A Multilabel approach using Binary Relevance and One-versus-Rest Least Squares Twin Support Vector Machine for Scene Classification

Divya Tomar
Indian Institute of Information Technology, Allahabad
Uttar Pradesh, India
divyatomar26@gmail.com

Sonali Agarwal
Indian Institute of Information Technology, Allahabad
Uttar Pradesh, India
sonali@iiita.ac.in

***Abstract*—The classification of an image scene having multiple class labels produces significant challenge to the researchers. A semantic scene may be described by multiple objects or by multiple classes. For example, a beach scene may also contain mountain or buildings in the background. This research work proposes a multi-label scene classification model by using Binary Relevance (BR) based one-versus-rest Least Squares Twin Support Vector Machine (LSTSVM). Fifteen evaluation metrics have been used to analyze and compare the result of the proposed scene classification model with the six existing multi-label classifiers. Experimental results demonstrate the superiority and usefulness of the proposed model in the classification of multi-label scene over the existing multi-label approaches.**

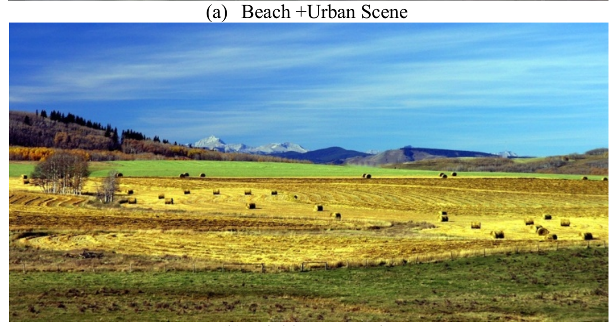
Keywords-Scene Classification; Multi-label Learning; Least Squares Twin Support Vector Machine; Binary Relevance.

I. INTRODUCTION

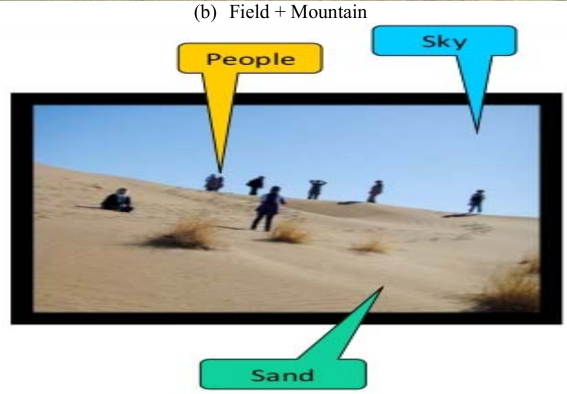
Scene classification is a challenging task of image classification in which images are categorized into classes such as mountains, roads, sky or beach. One of the most useful applications of it is automatic image annotation in digital album. Various researches have been done in the field of identification of a scene from images [1-10]. Only few of them considered this task as a multi-label classification problem [11]. An image may contain various scenes or objects at the same time. For example, a field scene can be described by a mountain and fall foliage or a sand scene includes people, sky etc. Figure 1 shows some examples of multi-label scene classification problem. Therefore, the objective of this study is to consider the task of scene classification as a multi-label classification task and develop a scene classification system using image database.

In multi-label classification problem, the objective is to develop a model in which a set of class labels are assigned to each instance or data sample as opposed to unilabel multi classification system where a data sample belongs to only single class label [12-14]. The researchers are attracted to multi-label learning because of its applicability in various fields. Some real world applications require a multi-label classifier, for example, in movie recommendation system a movie may belong to different categories (romantic-thriller, comedy-romantic or comedy-action etc.), a text also belongs to a variety of categories.

(a) Beach +Urban Scene



(b) Field + Mountain



(c) Sky +People +Sand
Figure 1. Multi-label scene classification

Other than these applications, multi-label classifier is useful for email-classification, medical diagnosis where a person may

978-1-5090-0210-8/16 $31.00 © 2016 IEEE
DOI 10.1109/CICT.2016.17

diagnosis with several diseases at the same time, emotion detection, image annotation, sentiment classification or personality prediction [15-24]. A multi-label classifier learns from a training dataset in which multiple class labels are assigned to a data sample at the same time. Thus multi-label classifier predicts all the possible class labels for a new data sample rather than a single class label.

