enSVM: A classification framework using ensemble of SVMs

Vikas Sharma[1], Dibyajyoti Chutia[1], Diganta Baruah[2], Jonali Goswami[1], Udayan Baruah[2] and P.L.N. Raju[1]
[1]North Eastern Space Applications Centre, Department of Space, Government of India, Umiam, Shillong, Meghalaya - 793103, India
[2]Information Technology Department, Sikkim Manipal Institute of Technology, Majhitar, East Sikkim -737136, India
Email: d.chutia@nesac.gov.in

(Received: Oct 15 2018; in final form: Oct 25, 2018)

Abstract: Empirically, Support Vector Machines (SVMs) have been reported with higher performance on many benchmark datasets including remotely sensed (RS) data. In SVMs, it is prerequisite to obtain a large decision boundary by representing the data points in a n-dimensional feature space using kernel methods. SVMs work well with a minimal number of training samples when appropriate kernels are used to optimize the hyperplane. In addition, SVMs are being used to build ensemble classification methods. The capability of SVM for constructing a set of the diverse base classifier is not yet fully exploited for classification of RS data. The main objective of this work is to develop an ensemble of SVMs (enSVM) with enhanced predictive ability with minimal number of base classifiers. The proposed enSVM is a collection of SVMs as base classifier where each of the base classifiers is assigned independently a random subspace of features to build the predictive model. Each individual classifier cast a unique vote and the final classification is based on the majority of votes of all the base classifiers. An investigation was carried out on the two sets of satellite data of QuickBird and Landsat Enhanced Thematic Mapper Plus (Landsat ETM+) sensors pertaining to different landscape with different land cover classes. It was observed that the enSVM outperformed the SVM, Maximum Likelihood Classifier (MLC), Multi Layer Perceptron (MLP) and achieved comparable results with the most powerful random forest (RF) classifier.

Keywords: Classification, Support Vector Machines, Ensemble, Random subspace, enSVM

1. Introduction

Remote sensing (RS) community paying confident attention to ensemble methods as they have shown significant potential to classify heterogeneous, high dimensional, noisy, missing, and complex RS datasets (Opitz and Maclin, 1999; Han et al., 2012; Huang et al., 2013; Chutia et al., 2015; Chutia et al., 2017; Rawat et al., 2018). The focal intuition of ensemble principle is to create multiple hypotheses generated by weak learners to achieve higher classification accuracy by combining or aggregating their predictions (Opitz and Maclin, 1999). Diversity in multiple hypotheses is prerequisite for ensemble methods. Empirically, ensemble methods tend to produce better results when there is a significant diversity among the models. Diversity can be achieved through the fusion of different classifiers, randomization of training data, and randomization in feature space etc. (Hansen and Salamon, 1990; Kuncheva and Whitaker, 2003; Brown et al., 2004).

Bagging (Breiman, 1996) and Boosting (Freund and Schapire, 1996) are more commonly used ensembles methods in machine learning applications. Bagging being a variant of meta classifier combines the predictions of base classifiers for improving the unstable estimation and the predictive accuracy of the classifiers. It uses the bootstrap technique for randomization of the training data. Bagging creates the small random bags with the replacement of training dataset from the original datasets in order to create the diversity among the base classifiers. It is very popular and more frequently used in decision tree algorithms (DT, Quinlan, 1986) as it avoids the overfitting and reduces the variance [Chan et al., 2001]. However, the boosting algorithm reduces the bias with the variance to achieve higher classification accuracy. It learns from the errors and assigns the corresponding weights to each weak base classifier in order to get better predictive accuracy. In boosting, AdaBoost (Freund and Schapire, 1996) is the most popular and successful version that follows the weight adjustment procedure to classify a novel instance. In every iteration, Adaboost increase the weights on the misclassified instance and decreases the weights on correctly classified instances in order to give a chance to misclassified or unclassified data (Freund and Schapire, 1996; Chan et al., 2001). The main drawback of bagging model is that they suffer from the overfitting while dealing with the noisy data. On the other hand, the large number of outliers (highest weight instance) can reduce the performance of AdaBoost (Freund and Schapire, 1999). For overcoming overfitting problem Breiman proposed random forest (RF) algorithm in 2001 (Breiman, 2001). Currently, the RF has been successfully implemented in various applications (Casanova et al., 2014; Yang et al., 2008; Goldstein et al., 2010; Sylvester et al., 2017). The RF is based on the principal of bagging, where a number of DT are independently constructed as a base classifier through random sampling of the training dataset. The major concern in RF is the selection of appropriate size of the base classifiers; number of random subset of trees. In the current scenario, rotation forest (Rodriguez et al., 2006) is getting popular and found comparable with RF in many instances (Peijun et al., 2015; Xia et al., 2014; Rawat et al., 2018). Unlike RF, rotation forest splits the features into several disjoint subsets and apply the data transformation on each subset using principal component analysis (PCA, Jolliffe, 1986). In the second stage, new training dataset for the DT is formed by concatenating the linear extracted features contained in each subset. Thus, rotation forest enhances the classification accuracy with diversity within the ensemble (Rodriguez et al., 2006; Du et al., 2015; Kuncheva and Rodriguez, 2007; Liu and Huang, 2008). Comparatively, rotation forest is a

