negative) cases. Specificity is a measure of how well a system can make a negative identification or indicate when something inspected is not what is being sought but something else.

The NANO algorithm presents herein two experiments using two representative datasets. All mammographic datasets are composed of (1024×1024) size. For computational complexity (i.e., image segmentation and feature extraction) each image is resized into fixed size of (16×16), (32×32) and (256×256). As researchers have shown in this section, it is obvious that the consistent features from NANO algorithm based on average global inconsistency and average global cut point measures can be used in the medical image diagnosis. The proposed method is validated using real datasets and compared with PreSAGe (Ribeiro *et al*., 2008), StARMiner (Ribeiro *et al*., 2009b) and ReliefF (Robnik-Sikonja and Kononenko, 2003) Feature Selection algorithms.

Medical image analysis

Image segmentation: The fundamental step of intelligent medical image analysis is segmentation which partitions an image into individual regions of interest. In this research, the image segmentation is achieved in three phases, namely image preprocessing, edge detection and edge refinement. In the image preprocessing phase, researchers identify accurate Area of Interest (AOI) in the image based on level set method and seeded region growing technique. The proposed AOI identification method contains two phases namely; breast contour identification and pectoral muscle removal. First, the proposed method identifies the breast contour identification (breast profile orientation) based on level set method. Second, the proposed method segments the pectoral muscle using the seeded region growing technique. The identified AOI image is submitted to the edge detection phase for further identification of an accurate region edges in the image and the same is carried out as detailed by Li *et al*. (2011) and Adams and Bischof (1994).

In the edge detection phase, researchers identify accurate region edges based on Orthogonal Polynomials Model (Bhattacharyya and Ganesan, 1997; Krishnamoorthi and Kannan, 2009). The Edge Detection Method performs two different tasks in a single step such as orthogonal feature components extraction and edge detection. In the orthogonal feature components extraction stage, a class of orthogonal polynomials obtained from point-spread operators for different sizes of image window is proposed. A simple computational procedure for constructing a complete set of difference operators from these point-spread operators is employed in the Edge Detection Method. Based on the polynomials operators the Edge Detection Method extracts a set of orthogonal feature components as DC energy feature coefficients, AC edge feature coefficients and AC texture feature coefficients from medical image. Then, the extracted orthogonal feature components are utilized to identify the region edges in medical image. In the edge detection stage, researchers conduct the Nair test (Bishop and Nair, 1939) and the F-test (Fisher and Yates, 1997) to separate out the responses toward edge and noise in the orthogonal feature components due to polynomials operators. Finally, the image edges are detected by maximizing Signal to Noise Ratio (SNR). The extracted region edges are submitted to the edge refinement phase for further identification of an accurate region of objects in the image. This procedure is detailed by Krishnamoorthi and Kannan (2009).

In the edge refinement phase, edge based active contour model with level set formulation method based on Orthogonal Polynomials Model (Krishnamoorthi and Kannan, 2009) and level set function (Li *et al*., 2011; Li *et al*., 2010) is employed. The extracted region edges in the edge detection phase are further refined using a variational level set Formulation Method (Li *et al*., 2010). The Edge Refinement Method is a variational level set formulation in which the regularity of the level set function is intrinsically maintained during the level set evolution. The level set evolution is derived as the gradient flow that minimizes energy functional with a distance regularization term and an external energy (edge informations) that drives the motion of the zero level set toward desired locations. The distance regularization term is defined with a potential function such that the derived level set evolution has a unique Forward And Backward (FAB) diffusion effect which can maintain a desired shape of the level set function, particularly a signed distance profile near the zero level set. This method yields a new type of level set evolution called Edge Based Active Contour Model with level set formulation. The distance regularization effect eliminates the need for reinitialization and thereby avoids its induced numerical errors. The Edge Refinement Method also allows the use of more general and efficient initialization of the level set function. In its numerical implementation, relatively large time steps can be used in the finite difference scheme to reduce the number of iterations while

ensuring sufficient numerical accuracy. Based on this procedure, the Edge Based Active Contour Model with level set formulation identifies accurate ROIs in the image. The identified ROIs images are submitted to the visual feature extraction for further automatic feature extraction and the same is carried out as detailed by Li *et al*. (2010).