The solution to multi-label classification problem is broadly categorized as-Problem Transformation and Algorithm adaptation approaches [14, 25-26]. In problem transformation method, the multi-label classification task is broken down into various unilabel classification tasks. In this approach, any existing classifier can solve the problem of multi-label classification as it is independent from the algorithm. Binary Relevance is one of the most widely used approaches of problem transformation method [27]. In this approach the dataset used for multi-label classification task is broken down into $|L|$ unilabel datasets for $|L|$ class labels. In this way dataset is generated for each class label and a binary classifier is constructed for each class label. Spyromitros et al. constructed a classifier with multiple class labels by using Binary Relevance problem transformation method where k-Nearest Neighbor was selected as a base classifier [28]. In another research work, Tanaka et al. proposed a multi-label classification system for functional genomics by adopting Decision Tree as a base classifier for Binary Relevance approach [29]. Classifier by Chaining (CC), Ranking by single label, ranking by pairwise comparison, calibrated label ranking (Quick Weighted algorithm for Multi-label), Label Powerset (RAkEL) are some examples of problem transformation approaches [30-33]. On the other hand, in algorithm adaptation approach the classifiers are modified so that they can work well in the multi-label classification scenario. AdaBoost.MH, AdaBoost.MR, Ensemble of Classifier Chain (ECC), Multi-label k-NN (ML k-NN) are some algorithm adaptation approaches [34].

This study adopts Binary Relevance and LSTSVM approach in order to develop a model for scene classification. In this research work, a multi-label image scene classification task is divided into several unilabel classification problems. LSTSVM is a recently proposed binary classifier which generates two non-parallel hyper-planes by solving two objective functions associated with equality constraint in order to classify the data samples of each class [35]. LSTSVM is the least square variant of Twin Support Vector Machine (TWSVM) which is proposed by Jayadeva et al. to solve the computational complexity problem of well-known Support Vector Machine (SVM) classifier [37]. TWSVM is four times faster as compared to the SVM as it requires the optimization of two smaller QPPs instead of a single complex QPP as in SVM. TWSVM has shown its superiority over SVM not only in terms of speed but also in terms of generalization ability [36-40]. Further, Kumar et al. replaced the inequality constraints of TWSVM associated with two QPPs by equality constraints and proposed a least square variant of TWSVM, termed as LSTSVM in which two QPPs of TWSVM is transformed into two linear equations which are easy to solve. Thus the implementation of LSTSVM classifier is easier as compared to the TWSVM. Therefore, this study develops a scene classification model in which LSTSVM acts as a base classifier for BR approach. The effectiveness of the proposed scene classification model is evaluated using fifteen performance evaluation metrics with six existing multi-label classification approaches. Results of the experiment demonstrate the superiority of the given scene classification model.

The paper is divided into four sections. Section 2 discusses the proposed Binary Relevance based one-versus-rest LSTSVM. In section 3, the dataset details, description of performance evaluation metrics and experimental results on scene dataset are given. Finally, section 4 contains concluding remarks.

II. Binary Relevance Based One-verus Rest Least Squares Twin Support Vector Machine

Consider a multi-label dataset D with N data samples and m features. The dataset is represented as $D = (x_i, Y_i), i = 1, \dots, N$, where x_i is a feature vector, $Y_i \subseteq L$ indicates the true class label set and $L = \{c_j, j = 1, \dots, M\}$ is the total label set. The objective of multiple class label learning is to obtain a multi-label classifier that optimizes the performance evaluation parameters. This study adopts Binary Relevance (BR) problem transformation approach to develop a multi-label scene classification model. Although BR does not take into account the label dependency, yet it provides several advantages over other existing multi-label classification approaches [27]. For example, BR approach has linear complexity with respect to the number of class labels and any binary classifier can be considered as a base learner. This approach divides the multi-label classification problem into $|L|$ number of unilabel classification problems by dividing the multi-label dataset into $|L|$ datasets each for one class label. It considers the data samples of j^{th} class label as the data samples of positive class and rest of the other data samples as the data samples of negative class for each dataset $D_j, 1 \leq j \leq |L|$. The class label for a new data sample is determined by combining the class labels which are positively predicted by each classifier.

