Multi-level Classification : A Generic Classification Method for Medical Datasets

M Srinivas*, R Bharath†, P Rajalakshmi† and C Krishna Mohan*
*VIsual learninG and InteLligence (VIGIL) group,
*Dept. of CSE, †Dept. of EE,
Indian Institute of Technology Hyderabad,
Email: { cs10p002, ee13p0007, raji, ckm } @iith.ac.in

***Abstract*—Classification of medical data is one of the most challenging pattern recognition problems. As stated in literature a single classifier is unable to solve all medical image classification problems due to high sensitivity to noise and other imperfections like data imbalance. So, several individual classifiers have been studied to solve the different types of classification problems arising in medical datasets but all have proven to be useful on some specific datasets. Hence, in this paper, we propose a generic multi-level classification approach for medical datasets using sparsity based dictionary learning and support vector machine approaches. The proposed technique demonstrates the following advantages: 1) gives better performance of classification accuracy over all datasets 2) solves imbalanced data problems 3) needs no fusion and ensemble methods in multi-level classification. The results presented on the 5 standard UCI medical datasets demonstrate that the efficacy of the proposed multi-level classification technique.**

Keywords*—Multi-level classification, Multiple classifiers, Single classifier, Fusion, On-line dictionary learning, Sparse representation, Support vector machines, Breast mass, Heart dataset, SPECTF dataset

I. Introduction

The increasing dependence on modern medical diagnostic techniques like radiology, histopathology and computerized tomography has led to an explosion in the number of medical images stored in hospitals. Several medical image classification systems are available today that classify medical images based on modalities, body parts, disease and orientation. However, one cannot utilize the information in these image collections unless they are organized for efficient search and retrieval of data. Pattern classification helps search algorithms by organizing medical image data. With a single classification technique, it is not possible to solve all medical image classification problems due to imbalanced data problems. So, different classification techniques have to be employed to classify specific datasets. Combinations of multiple classification techniques have been found to give better classification results than a single classifier. In this paper, we use two different types of classification approaches to correctly classify various types of medical image related problems. Here, support vector machine (SVM) and sparse representation based on-line dictionary learning (ODL) classification approaches are used to classify the different medical image related problems.

In [1], shape and texture features are extracted from breast MRI images and genetic algorithm is applied to select the best feature to be used for classification process. To improve classification performance three different classifiers, namely, multi-layer perceptron (MLP), generalized regression neural network (GRNN) and support vector machine (SVM) are combined to from a multi-classifier system. Influence of different types of distance measures on the performance of a multiple classifier system consisting of one-class classifiers were described by Bartosz et al. [2]. One of the problems in medical image classification is that medical datasets are often imbalanced (i.e. more samples of some classes compared to others). Bartosz et al. [3] introduced under sampling balanced ensemble method to solve the imbalance problem. The construction of multiple independent classifiers is typically a non-trivial problem. In [4], atlas-based segmentation and multiple classifiers methods are proposed to solve this non-trivial problem. The application of performance based decision fusion methods to multi-classifier atlas-based segmentation method is evaluated. Each of 20 subjects is segmented using each of the remaining 19 as the atlas. The resulting 19 segmentations per subject are combined into a final segment. The classification methods proposed in literature often have difficulties with breast cancer datasets. The main reason being that training data is imbalanced with more benign cases recorded than malignant ones.

In [5], a cost-sensitive ensemble classification algorithm is proposed. The data imbalance problem is addressed by employing cost-sensitive decision trees as base classifiers which are trained on random feature subspaces to ensure diversity, and an evolutionary algorithm for simultaneous classifier selection and fusion. Yok-Yen Nguwi et al. [6] introduced an unsupervised self-organizing learning with support vector ranking for imbalanced datasets. This model uses support vector machines for selecting variables so that the problem of imbalanced data distribution can be relaxed. Then, the ranker features are clustered using Emergent Self-Organizing Map (ESOM) so as to provide clusters for unsupervised classification. Marco Vannucci et al. [7] described a binary classification method named Labeled Som Classification Unbalanced Sets (LASCUS) that can be applied to uneven datasets and sensitive problems such as malfunction detection. LASCUS method is based on the use of a self-organizing map (SOM) and fuzzy inference system (FIS). The SOM creates a set of clusters to be associated either to frequent or unfrequented situations while the FIS determines such association on the basis of data distribution.