© Indian Society of Geomatics

complex model but it can provide excellent performance. However, RF is most stable and can provide consistently higher predictive accuracy with the noisy data and low computation cost.

Recently, the Random committee is found as an important platform for combining the base classifiers. Random committee constructs an ensemble of base classifiers and averages their results (Witten and Frank, 2005). The results of all base classifiers are based on the same data, but they are initialized by the different random number seed (Lira et al., 2007; Tatsis et al., 2013). On the other hand, random subspace creates diversity in the training data through feature space. The final prediction is the majority of votes of individual predictor. In recent times, an ensemble of traditional classifiers is in trend in various fields (Rahman et al. 2016; Kenduiywo et al., 2017; Liu et al., 2016; Yu et al., 2015; Sharma et al., 2018; Chutia et al., 2014; Vyškovský et al., 2016; Han et al, 2012; Lv et al., 2017). SVMs now effectively used as a base classifier in ensemble methods (Claesen et al., 2014; Sørensen and Nielsen, 2018; Wandekoken et al., 2011; Lo et al., 2015). Ensemble classifiers with DTs algorithm are mostly preferred as their capability already established on many benchmark data. However, appropriateness of traditional classifier in building ensemble approach needs attention.

1.2 Motivation

Nonparametric classifiers like SVMs are highly productive. But, the performance of SVMs are highly influenced by the selection of kernel parameters. A number of ensemble methods as illustrated in the earlier section have been proposed where DT classifiers have been used as base classifiers in most of the instances. However, the potential of SVMs with suitable kernel parameter and appropriate ensemble methods for erecting a set of the diverse hypothesis is not yet fully explored for RS data classification.

1.3 Objectives

The primary objective of this work is to develop an efficient and effective classification framework using ensemble technique for classification of RS imagery with the following contributions:

1) Selection of appropriate kernel parameters for SVM classifier;
2) To develop an ensemble of SVMs using random subspace method (enSVM);
3) To assess the performance of the enSVM in comparison with the MLP, MLC, RF and SVM.

Rest of the research article summarized as follows: the Section II gives the detailed information about the characteristics of datasets used in the investigation. The framework of the proposed enSVM is illustrated in the Section III. The analysis on the experimental results discussed in the Section IV followed by the concluding remark in the Section V.

2. Dataset used

To assess the predictive ability of the proposed framework, the investigation was carried out on the moderate resolution Landsat ETM+ (Test site-I) and high resolution QuickBird (Test site-II) multispectral sensors datasets. The more details about the respective satellite sensors are given in the Table 1 followed by a description on the test sites.

Table 1: Detailed information of both test sites

Particulars	Landsat ETM+	QuickBird
Satellite Name	Landsat 7	Digital Globe
Spectral Resolution	0.45µm - 0.90µm	0.45µm - 0.90 µm
Number of bands	04	04
Temporal Resolution	16 days	1-3.5 days
Spatial Resolution	30 meter	0.65 meter
Swath	183 km	16.4 Km

2.1 Test site-I

The Test site-I is pertaining to Sontipur area of Assam, India comprised of five major classes including the river Brahmaputra. The site is topographically plain terrain dominated by agricultural crop land followed by forest tree clad area, scrub forest, and sand nearby river Brahmaputra. Details of the classes with respective train and test samples for Test site-I is given in Table 2.

