Fusion of SMOTE and Outlier Detection Techniques for Land-Cover Classification Using Support Vector Machines

Rahul Gomes, Mostofa Ahsan and Anne Denton
Department of Computer Science
North Dakota State University
Fargo, ND, 58102
rahul.gomes@ndsu.edu, mostofa.ahsan@ndsu.edu, anne.denton@ndsu.edu

Abstract

Support Vector Machines is an excellent tool for classification. Since they can transform the dataset to higher degrees to produce decision boundaries, they outperform probabilistic models and clustering algorithms [1]. However, using classification algorithms on datasets without preprocessing is not a good practice [2]. The two most significant drawbacks being the presence of cloud cover and the dataset being imbalanced. Cloud cover introduces noise in the dataset reducing its overall accuracy [3]. Removal of imbalance in datasets has been an issue of importance for machine learning and data mining implementation techniques [4]. A dataset is said to be imbalanced if there is a significant difference in the number of instances of one attributes when compared to another. A typical example of an imbalanced GIS dataset is land-use classification where a study area can be mostly forest with very little urban settlement. This introduces bias in the dataset causing the output to be skewed towards the attribute having more instances compared to the sparse ones. The overall accuracy might be high, but the sensitivity and specificity might suffer for different classes.

To compensate for this problem, we propose a hybrid approach using Support Vector Machines that performs Synthetic Minority Over-Sampling Technique (SMOTE) [5] evaluation. Support Vector Machines is implemented using three different kernels. This study investigates the potential of such a hybrid approach to increase overall accuracy of the dataset as compared to the usage of stand-alone classification algorithm.

Introduction

Remote sensing analysis is centered around landcover estimation, classification [2], [3] and change detection [4], [5]. They form the foundation of various studies related to mapping crop diseases [6] to climate change analysis [7]. The foundation of remote sensing lies in supervised classification [8]. Most models work by training subset of selected pixels obtained from remotely sensed imagery and using the model to test the results on a testing dataset. Once accuracy can be estimated to a desired level, the model is applied on the entire dataset to generate a classified map of the study area. On investigating the accuracy of individual class labels using sensitivity and specificity values, there is a probability of minority classes suffering from low accuracies compared to majority classes. It is not uncommon to find a model with high overall accuracy occurring due to high true positives for a class label occupying majority of the dataset. So, even if the model is branded as accurate, it is not dependable if the study must be conducted on pixel values related to minority class labels.

Another prevalent problem is the presence of cloud cover in time-series GIS datasets [9]. Yield variance on crops [10], [11] over a desired time-frame of 5 to 10 years refers to collecting multiple raster datasets over the same place within a limited time corresponding to the crops' harvesting season. Very often, some time series data suffers from cloud cover. This noise in raster images makes classification a tough job for researchers. Although interpolation techniques are used to remove cloud cover, a dataset dwarfed almost entirely by cloud is practically useless for analysis.

Feature selection can be an effective tool to reduce the effect of noise in the datasets [12]. Data mining techniques such as random forests that use feature selection have been used effectively in remote sensing [13]. However, minority class labels suffer from classification more than majority labels even though they might have the same amount of noise in them [14]. This is because the majority class labels have sufficient records to identify noise as outliers compared to minority class labels. To address this anomaly, SMOTE can be applied on the records before classification. SMOTE sampling technique tries to balance the dataset by oversampling to the minority class labels to ensure that the class has enough training points to reduce the bias of majority class label. Although SMOTE creates new set of values, it does so by identifying the density of values clustered in the minority dataset. By doing so, SMOTE picks the actual values that contribute to minority class labels and outliers that negatively affect the outcome. SMOTE then creates records that are closer to actual values increasing the accuracy of the minority class label.

This study aims to find if SMOTE is an effective tool for landcover classification as well. Data corresponding to multiple years have been used that contain cloud cover. SVM classification will be applied with and without SMOTE to identify how it affects outcome.

Previous work

SMOTE has been widely used in various fields to reduce the problem of imbalanced dataset. An apt area for application of SMOTE is in fraud detection. Fraud detection such as credit cards, car insurance etc. deals with a highly unbalanced data that contains more legitimate records with very few fraud signatures. A hybrid approach was implemented using SMOTE and *k* reversed nearest neighbors (*k*RNNs) on an insurance fraud dataset [15]. Data mining approaches such as C4.5, Naïve Bayes and *k*-NN were tested on the unbalanced dataset and results were compared with the SMOTE applied dataset. The hybrid method was able to increase prediction accuracy for both fraud and legitimate labels. The highest increase was recorded for *k*-NN where accuracy increased from 85% to 99.9%. The hybrid method used extreme outlier detection in minority class and removed them before over-sampling the minority class. This ensured that the dataset produced new points around minority data sample showing higher density and ignored outliers.