Single classification method is not suitable for classification of various medical image datasets as can be seen in literature [3], [4], [5]. For example, WBC dataset is best classified by KNN classifier and other datasets with KNN classifier gives less classification performance. An extensive literature review revealed the following problems:

1. A single classifier system caters to only a specific medical dataset and performs poorly on others as can be seen in literature [3]. Moreover, it is very susceptible to noise in the data and the performance degrades considerably when noise data is fed as input to any of the individual classifiers.
2. Most of the medical dataset pose data imbalance problems. The imbalanced datasets usually give poor classification performance results with standard single classifiers [6]. A multi-level classifier combines correctly classified examples in the first level with the training data and supplies them as input to the next level classifier. So, if there is any data imbalance regarding less number of training samples it can alleviated by this method.
3. Main problem with multi-classifier system is how to select classifiers to form an ensemble, and how to fuse the individual decisions of the base classifiers into a single decision [5].

The problems stated above could be addressed by using multi-level classification approach which does not require ensemble or fusion methods for combine multiple classifiers. Combining the training data along with correctly classified test samples solves the problem of imbalanced data. The lack of data in training for a given class is compensated by the test samples incorporated in the training data after correct classification.

One of the dictionary learning algorithms, namely, on-line dictionary learning is used in conjunction with support vector machines in the proposed method. Support vector machine is a robust method that has been widely used for classification in various pattern recognition applications. This method was first proposed for classification and regression tasks by Vapnik [8]. We demonstrate the efficacy of our approach on various UCI datasets [9] meant for medical applications. Initially, sparsity based dictionary learning algorithm is applied to classify medical data. Next, correctly classified test data and training data are merged into a single training dataset and given as input to the SVM classifier.

The rest of the paper is organized as follows. Sections 2 gives a brief account on dictionary learning. Section 3 presents the proposed multi-level classification algorithm based on dictionary learning and support vector machine. Experiments on different medical applications are discussed in section 4. Finally, in section 5 we present the conclusions.

II. Sparse representation and On-line dictionary learning

Sparse representation has received a lot of attention from the research in signal and image processing. Sparse coding involves the representation of an image as a linear combination of some atoms in a dictionary [10]. It is a powerful tool for efficiently processing data in non-traditional ways. This is mainly due to the fact that closely related images tend to enjoy the property of being sparse in some dictionary. These dictionaries are often learned directly from training data. Several algorithms like ODL [11], K-SVD [12] and Method of Optimal Directions (MOD) [13] were developed to process this data. A sparsity measure is used to match the input query image with the appropriate class.

Learned dictionaries give sparse models to represent various datasets in of UCI medical data corpus. For a given number of classes N, we design an equal number of dictionaries to represent the classes. Each image is associated with a dictionary that provides the sparsest representation. For every image in the given set of images $\{\mathbf{y}_i\}_{i=1}^n$, ODL is used to seek a dictionary D that has the sparsest representation for the image. We define $l(\hat{\mathbf{D}}, \hat{\Phi})$ as the optimal value of the l_1 sparse coding problem [15]. This is accomplished by solving the following optimization problem.

$$l(\hat{\mathbf{D}}, \hat{\Phi}) = \arg\min_{\mathbf{D},\Phi} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \|\mathbf{Y}_i - \mathbf{D}\Phi_i\|_2^2 \; subject\ to \; \|\Phi_i\|_1 \leq \lambda, \tag{1}$$

where Y is the matrix whose columns are y_i and λ is the sparsity parameter. D denotes the learned Dictionary, Φ represents the sparse representation vectors, N denotes the number of classes and Y represents training database. The ODL algorithm alternates between sparse coding and dictionary update steps.

In the sparse coding step, the dictionary D is fixed and the representation vectors $\mathbf{\Phi_i}$ are identified for each example y_i . Several efficient pursuit algorithms have been proposed in the literature for sparse coding such as [13], [14] etc. The simplest one is the l_1-lasso algorithm [15]. In the next step, the dictionary is updated atom by atom.

III. Multi-level classification approach to Medical data

In this section, we explain the multi-level classification scheme. The motivation for present work is to overcome some problems involved in single and multiple classifier systems related to medical database classification problems which are stated in the previous section.