Extraction of visual features: The next step of intelligent medical image analysis is the extraction of visual features automatically extracted from ROIs. A total of 981-dimensional visual subband statistical and spectral orthogonal polynomials based texture features were computed for each image. It includes features generated by Orthogonal polynomials based texture feature (113 features), subband statistical and spectral orthogonal polynomials based texture feature (448 features), bivariate discrete orthogonal polynomials based texture feature (280 features) and gradient gray level cooccurrence probabilities based texture feature (140 features) as described by Kinoshita *et al*. (2007), Bhattacharyya and Ganesan (1997), Krishnamoorthi and Kannan (2009), Krishnamoorthy and Sathiya Devi (2008), Zhu (2012), Felipe *et al*. (2003) and Haralick *et al*. (1973).

Experiments on the NANO algorithm: The dataset BI-RADS consists of 446 images taken from mammograms collected from the Breast Imaging Reporting and Data System (BI-RADS) of the Department of Radiology of University of the Vienna, available at http://www.birads.at. Each image has a diagnosis composed of three main parts:

- Morphology: mass (circumscribed, indistinct, speculated); architectural distortion; asymmetry dens.; calcifications (amorph, pleomorph, linear, benign)
- BI-RADS: six levels (0-5)
- Histology: benign lesions (breast tissue, cyst, calcifications, ductal hyperplasia, fibrosis, fibroadenoma, fatty tissue, haematoma, harmartomo, lymphangioma, lymphatic node, mastitis, mastopathia, papilloma, sclerosing adenosis and scar); high-risk lesions (atypical ductal hyperplasia, lobular carcinoma *in situ*, phyllodestumor and radial scar) and malignant lesions (Ductal Carcinoma *In Situ* (DCIS), Invasive Ductal Cancer (IDC), Invasive Lobular Cancer (ILC), Invasive tub. cancer and Muc.cancer)

The BI-RADS categorization was developed by the American College of Radiology (ARC) to standardize mammogram reports and procedures. The BI-RADS categorization is summarized in Table 1.

Table 1: BI-RADS assessment categorization

Categories	Description
0	Need additional imaging evaluation
1	Negative
2	Benign finding
3	Probably benign finding (<2% malignant) short interval follow-up suggested
4	Suspicious abnormality (2-95% malignant) biopsy should be considered
5	Highly suggestive of malignancy (>2% malignant) appropriate action should be taken

A total of 496 images are used to test the performance of the proposed method. The dataset is divided into two sets: the training set is composed of 332 images (67% of BI-RADS dataset) and the test set is composed of 164 images (33% of BI-RADS dataset).

The 981 features of the training images and the BI-RADS classifications (class labels) are submitted to the NANO algorithm. The input parameter MRPerInt = 8 which is tuning parameter set by the user and remaining input parameters $\varsigma_{T_{rmin}} = (0 \le \varsigma_{T_{rmin}} \le 1)$, $\overline{\varsigma}_{T_r}$ and $\overline{\xi}_{T_r}$ are automatically computed by the NANO algorithm. The algorithm NANO automatically produced 127 selected features as the most relevant ones, obtaining a reduction of 87% in the feature vector size.

Researchers measure the effectiveness of NANO algorithm in the task of feature selection. For comparison of the proposed algorithm researchers also apply PreSAGe, StARMiner and ReliefF well-known feature selection algorithms. First, PreSAGe feature selection algorithm using the following input parameters: minperint = 8, mintofuse = 0.8 and valreduct = 17% which are tuning parameters set by the user. PreSAGe algorithm return 166 features as the most relevant ones, obtaining a reduction of 83% in the feature vector size. Second, StARMiner, a well-known feature selection algorithm in the image communities return 245 most relevant features, obtaining a reduction of 75% in the feature vector size. The selected features by StARMiner were also taken to compose a feature vector. Finally with ReliefF, a well-known filter feature selection algorithm return 333 most relevant features, obtaining a reduction of 66% in the feature vector size. The selected features by ReliefF were also taken to compose a feature vector.