Table 2: Details of classes with train and test samples for test site-I

Class Name	Train sample	Test sample
River/Waterbody-Perennial	3165	1513
Forest tree clad area	259	1119
Agriculture cropland	870	1114
Sand	225	1809
Scrub Forest	118	1502
Total number of samples	4637	7057

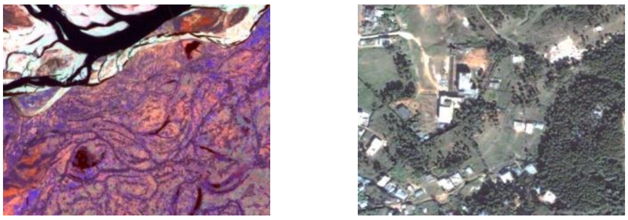
2.2 Test site-II

The Test site-II is anurban area of Shillong, capital of Meghalaya state of India representing the five major classes. The site-II is topographically hilly terrain and surrounded by pine trees followed by urban residential areas, open scrub, open surfaces, and shadows. Further information about the classes and their respective train and test samples can be found in the Table 3.

Table 3: Details of classes with train and test samples for test site–II

Class Name	Train sample	Test sample
Urban residential areas	980	1485
Pine trees	845	1103
Shadows	975	958
Open scrub	909	756
Open surfaces	1078	1312
Total number of samples	4787	5614

The satellite image of the Test site-I and Test site-II are shown in a and b respectively of Figure 1.



a) Test site-I (Landsat ETM+) **b)** Test site-II (QuickBird)

Figure 1: Satellite images of test sites

3. Methodology

The proposed enSVMis based on the principle of bagging, where each of the base classifier is initialized independently with a randomly selected subset of features. A set of SVMs are used as base classifier and each of the individual SVM casts a vote based on the decision boundary of hyperplane defined by the random subspace of the features. The final classification result is the majority of voting of all the SVM base classifiers.

The entire approach is comprised of the three major components- i) selection of kernel parameter for SVM, ii) building of enSVM framework using random subspace and iii) assessment of enSVM using a set of accuracy assessment parameters.

3.1 Selection of kernel parameters for SVMs

SVM is a nonparametric classifier that uses the hyperplane to separate the data into the predefined classes. It tries to find a hyperplane with the help of support vectors and creates the decision boundary with a maximum distance between two classes.

For an instance, D is the set of all input data, and X_n is the input space with target class $Y_n$then-

$$D(x) = Sign\,(\textstyle\sum_{i}^{n}.\alpha_i y_i (x_i . x) + b) \quad (1)$$

SVM uses the kernel function to optimize the hyperplane for classification of multiclass data. Selection of kernel parameters is one of the important aspectsin SVMs. Kernel functions are being used to reduce the time complexity of SVMs by using the inner product of two transformed data vectors in the feature space (Cortes and Vapnik, 1996; Sharma et al., 2016).
Popular kernels those can be used with SVMs are given below.

Linear kernel :$K(x_i, x_j) = {x_i}^T . x_j$ (2)

Polynomial kernel :$K(x_i, x_j) = (r + \gamma {x_i}^T . x_j)^d, \gamma > 0$ (3)

RBF kernel :$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$ (4)

Sigmoid kernel :$K(x_i, x_j) = tan\,h\,(\gamma {x_i}^T . x_j + r)$(5)

where, x_i = support vector of length m, T = Transformation, $\gamma$is the gamma function, d = degree of polynomial, and$r$ = bias. The more information about the SVM can be found in Cortes &Vapnik 1996 (Cortes and Vapnik, 1996).

A number of studies suggested that the polynomial kernel can achieve better predictive accuracy than the other kernels for classification of satellite datasets (Sharma et al. 2016; Kumar et al., 2018; Akbari et al., 2012). An investigation was carried out to assess the comparative performance of SVM classifier with all four kernels on both the datasets. It was observed that SVM with the polynomial kernel outperformed the SVM with other kernels with Kappa Index Analysis (KIA) =0.77 and Overall Accuracy (OA) =82.38% for Test site - I dataset and KIA=0.86 and OA=88.64% for Test site - II [Table 4]. Based on the experimental results as specified in Table 4 it is proposed to use SVM with the polynomial kernel as a base classifier for building the enSVM.