The proposed multi-level classification scheme for medical datasets is depicted in Figure 1. In the training phase, dictionaries are developed based on sparsity of training feature vectors for each class using on-line dictionary learning and all the dictionaries are combined to form a single dictionary. During testing, the sparsity of a test data is computed with the dictionaries of each class using the l_1-lasso distance. The class which exhibits maximum sparsity is then assigned as the class for that test data. Then correctly classified results are merged with original training dataset to form a new training dataset. The updated training data and the original test data sets are given as input to support vector machine classifier to classify medical database. The three different phases of the proposed classification system, namely, feature extraction, sparse coding based on ODL and SVM are described in the following subsections.

A. Feature Extraction

In this paper, five different types of medical image datasets are used, namely, SPECTF (Heart), Heart-Statlog, Wisconsin Breast Cancer Diagnostic (WBCD), Pima Indians Diabetes (PIMA) and Wisconsin Breast Cancer (WBC) all from the UCI repository. Different medical datasets contain different type of feature values. A brief description of the features extracted from various datasets are presented below:
Dataset 1: Wisconsin Breast Cancer Diagnostic (WBC):

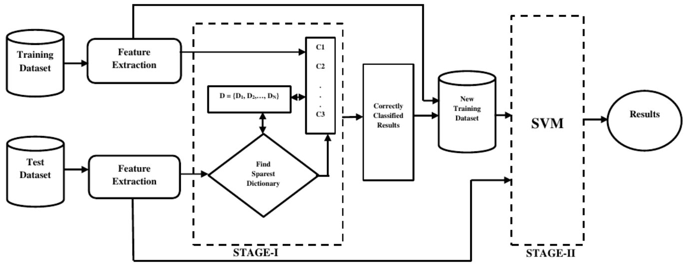

Fig. 1. Block diagram of the proposed Multi-level classification technique using on-line dictionary learning and support vector machine based classification approach.

This data set contains 30 continuous features, computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image, such as the perimeter, the area, the symmetry, and the number of concave portions of the contour.
Dataset 2: Pima Indians Diabetes:

The Pima Indians Diabetes data set contains 8 features. The features include age, number of times pregnant, diastolic blood pressure and body mass index, among others.
Dataset 3: SPECTF (Heart):

The SPECTF data set contains 44 continuous feature pattern was created for each patient.
Dataset 4: Wisconsin Breast Cancer (WBC):

The Breast data set contains 9 features. The features include clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses.
Dataset 5: Heart-Statlog :

The Breast data set contains 13 features. The features include age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electro cardio graphic results, maximum heart rate achieved, exercise induced angina, old peak, the slope of the peak exercise ST segment, number of major vessels.

Now the features extracted from the above datasets are given as input to form a sparse dictionary using on-line dictionary. The following subsection describes about dictionary construction and sparsity based classification approach.

B. On-line Dictionary Learning and Sparsity Based Classification

In the proposed method, we introduce at the first level, a sparsity based medical image classification algorithm by representing the test data a sparse linear combination of training data from a dictionary. On-line dictionary learning is a data-driven approach provides the best possible sparse representation for the images thereby improving the accuracy of classification. Class $C = [C_1, \dots, C_N]$ consists of training samples collected directly from the image of interest. The images related to the same classes are assumed to approximately lie in a low dimensional subspace. For a given N classes, the p^{th} class has K_p training images $\{\mathbf{y}_i^N\}$ i=1,..., K_p. Let b an image belongs to the p^{th} class, then it is represented by a linear combination of these training samples.

$$b = \mathbf{D}^p \Phi^p \tag{2}$$

where D_p is a $m \times K_p$ a dictionary whose columns are the training samples in the p^{th} class. And $\mathit{\Phi}_p$ is a sparse vector.

On-line dictionary learning and sparsity based classification method mainly consists of two steps:

- *Dictionary Construction:* Construct the dictionary for each class of training images using on-line dictionary learning algorithm [11]. Then, the dictionaries $D = [D_1, \dots, D_N]$ are computed using the equation.

$$(\hat{\mathbf{D}}_i, \hat{\Phi}_i) = \arg\min_{\mathbf{D}_i, \Phi_i} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \|\mathbf{C}_i - \mathbf{D}_i \Phi_i\|_2^2 + \lambda \|\Phi_i\|_1 \tag{3}$$

  satisfying $\mathbf{C}_i = \hat{\mathbf{D}}_i \hat{\Phi}_i, \quad i = 1, 2, \dots, N.$