In this study, we have used a one-versus-rest Least Squares Twin Support Vector Machine (LSTSVM) classifier as a base learner. LSTSVM has shown better predictive performance over existing classification approaches in terms of generalization ability and computational speed. LSTSVM generates two non-parallel hyper-planes, one for each class by solving two objective functions associated with equality constraints [35]. For each dataset D_j, consider a matrix A_j comprises the data samples of j^{th} class label and B_j consists the data samples of other class labels. Let the number of data samples in A_j matrix is N_j and the data samples of B_j matrix contains (N-N_j) data samples. One-versus-rest LSTSVM solves following two objective functions:

$$\min(w_{+1j}, b_{+1j}, \xi_j)\ \frac{1}{2}\left\|A_j w_{+1j} + e_{+1j} b_{+1j}\right\|^2 + \frac{c_{+1j}}{2}\xi_j^T \xi_j$$
$$\text{s.t.} -\left(B_j w_{+1j} + e_{-1j} b_{+1j}\right) + \xi_j = e_{-1j} \quad (1)$$

$$\min(w_{-1j}, b_{-1j}, \eta_j)\ \frac{1}{2}\left\|B_j w_{-1j} + e_{-1j} b_{-1j}\right\|^2 + \frac{c_{-1j}}{2}\eta_j^T \eta_j$$
$$\text{s.t.}\ \ \left(A_j w_{-1j} + e_{+1j} b_{-1j}\right) + \eta_j = e_{+1j} \quad (2)$$

and determines the following two hyper-planes corresponding to the j[th] class label:

$$f_{+1j} = \left(w_{+1j}.x\right) + b_{+1j} = 0 \quad (3)$$

and

$$f_{-1j} = \left(w_{-1j}.x\right) + b_{-1j} = 0 \quad (4)$$

Here, $w_{+1j}, w_{-1j} \in R^m$ are the normal vectors to the hyper-planes and b_{+1j} and b_{-1j} are the bias terms. $e_{+1j} \in R^{N_j}$ and $e_{-1j} \in R^{N-N_j}$are the vectors of 1's, $c_{+1j}, c_{-1j} > 0$ and ξ_j, η_j are trade-off constants and slack variables respectively. Lagrangian function corresponding to equation (1) is determined as:

$$\mathrm{L} = \frac{1}{2}\left\|A_j w_{+1j} + e_{+1j} b_{+1j}\right\|^2 + \frac{c_{+1j}}{2}\xi_j^T \xi_j - \alpha_j(-\left(B_j w_{+1j} + e_{-1j} b_{+1j}\right) + \xi_j - e_{-1j}) \quad (5)$$

where α_j is a lagrangian multiplier. Necessary and sufficient optimality KKT conditions are determined as:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{w}_{+1j}} = \mathrm{A}_j^{\mathrm{T}}\left(\mathrm{A}_j \mathrm{w}_{+1j} + \mathrm{e}_{+1j}\mathrm{b}_{+1j}\right) + \mathrm{B}_j^{\mathrm{T}}\alpha_j = 0 \quad (6)$$

$$\frac{\partial \mathrm{L}}{\partial \mathrm{b}_{+1j}} = \mathrm{e}_{+1j}^{\mathrm{T}}\left(\mathrm{A}_j \mathrm{w}_{+1j} + \mathrm{e}_{+1j}\mathrm{b}_{+1j}\right) + \mathrm{e}_{-1j}^{\mathrm{T}}\alpha_j = 0 \quad (7)$$

$$\frac{\partial L}{\partial \xi_j} = c_{+1j}\xi_j - \alpha_j = 0 \quad (8)$$

$$\frac{\partial L}{\partial \alpha_j} = \left(B_j w_{+1j} + e_{-1j} b_{+1j}\right) - \xi_j + e_{-1j} = 0 \quad (9)$$

Equations (6) and (7) lead to:

$$\begin{bmatrix}\mathrm{A}_j^{\mathrm{T}} \\ \mathrm{e}_{+1j}^{\mathrm{T}}\end{bmatrix}\left[\mathrm{A}_j\ \ \mathrm{e}_{+1j}\right]\begin{bmatrix}\mathrm{w}_{+1j} \\ \mathrm{b}_{+1j}\end{bmatrix} + \begin{bmatrix}\mathrm{B}_j^{\mathrm{T}} \\ \mathrm{e}_{-1j}^{\mathrm{T}}\end{bmatrix}\alpha_j = 0 \quad (10)$$