To build the precision versus recall graphs, researchers consider five cases of feature vectors to represent the images: using 981 original features using the 127 features selected by NANO using the 166 features selected by PreSAGe using the 255 features selected by StARMiner and using the 333 features selected by ReliefF. Similarity, queries are executed and the P&R graphs are constructed. Figure 6 shows the P&R graph obtained with the proposed technique. The P&R graph shows that even with a reduction of 87% of the feature

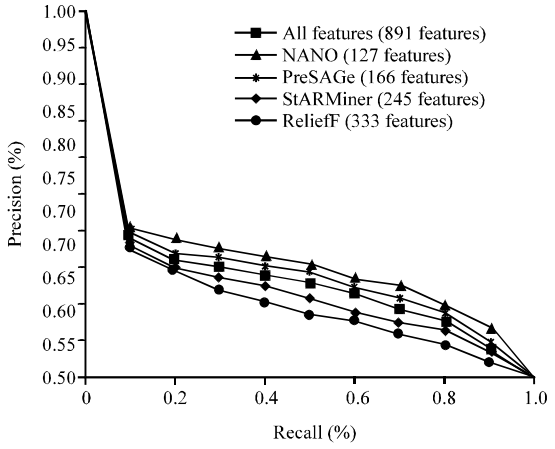

Fig. 6: Precision and recall curves for the BI-RADS image dataset: 981 all features, 127 selected by NANO, 166 selected by PreSAGe, 245 selected by StARMiner and 333 selected by ReliefF

vector size, the precision values are maintained. Moreover, NANO reaches higher values of precision than PreSAGe, StARMiner and ReliefF with regard to time consumption, PreSAGe took 27.3s to select the features and NANO took 23.8s (12.83% less time), StARMiner took 34.51s to select the features and NANO took 23.8s (31.05% less time), Relief took 45.22s to select the features and NANO took 23.8s (47.37% less time). This represents a significant difference for larger datasets. The algorithm NANO executes each feature only once and selects 192 features automatically.

Experiments on medical image classification: The images of the test set and consistent feature produced from NANO algorithm are submitted to IDEA (Ribeiro *et al*., 2009a) Method, C4.5 (Quinlan, 1993) and Naive Bayes (John and Langley, 1995) Classification algorithms. The diagnoses suggested by the Classification algorithms are compared with real diagnoses of images given by specialists and biopsy results. First with C4.5, the classifier constructs a decision tree in the training phase to predict the class labels. Second with Naive Bayes, the classifier that uses a probabilistic approach based on Bayes' theorem predicts the class labels. Finally, with IDEA Method that uses an association rules to predict the class label. Table 2 shows the results for high sensitivity (up to 98.64%) and high accuracy (up to 96.95%) with the BI-RADS dataset. Note that the consistent feature from NANO algorithm leads to higher values of sensibility, specificity and accuracy and also presents smallest error rates both in false positive rate and false negative rate. The results testify that the NANO algorithm is well suited in favor of the data discretization and selection of most consistent features in the real feature sets. The proposed method brings more confidence to the diagnosing process and improves the accuracy of the overall diagnosis process. The diagnosis results are encouraged.

Table 2: A comparison of the classification accuracy (Percentage) of the 127 features from the proposed NANO algorithm using IDEA Method, C4.5 and Naive Bayes Classification algorithms with BI-RADS dataset

Measurements	IDEA Method	C4.5	Naive bayes
Accuracy	96.95	91.46	86.59
Sensitivity	98.64	93.88	90.48
Specificity	82.35	70.59	52.94
False positive rate	17.65	29.41	47.06
False negative rate	1.36	6.12	9.52

CONCLUSION

In this study, a new Feature Selection and Discretization algorithm is proposed. The method discretization turns numeric attributes into discrete ones. Since, data discretization and feature selection have been carried out prior to the learning phase in the classifiers, they significantly reduce the processing effort of the Learning algorithm. Moreover, the proposed NANO algorithm has proved that it is capable of reducing inconsistency, irrelevant attributes and produces good classification accuracy. From the experiments, it is observed that the proposed method shows high sensitivity (up to 98.64%) and high accuracy (up to 96.95%). Hence, the proposed method brings more confidence to the classification process and improves the accuracy of the overall diagnosis process. Finally, the classification result shows that the proposed method could achieve better performance.

ACKNOWLEDGEMENT

Researchers would like to thank M.X. Ribeiro for the productive hints about this research and also provided IDEA prototype with BI-RADS dataset.