- *Classification:* In this classification process find the sparse vector $\mathit{\Phi}$ for given test image in the test dataset $B = [b_1, \dots, b_l]$. The dictionary of training samples $D = [D_1, \dots, D_N]$, the representation $\mathit{\Phi}$ satisfying $D\mathit{\Phi}$=B is obtained by solving the following optimization problem:

$$\begin{aligned} \Phi^j &= \arg\min_{\Phi} \tfrac{1}{2} \|\mathbf{b_j} - \mathbf{D}\Phi_j\|_2^2 \quad \text{subject to} \|\Phi_j\|_1 \le T_1, \\ \hat{i} &= \arg\min_i \|\mathbf{b_j} - \mathbf{D}\delta_i(\Phi^j)\|_2^2 \quad j = 1, \cdots, t \end{aligned} \tag{4}$$

  where δ_i is a characteristic function that selects the coefficients. Then b_j is assigned to C_i associated with the i^{th} dictionary. It means finding the sparsest dictionary for a given testing data using the l_1 -lasso algorithm. Then, test data is assigned to the class associated with this sparsest dictionary. The following subsection describes the second level classification approach using SVM.

C. Multi-level classification approach

In this approach, the new training dataset is formed based on the first level classification results obtained using dictionary learning. After first level classification, correctly classified results are merged with initial training dataset members to form a new updated training dataset. In the second level of classification, support vector machine learning classifier is used to classify test data based on new updated training dataset. Multi-level classification approach performs better than the single classification approach and more suitable for different types of medical image related datasets.

D. Advantages over other classification schemes

Multi-level vs Single-level : In a multi-level classification system the results of first classifier combined with original training data and form a new training data for second level classifier. This reduces dependency on noisy training data and increases robustness as compared to single-level classifiers. When the results from various classifiers are combined it reduces variation of estimation errors as well. Another problem with single classification techniques is, imbalanced datasets usually gives low classification performance. Thus, using multi-level classification approach addresses the imbalance data problems. *Multi-level vs Multi-classifier :* In multi-classifier systems, the choice of a combination method which is more suitable for an application is a difficult process, having to execute exhaustive testing to choose the best combination method. In some situations, small changes in the structure of the multi-classifier system can drastically change the performance of the combination method and, as a consequence, of the multi classifiers system. Moreover, in multi-classifier fusion and ensemble methods are required for combination of results from various classifiers. However, this process in not defined accurately and may differ from dataset to dataset. Multi-level eliminates this whole process.

E. Experimental results and discussion

In our experiments, we used five different type of medical datasets,namely, SPECTF (Heart), Heart-stalog, Wisconsin Breast Cancer Diagnostic (WBCD), Pima Indians Diabetes (PIMA) and Wisconsin Breast Cancer (WBC) selected from UCI database. Different medical data contains different type and number of feature values shown in Table I.

TABLE I. DATASETS USED IN THE EXPERIMENTS.

Dataset name	# of objects	# of attributes	# of classes
WBC	699	9	2
WBCD	569	32	2
SPECTF	267	44	2
PIMA	768	8	2
Heart-Statlog	270	12	2

In our experiments, we used two different type of breast cancer datasets. Both datasets represent binary classification problems (i.e benign and malignant) and both are highly imbalanced datasets. Wisconsin Breast Cancer original (WBC) is a well known and publicly available breast cancer dataset made available by the University of Wisconsin hospitals [16]. In total, there are 699 samples of which 241 are malignant and 458 are benign. Another breast cancer dataset is Wisconsin Breast Cancer Diagnostic (WBCD) data. WBCD data consist of 569 instances with 32 binary attributes. SPECTF (Single Proton Emission Computed Tomography) Heart data sets. SPECTF Heart dataset is composed as normal and abnormal classes. It consists of 267 instances with 44 attributes. There are 40 samples of each class in the training datasets and test datasets compose 172 normal samples and 15 abnormal samples. PIMA dataset contains the data from all female patients of at least 21 years old, and of Pima Indian heritage. The database consists of 768 instances, each with 8 attributes. Heart-Statlog dataset is composed as absence and presence classes. It consists of 270 instances with 12 attributes.

The performance of the proposed system is evaluated by measuring classification accuracy and sensitivity. Sensitivity is statistical measure for the performance of a binary classification test. The sensitivity is calculated from true positive (TP), false negative (FN), false positive (FP), and true negative (TN).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FN + FP)} \quad (5)$$

$$Sensitivity = \frac{(TP)}{(TP + FP)} \quad (6)$$

The proposed method gives best classification results than existing single and multiple classifier methods applied on five different type of UCI medical datasets. Experiments on the WBCD, WBC, PIMA and Heart-Statlog datasets were run through 5 -fold cross validation. The best results obtained from these experiments are presented in Table II-VI.