Let $u_{+1j} = \begin{bmatrix}\mathrm{w}_{+1j} \\ \mathrm{b}_{+1j}\end{bmatrix}$, $X_j = \left[\mathrm{A}_j\ \ \mathrm{e}_{+1j}\right]$ and $Y_j = \left[\mathrm{B}_j\ \ \mathrm{e}_{-1j}\right]$. With these notations equation (10) can be redefined as:

$$X_j^T X_j u_{+1j} + Y_j^T \alpha_j = 0\ \text{ or }\ u_{+1j} = -(X_j^T X_j)^{-1} Y_j^T \alpha_j \quad (11)$$

α_j is obtained by solving equations (8), (9) and (11) as:

$$\alpha_j = c_{+1j}(e_{-1j} + Y_j u_{+1j}) \quad (12)$$

Equations (11) and (12) lead to:

$$u_{+1j} = \begin{bmatrix}\mathrm{w}_{+1j} \\ \mathrm{b}_{+1j}\end{bmatrix} = -(Y_j^T Y_j + \frac{1}{c_{+1j}} X_j^T X_j)^{-1} Y_j^T e_{-1j} \quad (13)$$

In the same way,

$$u_{-1j} = \begin{bmatrix}\mathrm{w}_{-1j} \\ \mathrm{b}_{-1j}\end{bmatrix} = (X_j^T X_j + \frac{1}{c_{-1j}} Y_j^T Y_j)^{-1} X_j^T e_{+1j} \quad (14)$$

The decision function regarding class assignment to a new data sample is given below:

$$f(x) = arg \min_{i=+1,-1} \frac{|w_{ij}.x + b_{ij}|}{\|w_{ij}\|} \quad (15)$$

LSTSVM also works well for the data samples which are not separable by linear class boundaries. To make easier separation it uses kernel function to transform the data samples into higher dimensional space. Non-linear one-versus-rest LSTSVM solves following two objective functions:

$$\min(\mu_{+1j}, \gamma_{+1j}, \xi_j)\ \frac{1}{2}\left\|K(A_j, D_j^T)\mu_{+1j} + e_{+1j}\gamma_{+1j}\right\|^2 + \frac{c_{+1j}}{2}\xi_j^T \xi_j$$
$$\text{s.t.}\quad -\left(K(B_j, D_j^T)\mu_{+1j} + e_{-1j}\gamma_{+1j}\right) + \xi_j = e_{-1j} \quad (16)$$

$$\min(\mu_{-1j}, \gamma_{-1j}, \eta_j)\ \frac{1}{2}\left\|K(B_j, D_j^T)\mu_{-1j} + e_{-1j}\gamma_{-1j}\right\|^2 + \frac{c_{-1j}}{2}\eta_j^T \eta_j$$
$$\text{s.t.}\ \left(K(A_j, D_j^T)\mu_{-1j} + e_{+1j}\gamma_{-1j}\right) + \eta_j = e_{+1j} \quad (17)$$

and generates following kernel surfaces:

$$f_{+1j} = \left(\mu_{+1j}.K(x^T, D_j^T)\right) + \gamma_{+1j} = 0 \quad (18)$$

and

$$f_{-1j} = \left(\mu_{-1j}.K(x^T, D_j^T)\right) + \gamma_{-1j} = 0 \quad (19)$$

where K is appropriately chosen kernel function and $D_j = [A_j\ B_j]^T$. Similar to equations (13) and (14), kernel surface parameters are determined as:

$$\begin{bmatrix}\mu_{+1j} \\ \gamma_{+1j}\end{bmatrix} = -(Q_j^T Q_j + \frac{1}{c_{+1j}} P_j^T P_j)^{-1} Q_j^T e_{-1j} \quad (20)$$

In the same way,

$$\begin{bmatrix}\mu_{-1j} \\ \gamma_{-1j}\end{bmatrix} = (P_j^T P_j + \frac{1}{c_{-1j}} Q_j^T Q_j)^{-1} P_j^T e_{+1j} \quad (21)$$

Here, $P = \left[K(A_j, D_j^T)\ \ e_{+1j}\right]$ and $Q = \left[K(B_j, D_j^T)\ \ e_{-1j}\right]$. The new data sample is classified according to the following formula:

$$f(x) = arg \min_{i=+1,-1} \frac{|\mu_{ij}.x + \gamma_{ij}|}{\|\mu_{ij}\|} \quad (22)$$

In this way, binary classifiers are constructed for each multi-label dataset. Each binary classifier predicts the class label for a new sample. The positive class labels predicted by each classifier are combined to determine the final multi class labels for this given data sample. Therefore, a data sample belongs to those class labels which are positively predicted by each classifier.