Table 4: Comparative assessment of kernels

Kernels	Test site-I			Test site-II		
	OA (%)	KIA	T_r (sec)	OA (%)	KIA	T_r (sec)
Linear	81.46	0.76	0.52	87.27	0.82	7.71
Polynomial	82.38	0.77	0.14	88.64	0.86	1.48
RBF	78.71	0.70	0.50	86.58	0.83	8.46
Sigmoid	76.19	0.67	1.30	85.34	0.82	22.77

3.2 Building enSVM using random subspace

Random subspace creates the diversity in the features space by creating random subspace of features for each base classifier in order to achieve higher predictive accuracy (Ho, 1998). In enSVM model, training dataset has been feed to the random subspace algorithm where a set of training dataset defined by a set of random features was created based on the size of the enSVM. Then each of the randomly generated training datasets is provided to each of the SVM.

Let, p_K is a collection of SVM classifiers, enSVM $\rightarrow \{p(x, s_k^r), k = 1, \ldots, K\}$ of sizeK, where $p(x, s_k^r)$is each individual base classifier trained by a random subspace of feature $s_k^r \in F^R$drawn with replacement.$F^R$is the original set of feature sets of the training datasets T with R number of features and r is the number of features in s_k^r where $r < R$ and $s_{k=i}^r \neq s_{k=j}^r\ \forall$ k. Each of the SVM base classifiers is represented as $p(x, s_k^r)$ with s_k^rrandomly selected predictors. Each individual $p(x, s_k^r)$ casts a vote for an unknown input x independently. Let $\hat{C}_{s_k^r}(x)$ is class prediction of each $p(x, s_k^r)$, the final classification of input x is the majority of the voting of all the $p(x, s_k^r)$ classifiers i.e.,$\hat{C}_{s_k^r}(x) = \text{majority vote}\,\{\hat{C}_{s_k^r}(x)\}_1^K$.More detailed information about the random subspace method can be found in (Ho, 1998).

4. Results and Discussions

All experiments were executed in the High Performance Cluster Computing (HPC) environment with 20 core Intel (R) Xeon (R) CPU E5-2680 V2 processor and 48 GB

RAM system configuration. The training samples of Test site - I associated with 80 features whereas 142 features were associated with Test site - II dataset. Random subspace (space=0.50) was used to create diversity among the base classifiers by creating a set of training datasets with a random subspace of optimal features.

A set of parameters such as KIA, Receiver Operating Characteristic (ROC), and Time Complexity (T_r) have been identified for performance analysis. A detailed discussion on the experimental observations are summarized in the context of the following:

1) The optimal size of enSVM;
2) Performance of enSVM against training size;
3) Comparison with other counterparts like MLC, MLP, SVM, and RF.

4.1 Optimal size of enSVM

The size of enSVM (K) or the optimal number of the base classifier is an important parameter of the enSVM. The overall performance of the predictive model is highly affected by the size of the K. The enSVM was executed with K=1 with an increment of 1 till K=30 or it achieved the highest accuracy. Afterwards, the enSVM was executed with K=10 with an increment counter of 10 maximum up to 200 or till it achieved the highest accuracy. It was observed that the enSVM achieved the highest accuracy at K=7 for Test site - I and at K=16 for Test site – II (Figure 2). The red color depicts the experimental observation of the OA on Test site - I against the variable size of K; whereas, the OA of enSVM for the Test site - II is represented by the blue colour. For Test site – I, enSVM was able to give highest accuracy with the 7 base classifier. The performance of enSVM was found higher and consistent on the high spatial resolution satellite dataset (Test site - II) as compared to the moderate spatial resolution dataset (Test site - I). It was also found that the larger size of K cannot ensure the higher performance rather the model can achieve better performance with a minimal size of K. The challenge was to find the optimal value for K. The following two observations are highlighted below:

1) The performance of enSVM is not directly proportional to the higher value of K. However, larger size of K can cause higher computational expenses.
2) During the execution of enSVM, each of the individual SVMs was initialized with a random subspace of feature. The diversity of model is not determined by the size of K rather it depends on the degree of randomization of the feature assign to each SVM.