Table II shows the performance evaluation obtained on WBCD data. In this table, F-BP, F-kNN, F-SVM, F-Bayes and Multi-agent classifiers give good classification results because they use multi-classification technique. However, determining the right ensemble for fusion is a difficult task. This problem is alleviated in our implementation which utilizes multi-stage classification and hence, the true performance of both classifiers is explored giving rise to 99.1% accuracy. Table III depicts the results of the proposed method and various classifiers on Wisconsin Breast Cancer original (WBC) dataset. It can be noted that the proposed method gives 98 % highest performance compared with the existing single and multiple classification methods.

TABLE II. COMPARISON OF RESULTS WITH STATE-OF-THE-ART ON WISCONSIN BREAST CANCER DIAGNOSTIC DATASET.

Author	Method/Classifier	Accuracy(%)
Balasubramanian, V. et al. 2009 [17]	Random Sampling	92.8
	Margin-based SVM	83.6
	Query by Committee	80
	Ho-WechslersInitial QBT	46.8
	Proposed GQBT	28
Fangqing Peng et al. 2009	Multi-agent	96.58
	F-Bayes	96.32
	F-BP	96.11
	F-kNN	96.26
	F-SVM	96.04
Jing Wei et al. 2013 [18]	k2	94.03
	SDBNS	95.59
	ECFBN	95.76
Proposed	**ODL+SVM**	**99.1**

TABLE III. COMPARISON OF RESULTS WITH STATE-OF-THE-ART ON WISCONSIN BREAST CANCER ORIGINAL (WBC).

Author	Method/Classifier	Accuracy(%)
Myungraee Cha et al. 2014 [19]	support vector data description	94.8
	density weighted SVDD	96.2
	Liu et al. (2013) [20]	96.6
Yuwono, M. et al. 2012 [21]	Multi-agent	96.58
	RCE	96
	RCE+	96.08
	Swarm RCE+	95.89
Duch, W. et al. 2012 [22]	K2MLP	97
Yuanyuan Guo. et al. 2012 [23]	1-NN	92.46
	LLGC	65.52
	SVM	96
	TSVM	97
Ramos-Pollan, R. 2010 [24]	Grid based	95.8
Sheng-Yi Jiang et al. 2009 [25]	C4.5	96.09
	RIPPER	95.99
	Naive-Bayes	97.32
Proposed	**ODL+SVM**	**98**

Table IV represents the performance of proposed method with existing methods on Heart-StatLog dataset. The performance of the proposed method gives 88%, best classification accuracy when compared to other single and multiple classification methods including the state-of-the-art. Table V enlists the various classification schemes applied on the Pima indians diabetes dataset including our proposed method. The PIMA database consists of 768 instances, each with eight attributes. A total of 268 patients were diagnosed as having diabetes and 500 patients are healthy person without diabetes. This dataset contains data imbalance problem. Experimental results on in this dataset, performance of proposed method closely to state-of-the-art.

TABLE IV. COMPARISON OF RESULTS WITH STATE-OF-THE-ART ON HEART-STATLOG DATASET.

Author	Method/Classifier	Accuracy(%)
Christoph F. Eick et al. 2004 [26]	Nearest representative	83.8
	Wilson	80.4
	1-NN	76.7
	C4.5	78.2
Rodda, S et al 2007 [27]	Associative Classifier	82.81
Sheng-Yi Jiang et al. 2009 [25]	C4.5	81.48
	RIPPER	82.33
	Naive-Bayes	84.33
Kemal Polat et al. 2009 [28]	Combining of RBF kernel F-score feature selection and LS-SVM classifier	83.00
Yuanyuan Guo. et al. 2010 [23]	1-NN	53.26
	LLGC	70.40
	SVM	57.80
	TSVM	83.93
Koji Ouchi et al. 2011 [29]	Logistic Regression with a ridge estimator	83.7
	Naive Bayes	83.7
Wodzisaw Duch et al. 2012 [22]	LVQ	85.07
Proposed	**ODL+SVM**	**88**

TABLE V. COMPARISON OF RESULTS WITH STATE-OF-THE-ART ON PIMA INDIANS DIABETES DATASET.