III. NUMERICAL EXPERIMENT

A. Dataset Description

The experiment has been performed on Scene dataset taken from Mulan's repository [41]. Scene dataset contains 2407 instances or images of six class labels-Beach, Urban, Mountain, Field, Sunset and Fall Foliage. The spatial color moments of each scene image are considered as features and each image is broken down into 49 blocks. The dataset has 294 attributes, 15 distinct class labels and the domain of scene dataset is image [11,41]. The detail description of scene dataset is given in table 1 as follows:

TABLE I. DATASET DETAILS

Class	Total
Beach	369
Urban	405
Field	327
Mountain	405
Fall Foliage	360
Sunset	364
Fall Foliage + Field	23
Fall Foliage + Mountain	13
Beach + Field	1
Beach + Mountain	38
Beach + Urban	19
Field + Mountain	75
Field + Urban	75
Mountain + Urban	1
Field + Mountain + Fall Foliage	1
Total	2407

The objective of the proposed model is to recognize the above mentioned scenes by learning from 2407 images.

B. Evalaution Metrics

The performance of multi-label classifier has been evaluated by using different metrics than those used in conventional unilabel classifier. For a given data sample x_i, its predicted class label set is denoted by P_i and the predicted rank corresponding to the class label is represented by $r_i(c)$. The top rank(1) is given to the most relevant class label and lowest rank (M) is assigned to the least relevant class label. In this study, we have evaluated the outcome of multi-label classification approaches by using example based and label based metrics. The example based metrics calculates the average differences between true and predicted class labels for each test data sample and then average over all data samples in the test dataset. In label based evaluation metrics, first the performance of each class label is measured separately and then averaged over all class labels. The seven example based metrics-Hamming Loss, Ranking Loss, Accuracy, Precision, Recall, F-measure and subset accuracy and eight label based metrics-Macro Accuracy, Macro Recall, Macro Precision, Macro F-measure, Micro Accuracy, Micro Recall, Micro Precision and Micro F-measure are used in this study for the performance evaluation. Hamming Loss is obtained as:

$$Hamming\ Loss = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i \Delta P_i|}{M} \quad (23)$$

Here, symbol Δ shows the symmetric difference of true and predicted class labels. It corresponds to the XOR Boolean logic operation. Hamming Loss lies between 0 and 1 and the result corresponds to 0 is considered as the best result.

Recall and Precision are defined as:

$$Recall = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i \cap P_i|}{|Y_i|} \quad (24)$$

$$Precision = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i \cap P_i|}{|P_i|} \quad (25)$$

F-measure is obtained by taking the harmonic mean of Recall and Precision as:

$$F\_Measure = \frac{1}{N}\sum_{i=1}^{N}\frac{2 \times Recall \times Precision}{Recall + Precision} = \frac{1}{N}\sum_{i=1}^{N}\frac{2 \times |Y_i \cap P_i|}{|Y_i| + |P_i|} \quad (26)$$

Subset accuracy measures the fraction of correctly recognized data samples and is defined as:

$$Subset\ Accuracy = \frac{1}{N}\sum_{i=1}^{N} I(Y_i = P_i) \quad (27)$$

Ranking Loss measures the fraction of reversely ordered class label pairs in which higher rank is assigned to the irrelevant class label and lower rank is given to the relevant one. Ranking Loss is formulated as:

$$Ranking\ Loss = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|Y_i||\bar{Y}_i|}|\{(c_a, c_b): r_i(c_a) \le r_i(c_b), (c_a, c_b) \in Y_i \times \bar{Y}_i\}| \quad (28)$$

where $\bar{Y}_i$ is the complementary class label set of Y_i with respect to the total label set L. Ranking Loss lies between 0 and 1 where 0 corresponds to the best result.

Accuracy can be defined as the number of predicted true class labels with regard to the total class labels. It is defined as:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i \cap P_i|}{|Y_i \cup P_i|} \quad (29)$$

It lies between 0 and 1 and the classifier with 1 accuracy score is considered as the best performing classifier.