4.2 Performance of enSVM against training size

SVM has been found effective classification option where the size of the training dataset is limited. It is required to assess the impact of the training data size on the performance of enSVM. The performance of enSVM against different size of the training dataset are given in the Table 5. It was observed that the performance of the enSVM is increased when the training size is large, because the same dataset is splitted into the train and test datasets. For example, enSVM achieved OA=88.58% with KIA=0.85 for Test site - I and OA= 96.22% with KIA=0.94 for Test site - II respectively when the model was trained by 10% of the dataset and tested with the remaining 90%. Similarly, it achieved the highest performance with OA=93.11 for Test site-I and OA=99.19% with KIA=0.98 for Test site-II when the training and test ratio was 90:10. Even though, the performance of enSVM with less training dataset yields slightly lesser accuracy as compared to the larger size of training dataset but this can be accepted when the training dataset is very limited. The remaining part of experiment was carried out with the independently generated train and test datasets as mentioned in the Section - II. There are two major observations can be made here:

1) Like SVM, enSVM can also perform satisfactory when there is a lack of sufficient training dataset.
2) The evaluation of the actual predictive ability the classifier should be trained and tested with the independently generated training and test datasets.

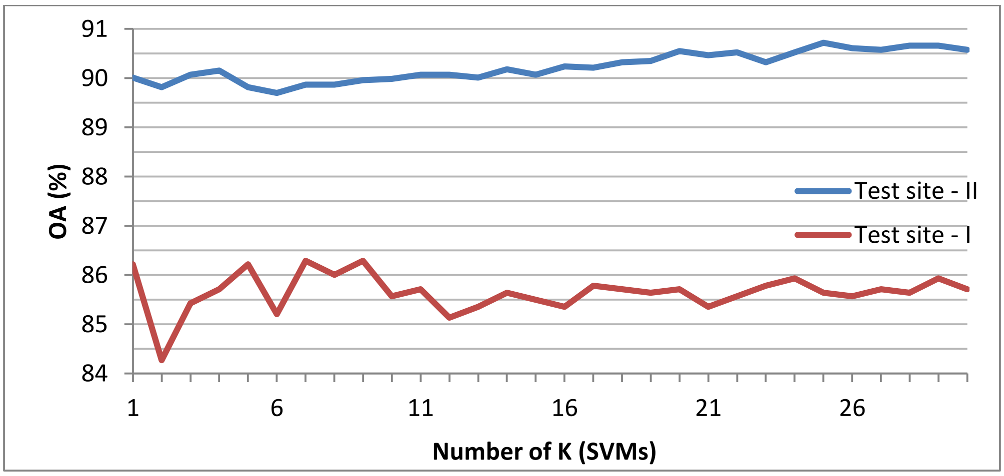


Figure 2: Overall Accuracy (OA) versus number of K (SVMs)

Table 5: Overall Accuracy (OA) versus size of training data

Datasets	Accuracy Parameters	Training data size (%)								
		10	20	30	40	50	60	70	80	90
Test site - I	OA(%)	88.58	90.83	91.79	92.06	92.77	93.16	93.59	93.83	93.11
	KIA	0.85	0.88	0.89	0.90	0.90	0.914	0.91	0.92	0.91
Test site - II	OA(%)	96.22	98.08	98.26	98.52	98.87	98.78	98.92	99.19	99.19
	KIA	0.94	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98

Table 6: Comparative assessment of classifier

Classifiers	Test site - I				Test site - II			
	OA	KIA	ROC	t (sec)	OA	KIA	ROC	t (sec)
SVM	82.38	0.77	0.91	0.13	88.64	0.86	0.96	1.39
enSVM	85.50	0.81	0.95	1.26	91.99	0.89	0.98	11.0
MLC	81.70	0.77	0.91	0.02	85.57	0.81	0.97	0.05
MLP	83.30	0.78	0.93	26.5	89.99	0.87	0.98	13.1
RF	84.34	0.79	0.94	0.03	90.24	0.88	0.98	0.10