Author	Method/Classifier	Accuracy(%)
Yuwono, M 2012 [21]	RCE	65.64
	RCE+	65.64
	Swarm RCE+	65.6
Duch, W. et al. 2012 [22]	Naive Bayes	75.3
	SVML	77.08
Ouchi, K. et al. 2011 [29]	logistic regression with a ridge estimator	77.21
Yuanyuan Guo et al. 2010 [23]	1-NN	64.84
	LLGC	65.10
	SVM	70
	TSVM	71
Sheng-Yi Jiang et al. 2009 [25]	C4.5	77.73
	RIPPER	77.30
	Naive Bayes	77.28
Chen, S.-C et al. 2006 [30]	SA+BPN (simulated annealing (SA) +BP)	82.16
Zhongwei Li et al. 2006 [31]	Cascade Structure	79.89
Christoph F. Eick et al. 2004 [26]	Nearest representative	74.5
	Wilson	73.4
	1-NN	69
	C4.5	74.5
Proposed	**ODL+SVM**	**82**

Table VI represents the performance of proposed method with existing methods on SPECTF (Heart) dataset. The performance of the proposed method gives 97.8%, best classification accuracy when compared to other single and multiple classification methods including the state-of-the-art. In Table VII, the performance of various single classifiers on the different medical datasets is presented. It can be noted that on PIMA dataset QDA give best performance. However, in all other datasets the proposed classification scheme out performs all others, making it reliable for use over a variety of medical applications.

Figs. 2 shows the sensitivity measure of different type of classifiers on various UCI medical datasets. Proposed method gives best sensitivity results among all classifiers.

TABLE VI. COMPARISON OF RESULTS WITH STATE-OF-THE-ART ON SPECTF (HEART) DATASET.

Author	Method/Classifier	Accuracy(%)
Myungraee Cha et al. 2013 [19]	support vector data description	82.7
	Density Weighted SVDD	95.4
	Liu et al. (2013) [20]	90
Jing Wei et al. 2013 [18]	k2	94.03
	SDBNS	95.59
	ECFBN	95.76
Kumar, R. et al. 2013 [32]	mc-MKC	79.9
	mc-SVM	79.1
Duch, W et al. 2012 [22]	SVMG	80.18
Cui Li-lin et al. 2010 [33]	TCM-IKN N	90
Tian, D. et al. 2007 [34]	C-GAME+Johnson+c4.5	84.4
	RMEP+Johnson+c4.5	41
	C4.5	81.7
Proposed	**ODL+SVM**	**97.8**

TABLE VII. COMPARISON OF RESULTS USING INDIVIDUAL CLASSIFIERS ON DIFFERENT MEDICAL DATASETS.

Method/Dataset	**WBCD**	**WBC**	**Heart-StatLog**	**PIMA**	**SPECTF**
KNN	94.6	96.5	72.5	68.5	70.5
Neural Network	89.3	86.1	82.3	83.3	73.26
Naive Bayes	92	97.2	71.2	75.9	81.8
LDA	89.3	91.6	74.5	81.4	58.2
QDA	92.1	90.2	74.5	**83.3**	53.4
SVM	96.2	85.4	76	81.3	73.4
ODL	96.5	96.5	79	81.4	94.1
(ODL+SVM)	**99.1**	**98**	**88**	82	**97.8**

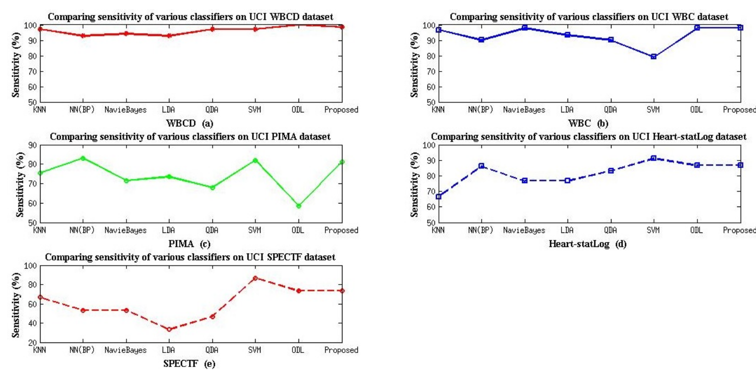

Fig. 2. Sensitivity measure of proposed (ODL+SVM) method on various UCI medical datasets

F. Conclusion

In this paper, a multi-level classification approach using on-line dictionary learning and SVM classification methods for UCI medical data classification is proposed. In all the datasets barring one (PIMA), the performance of the proposed scheme is significantly better than the single classifiers. On-line Dictionary learning being a data-driven approach provides the best possible sparse representation for the images thereby improving the accuracy of classification. Also, it works better than other multiple classifier schemes which suffer from the problem of ensemble selection. Thus, this method proves to be an all-round strategy for medical image classification from various sources.