Macro Accuracy, Macro Precision, Macro Recall and Macro F-Measure are defined as:

$$Macro\ Accuracy = \frac{1}{|L|}\sum_{i=1}^{|L|}\frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i} \quad (30)$$

$$Macro\ Precision = \frac{1}{|L|}\sum_{i=1}^{|L|}\frac{TP_i}{TP_i + FP_i} \quad (31)$$

$$Macro\ Recall = \frac{1}{|L|}\sum_{i=1}^{|L|}\frac{TP_i}{TP_i + FN_i} \quad (32)$$

$$Macro\ F - Measure = \frac{1}{|L|}\sum_{i=1}^{|L|}\frac{2 \times Recall_i \times Precision_i}{Recall_i + Precision_i} \quad (33)$$

Micro Accuracy, Micro Precision, Micro Recall and Micro F-Measure are formulated as:

$$Micro\ Accuracy = \frac{\sum_{i=1}^{|L|} TP_i + \sum_{i=1}^{|L|} TN_i}{\sum_{i=1}^{|L|} TP_i + \sum_{i=1}^{|L|} TN_i + \sum_{i=1}^{|L|} FP_i + \sum_{i=1}^{|L|} FN_i} \quad (34)$$

$$Micro\ Precision = \frac{\sum_{i=1}^{|L|} TP_i}{\sum_{i=1}^{|L|} TP_i + \sum_{i=1}^{|L|} FP_i} \quad (35)$$

$$Micro\ Recall = \frac{\sum_{i=1}^{|L|} TP_i}{\sum_{i=1}^{|L|} TP_i + \sum_{i=1}^{|L|} FN_i} \quad (36)$$

$$Micro\ F - Measure = \frac{2 \times Micro\ Precison \times Micro\ Recall}{Micro\ Precision + Micro\ Recall} \quad (37)$$

C. Results and Discussion

This study compares the performance of Binary Relevance one-versus-rest LSTSVM based scene classification system with six existing multi-label classification approaches-Binary Relevance Support Vector Machine, ML-kNN, ML C4.5, RAkEL, CC, QWML. SVM is used as a base classifier for RAkEL, CC and QWML. The performance of proposed BR-LSTSVM scene classification model is tested against fifteen evaluation metrics using ten-fold cross validation approach.

TABLE II. PERFORMANCE COMPARISON FOR SCENE DATASET

Evaluation Metrics	ML-kNN	ML C4.5	RAkEL	QWML	CC	BR-SVM	BR One-versus-rest LSTSVM
Hamming Loss $\downarrow$	0.099	0.141	**0.079**	0.087	0.084	0.082	0.110
Recall $\uparrow$	0.655	0.582	0.740	0.709	0.726	0.711	**0.746**
Precision $\uparrow$	0.661	0.620	**0.768**	0.711	0.758	0.718	0.712
F-Measure $\uparrow$	0.658	0.587	**0.754**	0.710	0.742	0.714	0.729
Subset Accuracy $\uparrow$	0.573	0.533	**0.694**	0.630	0.685	0.639	0.644
Ranking Loss $\downarrow$	0.093	0.169	0.104	0.103	0.064	0.068	**0.063**
Accuracy $\uparrow$	0.629	0.569	0.734	0.683	0.720	0.689	**0.726**
Macro Accuracy $\uparrow$	0.702	0.624	0.758	0.763	0.745	0.721	**0.887**
Macro Precision $\uparrow$	0.784	0.635	0.835	0.832	0.817	0.844	**0.964**
Macro Recall $\uparrow$	0.647	0.573	0.727	0.701	0.716	0.703	**0.904**
Macro F-Measure $\uparrow$	0.692	0.596	0.777	0.759	0.762	0.765	**0.932**
Micro Accuracy $\uparrow$	0.684	0.605	0.754	0.760	0.733	0.720	**0.889**
Micro Precision $\uparrow$	0.691	0.619	0.831	0.832	0.814	0.843	**0.964**
Micro Recall $\uparrow$	0.634	0.570	0.721	0.692	0.708	0.694	**0.907**
Micro F-Measure $\uparrow$	0.661	0.593	0.772	0.756	0.757	0.769	**0.934**

All these multi-label classifiers are implemented in Matlab 2012a on a Personal Computer. Tradeoff constants are selected using Grid Search approach form the range of $c_{+1j}, c_{-1j} \in \{10^{-6}, \ldots, 10^{2}\}$. Table 2 indicates the result analysis of the proposed scene classification model with the existing multi-label classification approaches. In table 2, the downward symbol attached with the evaluation metrics denotes that the score of corresponding metric should be as less as possible i.e., a multi-label classifier is said to be the best performing classifier if it scores low value for that metric. However, the score of the metric corresponding to the upward symbol should be as high as possible. The result of best performing multi-label classifier for each evaluation metric is shown by the bold value. From table 2, it is observed that the RAkEL performs better on four evaluation metrics such as Hamming Loss, Precision, F-Measure and Subset Accuracy while the proposed approach shows better performance on the remaining eleven metrics as compared to the existing six multi-label classification approaches. Therefore, it can be concluded that the proposed Binary Relevance one-versus-rest LSTSVM is an effective model of multi-label scene classification.