4.3 Comparison with other counterparts i.e., MLC, MLP, SVM, and RF

The enSVM was executed with the optimal number of K i.e. 7 for Test site-I and 16for Test site - II as illustrated in the earlier section. A set of other counterparts of enSVM such as SVM, MLC, MLP, and RF have been utilized for classification of both the datasets. The comparative performance of all the classifiers are given in the Table6.The final classified outputs of enSVM on both the datasets sites are depicted in Figure 3 and Figure 4. It was found that the enSVM outperformed all the classifiers for both the datasets sites with KIA =0.81 for Test site - I and K=0.89 for Test site - II [Table 6]. The performance of enSVM was significantly enhanced than a single SVM classifier (K =0.77 for Test site - I and K=0.86for Test site - II) (Table 4). In addition, RF was found comparatively effective; it can create more diversity among the base classifiers through bagging and random selection of the best feature at each node. Comparatively enSVM suffers from little computational expenses due to the size of the ensemble (Table 6). The performance of MLP is higher than the MLC; however it causes large computational expenses during the training phase. The ensemble approach can provide higher performance with the traditional powerful classifiers like SVM. However, enSVM can create comparatively less diversity than RF. It will be more sensitive when there is little noise in the data and a random subspace method fails to assign the relevant subspace of features to each individual base classifier. However, if appropriate kernel parameter is selected for SVM, enSVM can performance than other powerful classifier like RF.

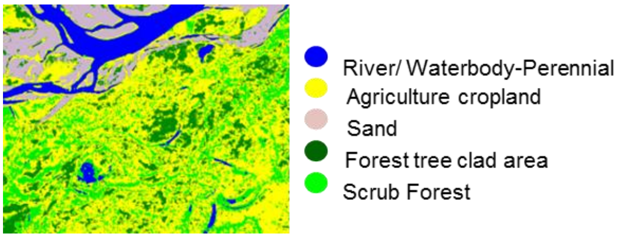


Figure 3:Classified image of test site-I with legends

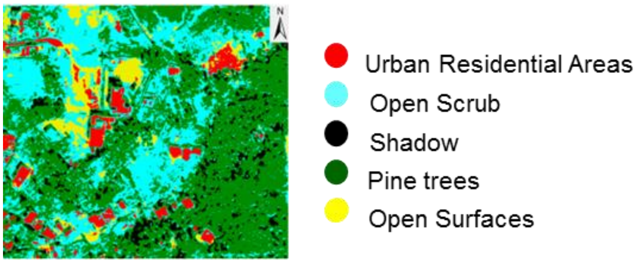


Figure 4: Classified image of test site-II with legends

5. Conclusion

The proposed enSVM is based on the principle of bagging of training dataset in feature dimension where each individual SVM classifier decides the final classification based on the majority of their votes. The performance depends on the random subspace of relevant feature set assigned to each SVM classifier. The observations achieved during the investigation were found quite encouraging. However, random subspace of feature could be defined from the optimal set of relevant features using a feature selection technique in order to enhance the predictive ability as well as the computational complexity.

Acknowledgment

The authors would like to thanks North Eastern Space Applications Centre (NESAC), Department of Space, Government of India, Shillong, Meghalaya for providing the scientific environment, necessary data, High-Performance Cluster Computing (HPC) facility with necessary infrastructure during this research study.

References

Breiman, L. (1996). Bagging Predictors, Machine Learning, 24(2), 123-140 https://doi.org/10.1023/A:1018054314350

Breiman, L. (2001). Machine Learning, 45(1), 5-32 https://doi.org/10.1023/A:1010933404324.

Brown, G., J. Wyatt, R. Harris.andX. Yao. (2004). Diversity creation methods: a survey and categorisation, Information Fusion, 6(1), pp.5-20. https://doi.org/10.1016/j.inffus.2004.04.004.

Casanova, R., S. Saldana, E.Y. Chew, R.P. Danis, C.M. Greven and W.T. Ambrosius. (2014). Application of random forests methods to diabetic retinopathy classification analyses, PLOS ONE, 9(6), pp. 1-8.https://doi.org/10.1371/journal.pone.0098587.

Chan, J., C. Huang and R. Defrie. (2001). Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting,IEEE Transactions on Geoscience and Remote Sensing, 39(3), 663-695.doi: 10.1109/36.91112.