REFERENCES

[1] F. Keyvanfard, Mahdi Aliyari Shoorehdeli, Mohammad Teshnehlab, "Feature selection and classification of breast MRI lesions based on Multi classifier," in Proc. IEEE Int. Symposium. Artificial Intelligence and Signal Processing (AISP) 2011.

[2] Bartosz. Krawczyk, Micha Wo niak, “Experiments on distance measures for combining one-class classifiers,” in Proc. IEEE Int. Conf. Computer Science and Information Systems, pp. 89-92, 9-12 Sept. 2012.

[3] Bartosz. Krawczyk, Gerald. Schaefer, “Ensemble fusion methods for medical data classification,” in Proc. IEEE Int. Symposium. Neural Network Applications in Electrical Engineering, pp. 143-146, Sept. 2012.

[4] T. Rohlfing, D. B. Russakoff, R. Brandt, R. Menze, C. R. Maurer,“Perfomance based multi classifier decision fusion for atlas based segmentation of biomedical images,” in Proc. IEEE Int. Symposium. Biomedical Imaging: Nano to Macro., pp. 401-407, 15-18 April 2004.

[5] Bartosz. Krawczyk, Gerald. Schaefer, Micha. Woniak, “A Cost-sensitive Ensemble Classifier for Breast Cancer Classification,” in Proc. IEEE Int. Symposium. Applied Computational Intelligence and Informatics (SACI)., pp. 427-430, 23-25 May 2013.

[6] Yok-Yen. Nguwi, Siu-Yeung. Cho, “An unsupervised self-organizing learning with support vector ranking for imbalanced datasets,” in Proc. Expert Systems with Applications., Vol. 37, no. 12, pp. 8303-8312, December 2010,

[7] Marco. Vannucci, Valentina. Colla, “Novel classification method for sensitive problems and uneven datasets based on neural networks and fuzzy logic,”in Proc. Applied Soft Computing., Vol. 11, no. 2, pp. 2383-2390, March 2011.

[8] Vapnik. V. (1995). The Nature of Statistical Learning Theory. New York.

[9] Frank. A, Asuncion. A, (2010). UCI machine learning repository, https://archive.ics.uci.edu/ml/datasets.html

[10] I. Ramirez, P. Sprechmann, Guillermo. Spiro,“Classication and Clustering via Dictionary Learning with Structured Incoherence and Shared Features,” in CVPR, June 2010.

[11] J. Mairal, F. Bach, J. Ponce, G. Sapiro, “Online dictionary learning for sparse coding,” In Proc. of the 26th Int. Conf. on Machine Learning, ICML 09 (ACM), pp. 689-696, USA, 2009.

[12] M. Aharon, M. Elad, A. M. Bruckstein,“The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation,” IEEE Trans. Signal Process., vol. 54 no. 11, pp. 4311-4322, 2006.

[13] K. Engan, S. Aase, J. Hakon-Husoy, “Method of optimal directions for frame design,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 5, pp. 2443-2446, 1999.

[14] S. Mallat, Z. Zhang,“Matching Pursuits with Time Frequency Dictionaries,” IEEE Trans. Signal Processing., Vol. 41, no.12, pp. 3397-3415, 1993.

[15] H. Lee, A. Battle, R. Raina, A. Y. Ng, “Efcient sparse coding algorithms,” in Proc. Advances in Neural Information Processing Systems, vol. 19, pp. 801-808, 2007.

[16] W. Wolberg, O. Mangasarian, ”Multisurface method of pattern separation for medical diagnosis applied to breast cytology,” in Proc. of the National Academy of Sciences, vol. 87, pp. 9193-9196, 1990.

[17] Vineeth. Balasubramanian, Shayok. Chakraborty, Sethuraman. Panchanathan, “Generalized Query by Transduction for Online Active Learning,” in Proc. of IEEE Computer Vision Workshops, 2009.