IV. CONCLUSION

This study considers the task of scene classification from images as multi-label classification problem and constructed a multi-label scene classification model. Binary Relevance based one-versus-rest Least Squares Twin Support is used to recognize the scenes from given images. The result of experiment demonstrates the suitability of the proposed model. The proposed model shows better performance in terms of eleven evaluation metrics in comparison with the other six existing multi-label classification algorithms. The scene dataset contains large number of features, so in future it would be interesting to consider the feature selection approaches for the selection of significant features.

REFERENCES

[1] L.J. Li, H. Su, Y. Lim, and L. F.Fei. "Objects as attributes for scene classification." In Trends and Topics in Computer Vision, pp. 57-69. Springer Berlin Heidelberg, 2012.

[2] O. Maron and A. L. Ratan. "Multiple-Instance Learning for Natural Scene Classification." In ICML, vol. 98, pp. 341-349. 1998.

[3] C. Carson, H. Greenspan and J. Malik. Recognition of images in large databases using a learning framework. University of California, Berkeley, Computer Science Division, 1997.

[4] M. Bautell, J. Luo and R. T. Gray. "Sunset scene classification using simulated image recomposition." In Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on, vol. 1, pp. I-37. IEEE, 2003.

[5] M. R. Boutell, J. Luo, and C. M. Brown. "Scene parsing using region-based generative models." Multimedia, IEEE Transactions on 9, no. 1, pp.136-146, 2007.

[6] J. Luo, A. Singhal, and W. Zhu. "Natural object detection in outdoor scenes based on probabilistic spatial context models." In Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on, vol. 2, pp. II-457. IEEE, 2003.

[7] Y. Song, and A. Zhang. "Analyzing scenery images by monotonic tree." Multimedia Systems 8, no. 6, pp. 495-511, 2003.

[8] BVV Sri Raj Dutt, P. Agrawal, and S. Nayak. "Scene Classification in Images."

[9] M. Boutell, A. Choudhury, J. Luo, and C. M. Brown. "Using semantic features for scene classification: How good do they need to be?." In Multimedia and Expo, 2006 IEEE International Conference on, pp. 785-788. IEEE, 2006.

[10] A. Bosch, A. Zisserman, and X. Muñoz. "Scene classification via pLSA." In Computer Vision–ECCV 2006, pp. 517-530. Springer Berlin Heidelberg, 2006.

[11] M. R. Boutell,J. Luo, X. Shen, and C. M. Brown. "Learning multi-label scene classification." Pattern recognition 37, no. 9, pp.1757-1771, 2004.

[12] G. Tsoumakas, I. Katakis, and I. Vlahavas. "Mining multi-label data." In Data mining and knowledge discovery handbook, pp. 667-685. Springer US, 2010.

[13] W. Bi and J. Kwok. "Efficient Multi-label Classification with Many Labels." In Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 405-413. 2013.

[14] G. Tsoumakas and I. Katakis. "Multi-label classification: An overview." International Journal of Data Warehousing and Mining (IJDWM) 3, no. 3, pp. 1-13, 2007.

[15] T.Y.Wang, and H.M. Chiang. "Solving multi-label text categorization problem using support vector machine approach with membership function."Neurocomputing 74, no. 17, pp.3682-3689, 2011.

[16] J. Y. Jiang, S. C. Tsai, and S.J. Lee. "FSKNN: multi-label text categorization based on fuzzy similarity and k nearest neighbors." Expert Systems with Applications 39, no. 3, pp. 2813-2821, 2012.

[17] B. Yang, J.T. Sun, Tengjiao Wang, and Zheng Chen. "Effective multi-label active learning for text classification." In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 917-926. ACM, 2009.

[18] D. Zufferey, T. Hofer, J. Hennebert, M. Schumacher, R. Ingold, and S. Bromuri. "Performance comparison of multi-label learning algorithms on clinical data for chronic diseases." Computers in Biology and Medicine 65, pp. 34-43, 2015.