REFERENCES

Adams, R. and L. Bischof, 1994. Seeded region growing. IEEE Trans. Pattern Anal. Machine Intell., 16: 641-647.

Bacauskienea, M., A. Verikasa, A. Gelzinisa and D. Valinciusa, 2009. A feature selection technique for generation of classification committees and its application to categorization of laryngeal images. Pattern Recognit., 42: 645-654.

Beyer, K., J. Goldstein, R. Ramakrishna and U. Shaft, 1999. When is nearest neighbors meaningful. Proceedings of the 7th International Conference on Database Theory, January 10-12, 1999, London, UK., pp: 217-235.

Bhattacharyya, P. and L. Ganesan, 1997. An orthogonal polynomials based framework for edge detection in 2-D monochrome images. Pattern Recognit. Lett., 18: 319-333.

Bishop, D.J. and U.S. Nair, 1939. A note on certain methods of testing for the homogeneity of a set of estimated variances. J. R. Stat. Soc., 6: 89-99.

Clausi, D.A. and H. Deng, 2005. Design-based texture feature fusion using gabor filters and co-occurrence probabilities. IEEE Trans. Image Process., 14: 925-936.

Das, S., 2001. Filters, wrappers and a boosting-based hybrid for feature selection. Proceeding of the 18th International Conference on Machine Learning, 28 June-July 1, 2001, San Francisco, CA., USA., pp: 74-81.

Dash, M. and H. Liu, 1997. Feature selection for classification. Intell. Data Anal., 1: 131-156.

Felipe, J.C., A.J.M. Traina and C. Traina, 2003. Retrieval by content of medical images using texture for tissue identification. Proceedings of the 16th IEEE Symposium Computer-Based Medical Systems, June 26-27, 2003, New York, pp: 175-180.

Fisher, R.A. and F. Yates, 1997. Statistical Tables for Biological. Agricultural and Medical Research, London.

Ganesan, K., U.R. Acharya, C.K. Chua, L.C. Min, K.T. Abraham and K.H. Ng, 2013. Computer-aided breast cancer detection using mammograms: A review. IEEE Rev. Biomed. Eng., 6: 77-98.

Hall, M.A. and G. Holmes, 2003. Benchmarking attribute selection techniques for discrete class data mining. IEEE Trans. Knowledge Data Eng., 15: 1437-1447.

Hall, M.A., 2000. Correlation-based feature selection for discrete and numeric class machine learning. Proceedings of the 17th International Conference on Machine Learning, 29 June-July 2, 2000, California, pp: 359-366.

Haralick, R.M., K. Shanmugam and I. Dinstein, 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernet., SMC-3: 610-621.

Hoare, C.A.R., 1962. Quicksort. Comput. J., 5: 10-16.

Hsu, C.N., H.J. Huang and S. Dietrich, 2002. The annigma-wrapper approach to fast feature selection for neural nets. IEEE Trans. Syst. Man Cybern. Part B: Cybern., 32: 207-212.

Huang, S.H., L.R. Wulsin, H. Li and J. Guo, 2009. Dimensionality reduction for knowledge discovery in medical claims database: Application to antidepressant medication utilization study. Comput. Methods Programs Biomed., 93: 115-123.

Jain, A. and D. Zongker, 1997. Feature selection: Evaluation, application and small sample performance. IEEE Trans. Pattern Anal. Mach. Intell., 19: 153-158.

John, G.H. and P. Langley, 1995. Estimating continuous distributions in bayesian classifiers. Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, August 18-20, 1995, San Mateo, CA., pp: 338-345.

Kerber, R., 1992. ChiMerge: Discretization of numeric attributes. Proceedings of the 10th International Conference on Artificial Intelligence, July 12-16, 1992, San Jose, California, pp: 123-128.

Kim, Y., W. Street and F. Menczer, 2000. Feature selection for unsupervised learning via volutionary search. Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 20-23, 2000, Boston, MA., pp: 365-369.

Kinoshita, S.K., P.M. de Azevedo-Marques, R.R. Pereira Jr., J.A.H. Rodrigues and R.M. Rangayyan, 2007. Content-based retrieval of mammograms using visual features related to breast density patterns. J. Digital Imaging, 20: 172-190.

Kira, K. and L.A. Rendell, 1992. A practical approach to feature selection. Proceedings of the 9th International Conference Workshop on Machine Learning, July 12-16, 1992, Aberdeen, Scotland, pp: 249-256.