[18] Jing Wei, Hua. Yu, Juyun. Wang, “The Research of Bayesian Method from Small Sample of High-dimensional Dataset in Poison Identification,” in Proc. IEEE Int. Conf. Software Engineering and Service Science (ICSESS), 23-25 May 2013.

[19] Myungraee. Cha, Jun. Seok. Kim, Jun-Geol. Baek, “Density weighted support vector data description,” Expert Systems with Applications Vol. 41, No. 7, pp. 3343-3350, 1 June 2014.

[20] Liu, B, Xiao. Y. Cao. L. Hao, Z. Deng, F, “SVDD-based outlier detection on uncertain data,” Knowledge and Information Systems, Vol.34, No.3, pp. 597-618, 2013.

[21] Mitchell. Yuwono, Steven. W. Su, Bruce. D. Moulton, Hung. T. Nguyen, “Optimization Strategies for Rapid Centroid Estimation,”in Proc. IEEE Int. Conf. 34th Annual International Conference of the IEEE EMBS San Diego, California USA, 28 August - 1 September, 2012.

[22] Wodzisaw. Duch, Norbert. Jankowski, Tomasz. Maszczyk, “Make it cheap: learning with O(nd) complexity,” in Proc. IEEE World Congress on Computational Intelligence June, 10-15, 2012.

[23] Yuanyuan Guo, Xiaoda Niu and Harry Zhang,”An Extensive Empirical Study on Semi-supervised Learning,”in Proc. IEEE Int. Conf. 2010 IEEE International Conference on Data Mining, 2010.

[24] Ral Ramos-Polln, Jos M. Franco, Jorge Sevilla, Miguel A. GuevaraLpez, Naimy Gonzlez de Posada, Joanna Loureiro and Isabel Ramos,“ Grid Infrastructures for Developing Mammography CAD Systems,” in Proc. IEEE Int. Conf. EMBS 2010.

[25] Sheng-yi Jiang, Xia Li, Qi Zheng, Lian-xi Wang, “Approximate Equal Frequency Discretization Method,” in Proc. IEEE Int. Conf. Global Congress on Intelligent Systems, 2009.

[26] Christoph F. Eick , Nidal Zeidat, and Ricardo Vilalta, “Using Representative-Based Clustering for Nearest Neighbor Dataset Editing,” in Proc. IEEE Int. Conf. on Data Mining, 2004.

[27] Sireesha Rodda M. Shashi, “An Improved Associative Classifier,” in Proc. IEEE Int. Conf. Computational Intelligence and Multimedia Applications, 2007.

[28] Kemal Polat, Salih Gunes, “A new feature selection method on classification of medical datasets: Kernel F-score feature selection,” Expert Systems with Applications Vol. 36, No. 7, pp. 10367-10373 , 2009.

[29] Koji Ouchi, Atsuyoshi Nakamura,Mineichi Kudo,“An efficient construction and application usefulness of rectangle greedy covers,” Pattern Recognition Vol. 47, No. 3, pp. 1459-1468, March 2014.

[30] S.-C. Chen, S.-W. Lin, T.-Y. Tseng and H.-C. Lin, “Optimization of Back-Propagation Network Using Simulated Annealing Approach, ” in Proc. IEEE Int. Conf. Systems,Man, Cybernetics, Oct. 8-11, 2006.

[31] Zhongwei Li, “A Support Vector Machine training Algorithm based on Cascade Structure,” in Proc. IEEE Int. Conf. on Innovative Computing, Information and Control (ICICIC’06), 2006.

[32] Ritwik Kumar, Ting Chen, Moritz Hardt, David Beymer, Karen Brannon, Tanveer Syeda-Mahmood, “Multiple Kernel Completion and its Application to Cardiac Disease Discrimination,” in Proc. IEEE Int. Conf. Biomedical Imaging, April 7-11, 2013.

[33] Cui Li-lin , Zhu Hai-chao , Zhang Lin-ke, Luan Rui-peng, “Improved k Nearest Neighbors Transductive Confidence Machine for Pattern Recognition, ” IEEE Int. Conf. on Computer Design And Appliations 2010.

[34] David Tian, John Keane and Xiao-jun Zeng, “Core-generating Approximate Minimum Entropy Discretization for Rough Set Feature Selection: An Experimental Investigation, ”in Proc. IEEE Int. Conf. on Fuzzy Systems Conference, 2007.