[19] J. Tang, Z. J.Zha, D. Tao, and T. S. Chua. "Semantic-gap-oriented active learning for multilabel image annotation." Image Processing, IEEE Transactions on 21, no. 4, pp. 2354-2360, 2012.

[20] A. Fakhari, and A. M. E. Moghadam. "Combination of classification and regression in decision tree for multi-labeling image annotation and retrieval." Applied Soft Computing 13, no. 2, pp.1292-1302, 2013.

[21] Y. Pang, Z. Ma, Y. Yuan, X. Li, and K. Wang. "Multimodal learning for multi-label image classification." In Image Processing (ICIP), 2011 18th IEEE International Conference on, pp. 1797-1800. IEEE, 2011.

[22] M.S. Liu, and J.H. Chen. "A multi-label classification based approach for sentiment classification." Expert Systems with Applications 42, no. 3, pp.1083-1093, 2015.

[23] A. C. ES Lima, and L. N. D. Castro. "A multi-label, semi-supervised classification approach applied to personality prediction in social media." Neural Networks 58, pp.122-130, 2014.

[24] D. Tomar and S. Agarwal, "Multi-label classifier for Emotion Recognition from Music", In third International Conference on Advanced Computing, Networking and Informatics, Springer International Publishing, 2015. In press.

[25] M. S. Sorower, "A literature survey on algorithms for multi-label learning." Oregon State University, Corvallis, 2010.

[26] M.L. Zhang and Z.H. Zhou. "A review on multi-label learning algorithms." Knowledge and Data Engineering, IEEE Transactions on 26, no. 8, pp.1819-1837, 2014.

[27] O. Luaces, J. Díez, J. Barranquero, J. José del Coz, and A. Bahamonde. "Binary relevance efficacy for multilabel classification." Progress in Artificial Intelligence 1, no. 4, pp. 303-313, 2012.

[28] E. Spyromitros, G. Tsoumakas and I. Vlahavas. "An empirical study of lazy multilabel classification algorithms." In Artificial Intelligence: Theories, Models and Applications, pp. 401-406. Springer Berlin Heidelberg, 2008.

[29] E. A. Tanaka, S. R. Nozawa, A. A. Macedo, and J. A. Baranauskas. "A multi-label approach using binary relevance and decision trees applied to functional genomics." Journal of biomedical informatics 54, pp. 85-95, 2015.

[30] J. Read, B. Pfahringer, G. Holmes and E. Frank. "Classifier chains for multi-label classification." Machine learning 85, no. 3, pp. 333-359, 2011.

[31] P. Prajapati, A. Thakkar and A. Ganatra. "A Survey and Current Research Challenges in Multi-Label Classification Methods." International Journal of Soft Computing 2 (2012).

[32] J. Fürnkranz, E. Hüllermeier, E. L. Mencía, and K. Brinker. "Multilabel classification via calibrated label ranking." Machine Learning 73, no. 2 pp. 133-153, 2008.

[33] E.L. Mencía, S.H. Park, and J. Fürnkranz. "Efficient voting prediction for pairwise multilabel classification." Neurocomputing 73, no. 7 , pp. 1164-1176, 2010.

[34] M.L. Zhang and Z.H. Zhou. "ML-KNN: A lazy learning approach to multi-label learning." Pattern recognition 40, no. 7, pp. 2038-2048, 2007.

[35] M. A. Kumar and M. Gopal, Least squares twin support vector machines for pattern classification, Expert Systems with Applications 36, pp.7535–7543, 2009.

[36] D. Tomar and S. Agarwal. "Twin support vector machine: a review from 2007 to 2014." Egyptian Informatics Journal 16, no. 1, pp. 55-69, 2015.

[37] Jayadeva, R. Khemchandani and S. Chandra, Twin Support vector Machine for pattern classification. IEEE Transaction Pattern Analysis and Machine Intelligence 29(5), pp. 905-910, 2007.

[38] C. Cortes and V. Vapnik, Support Vector Network. Machine Learning 20, no.3, pp.273-297, 1995.

[39] V. Vapnik, The nature of statistical Learning, @nd edn. Springer, New York, 1998.

[40] D. Tomar and S. Agarwal. "A comparison on multi-class classification methods based on least squares twin support vector machine." Knowledge-Based Systems 81, pp.131-147, 2015.

[41] Scene Dataset. http://mulan.sourceforge.net/datasets-mlc.html. Accessed on 20th March, 2015.