Chutia, D.,D.K. Bhattacharyya, R.Kalita, J. Goswami, S.S. Puyam and S. Sudhakar (2014). A model on achieving higher performance in the classification of hyperspectral satellite data: A case study on Hyperion data, Applied Geomatics, 6. doi:10.1007/s12518-014-0134-z.

Chutia, D., D.K. Bhattacharyya, K.K. Sarma, R. Kalita and S. Sudhakar. (2015). Hyperspectral remote sensing classifications: A perspective survey, Transactions in GIS, 20(4), 463-490. doi:10.1111/tgis.12164.

Chutia, D., D.K. Bhattacharyya, J. Sarma and P.L.N. Raju. (2017). An effective ensemble classification framework using random forests and correlation basedfeature selection technique, Transaction in GIS, 21(6), 1165-1178.https://doi.org/10.1111/tgis.12268.

Claesen, M., F.D. Smet, J.A.K. Suykens and B.D Moor. (2014). EnsembleSVM: A library for ensemble learning using support vector machines, Journal of Machine Learning Research, 14, 141-145.

Cortes, C and V. Vapnik. (1996). Support-Vector Networks, Machine Learning, 20(3),273–297. doi:10.1007.BF00994018.

Du, P., A. Samat, B. Waske, S. Liu and Z. Li. (2015). Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features, ISPRS Journal of Photogrammetry and Remote Sensing, 105, 38-53. doi.10.1016/j.isprsjprs.2015.03.002.

Freund, Y and R.E. Schapire. (1996). Experiments with a new boosting algorithm. Proceedings 13th International Conference on Machine Learning, pp. 148–156.

Freund, Y. and. R.E.Schapire. (1999). A Short introduction to boosting, Journal of Japanese Society for Artificial Intelligence, 14(5), pp.771-780.

Goldstein, BA., A.E. Hubbard, A. Cutle and L.F. Barcellos. (2010). An application of Random Forests to a genome-wide association dataset: Methodological considerations & new findings. 49(11), 1471-2156. https://doi.org/10.1186/1471-2156-11-49.

Han, M., X. Zhu and W. Yao. (2012). Remote sensing image classification based on neural network ensemble algorithm, Neurocomputing,.78, 133-138.

Hansen, L. K and P. Salamon. (1990). Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(10), 993-1001. doi: 10.1109/34.58871.

Ho, T.K. (1998). The random subspace method for constructing decision forests, IEEE Transaction on Pattern Analysis and Machine Intelligence, 20(8), 832–844.

Jolliffe, I.T. (1986). Principal Component Analysis. Ed. Springer-Verlag, New Work

Kenduiywo, BK., D. Bargiel and U. Soergel. (2017). Higher order dynamic conditional random fields Ensemble for crop type classification in radar images, in IEEE Transactions on Geoscience and Remote Sensing, 55(8),4638-4654.

Kumar, R.A., D. Chutia, J. Goswami, V. Sharma and P.L.N. Raju. (2018). Fusion of Hyperion and RapidEye data sets and comparison of classification accuracy over the Dhemaji and Lakhimpur districts of Assam India, Journal of Geomatics (JoG), 12, 35-46.

Kuncheva, L.I and C. Whitaker. (2003). Measures of diversity in classifier ensembles, Machine Learning, 51, 181-207.

Kuncheva, L.I and J.J. Rodriguez. (2007). An experimental study on rotation forest ensembles, Multiple Classifier Systems, 4472,459–468.

Lo, S.L., R. Chiong and D. Cornforth. (2015). Using support vector machine ensembles for target audience classification on Twitter, PLOS ONE, 10(4): e0122855.

Lira, M.M.S., R.R.B. Aquino, A.A. Ferreira, M.A. Carvalho, O.N. Neto and G.S.M. Santos. (2007). Combining multiple artificial neural networks using random committee to decide upon electrical disturbance classification, International Joint Conference on Neural Networks, pp. 2863-2868.

Liu, K.H and D.S. Huang. (2008). Cancer classification using rotation forest, Computers in Biology and Medicine, 38(5), 601–610. doi:10.1016/j.compbiomed.2008.02.007.

Liu, Y., D. Pi and Q. Cheng. (2016). Ensemble kernel method: SVM classification based on game theory, Journal of Systems Engineering and Electronics, 27(1), 251-259.doi: 10.1109/JSEE.2016.00025.