Kohavi, R. and G.H. John, 1997. Wrappers for feature subset selection. Artif. Intell., 97: 273-324.

Krishnamoorthi, R. and N. Kannan, 2009. A new integer image coding technique based on orthogonal polynomials. Image Vision Comput., 27: 999-1006.

Krishnamoorthy, R. and S. Sathiya Devi, 2008. Design of fusion texture feature with orthogonal polynomials model and co-occurrence property for content based image retrieval. Proceedings of the World Congress on Engineering and Computer Science, October 22-24, 2008, San Francisco, USA.

Li, C., C. Xu, C. Gui and M.D. Fox, 2010. Distance regularized level set evolution and its application to image segmentation. IEEE Trans. Image Process., 19: 3243-3254.

Li, C., R. Huang, Z. Ding, J.C. Gatenby, D.N. Metaxas and J.C. Gore, 2011. A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE Trans. Image Process., 20: 2007-2016.

Li, S. and J. Shawe-Taylor, 2005. Comparison and fusion of multiresolution features for texture classification. Pattern Recognit. Lett., 26: 633-638.

Liu, H. and L. Yu, 2005. Toward integrating feature selection algorithms for classification and clustering. IEEE Trans. Knowl. Data Eng., 17: 491-502.

Liu, H. and R. Setiono, 1997. Feature selection via discretization. IEEE Trans. Knowl. Data Eng., 9: 642-645.

Mudigonda, N.R., R.M. Rangayyan and J.E.L. Desautels, 2001. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. IEEE Trans. Med. Imaging, 20: 1215-1227.

Muller, H., N. Michoux, D. Bandon and A. Geissbuhler, 2004. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. Int. J. Med. Inform., 73: 1-23.

Oliveira, M.C., W. Cirne and P.M. de Azevedo Marques, 2007. Towards applying content-based image retrieval in the clinical routine. Future Gen. Comput. Syst., 23: 466-474.

Peng, H., F. Long and C. Ding, 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell., 27: 1226-1238.

Quah, K.H. and C. Quek, 2007. MCES: A novel monte carlo evaluative selection approach for objective feature selections. IEEE Trans. Neural Networks, 18: 431-448.

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, CA., USA.

Ribeiro, M.X., A.J. M. Traina, C. Traina and P.M. Azevedo-Marques, 2008. An association Rule-based method to support medical image diagnosis with efficiency. IEEE Trans. Multimedia, 10: 277-285.

Ribeiro, M.X., A.G.R. Balan, J.C. Felipe, A.J.M. Traina and C. Traina Jr., 2009a. Mining Statistical Association Rules to Select the Most Relevant Medical Image Features. In: Mining Complex Data, Zighed, D.A., S. Tsumoto, Z.W. Ras and H. Hacid (Eds.). Vol. 165, Springer, Berlin, Heidelberg, ISBN-13: 9783540880677, pp: 113-131.

Ribeiro, M.X., P.H. Bugatti, C. Traina Jr., P.M.A. Marques, N.A. Rosa and A.J.M. Traina, 2009b. Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques. Data Knowl. Eng., 68: 1370-1382.

Robnik-Sikonja, M. and I. Kononenko, 2003. Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learn., 53: 23-69.

Seng, C.H., A. Bouzerdoum, S.L. Phung and M. Amin, 2010. Automatic parameter selection for feature-enhanced radar image restoration. Proceedings of the IEEE Radar Conference, May 10-14, 2010, Washington, DC., pp: 1123-1127.

Senthilkumar, J., D. Manjula and R. Krishnamoorthy, 2009. NANO: A new supervised algorithm for feature selection with discretization. Proceedings of the IEEE International Advance Computing Conference, March 6-7, 2009, Patiala, India, pp: 1515-1520.

Tang, J., R.M. Rangayyan, J. Xu, I. El Naqa and Y. Yang, 2009. Computer-aided detection and diagnosis of breast cancer with Mammography: Recent advances. IEEE Trans. Inform. Technol. Biomed., 13: 236-251.

Zhu, H., 2012. Image representation using separable two-dimensional continuous and discrete orthogonal moments. Pattern Recognit., 45: 1540-1558.