Opitz, D and R. Maclin. (1999). Popular ensemble methods: An empirical study, Journal of Artificial Intelligence Research, 11, 169-198.doi: 10.1613/jair.614.

Quinlan, J.R. (1986). Induction of decision trees, Machine Learning, 1, 81-106.

Rahman, M.M., M.M. Islam, K. Murase and X. Yao. (2016). Layered ensemble architecture for time series forecasting, IEEE Transactions on Cybernetics, 46(1), 270-283.

Rawat, U., V. Sharma, D. Chutia, N. Nishant, U. Baruah, U and P.L.N. Raju. (2018). Random forest versus rotation forest – A comparative assessment towards classification of satellite data, 5^{th} IEEE International Conference on Computing for Sustainable Global Development, 14th - 16th March, Proceedings of the 12th INDIACom; pp. 1732—1736. INDIACom-2018; IEEE Conference ID: 42835.

Rodriguez, J.J., L.I. Kuncheva and C.J. Alonso. (2006). Rotation Forest: A new classifier ensemble method, IEEE Transactions on Pattern Analysis & Machine Intelligence,28, 1619-1630. doi:10.1109/TPAMI.2006.211.

Sharma, V., D. Baruah, D. Chutia, P.L.N. Raju and D.K. Bhattacharya. (2016). An assessment of support vector machine kernel parameters using remotely sensed satellite data, IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 1567-1570. doi:10.1109/RTEICT.2016.7808096.

Sharma, V., N. Choudhury, R. Mandal, D. Chutia, U. Baruah and R. Kumar. (2018). An empirical study on traditional classifiers for remotely sensed satellite data classification,5^{th} IEEE International Conference on Computing for Sustainable Global Development, 14th - 16th March, Proceedings of the 12th INDIACom; pp. 2032—2036. INDIACom-2018; IEEE Conference ID: 42835.

Sørensen, L and M. Nielsen. (2018). Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination, Journal of Neuroscience Methods,302, 66-74. doi: 10.1016/j.jneumeth.2018.01.003.

Sylvester, E.V.A., P. Bentzen, I.R. Bradbury, M. Clément, J. Pearce, J. Horne and R.G. Beiko. (2017). Applications of random forest feature selection for fine- scale genetic population assignment, Evolutionary Applications, 11, 153–165. https://doi.org/10.1111/eva.12524

Tatsis, V.A., C. Tjortjis and P. Tzirakis. (2013). Evaluating data mining algorithms using molecular dynamics trajectories, International Journal of Data Mining and Bioinformatics,8(2), 169–187.doi:10.1504/IJDMB.2013.055499.

Vyškovský, R., D. Schwarz. E. Janoušová and T. Kašpárek. (2016). Random subspace ensemble artificial neural networks for first-episode schizophrenia classification,Federated Conference on Computer Science and Information Systems (FedCSIS), Gdansk, pp.317-321.DOI:10.15439/2016F333.

Wandekoken, E.D., F.M. Varejão, R. Batista and T.W. Rauber. (2011). Support vector machine ensemble based on feature and hyperparametervariation for real-world machine fault diagnosis, soft computing in industrial applications, Advances in Intelligent and Soft Computing, Springer, 96, 271-282. https://doi.org/10.1007/978-3-642-20505-7_24.

Witten, I.H and E. Frank. (2005). Data Mining: Practical machine learning tools and techniques, San Francisco: CA: ELSEVIER, 2nd ed., chapter.7 and 10.

Huang, X and L. Zhang. (2013). An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery, IEEE Transactions on Geoscience and Remote Sensing, 51(1), 257-272. doi: 10.1109/TGRS.2012.2202912.

Yang, BS., X. Di and T. Han. (2008). Random forests classifier for machine fault diagnosis, Journal of Mechanical Science and Technology, 22(9), 1716–1725. https://doi.org/10.1007/978-1-84628-814-2_82.

Yu, Z., L. Li, J. Liu and G. Han. (2015). Hybrid adaptive classifier ensemble, IEEE Transactions on Cybernetics, 45(2), 177-190. doi: 10.1109/TCYB.2014.2322195.