In [16] authors quantified the relationship between over and under sampling dataset using SMOTE. Classification using SVM was done on three balanced datasets obtained from UCI machine learning repository. The classification generated accuracy close to 100%. This was followed by under sampling the majority class and recording the hyperplanes and oversampling the minority class and recording the hyperplanes. The dot product was obtained between the original hyperplane and the under sampled data. Another dot product was obtained for the original data with the oversampled one. Results indicate that oversampling the minority data is always beneficial to accuracy estimation proving that SMOTE is good for oversampling. However, under sampling extensively can reduce the accuracy estimation and increase the variance of the hyperplane as the model does not contain sufficient training data. Using different error costs based on data labels and SMOTE implementation the model generated smoother and more consistent hyperplanes.

A combination of genetic algorithm(GA) and SMOTE was used for feature selection and balancing dataset [12]. Employing the concepts of selection, crossover and mutation a GA tries to mimic evolution concepts to select a proper solution using fitness function [17], [18]. The GA was used to evaluate the dataset and identify features that were considered as more important in defining the dataset. Using the important features generate by the GA, SMOTE was applied on the dataset to oversample the minority label. A k=5 was selected for SMOTE where k denotes the number of neighbors selected for random sample generation. An appropriate over-sampling rate was selected to ensure the randomly generated sample data would be spread throughout the decision boundary of the SVM that would be applied later for classification. This method avoids dense cluster of data and increases prediction accuracy. The method achieved an accuracy of 72% for a lymphoma dataset used in the study.

SMOTE application in various fields have been successful and its application in removing noise in GIS datasets can be of great help in various remote sensing applications.

Support Vector Machines

Support Vector Machines is a non-parametric approach to classification. It works by creating decision boundaries between one class and the other. The decision boundary or the hyperplane that separates one class from the other is determined by points in the dataset that are responsible for separating the two class labels [19]. These points are also known as support vectors, hence, the term support vector machines. Original SVM algorithms apply one-against-one methodology making it an appropriate tool for binary classification. Recent versions also perform multi-class classification by doing multiple one-against-one iterations over several class labels and aggregating the decision boundaries obtained using certain reliability frameworks such as static and dynamic reliability measures [20]. The underlying difference in results obtained using SVM lies in the kernels. Three of the most frequently used SVM kernels include linear, polynomial and radial. The linear kernel as its name suggests works by creating linearly separable hyperplanes between the classes. Two hyperplanes exist as shown in equation 1.

$$m.x_i + c \geq 1 \quad \forall\, y = 1 \; and \; m.x_i + c \leq 1 \quad \forall\, y = -1 \qquad \text{Equation 1}$$

Here m denotes the orientation of the hyperplane and c signifies the distance of the hyperplane from the origin. The support vectors lying on the hyperplane can be denoted by equation 2.

$$m.x_i + c = \pm 1 \qquad \text{Equation 2}$$

Figure 1 shows the linearly separable SVM classification for a two-class problem. Here, green lines correspond to the two hyperplanes and the points lying on them signify the support vectors.

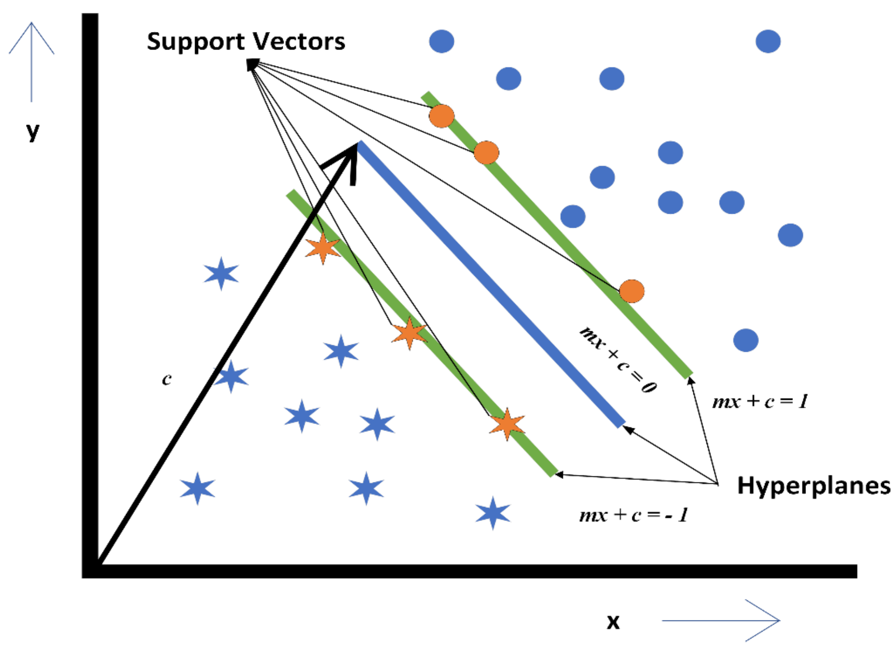

Figure 1: Diagrammatic representation of linear kernels in SVM.

Linearly separable kernels are good when the dataset is easily separable using a straight-line hyperplane. However, in some cases with multiple attributes, a single horizontal line cannot easily demarcate the difference in class labels. Polynomial kernel is a non-linear kernel that works by increasing the degree of attributes to a higher dimension so that a better hyperplane can be deduced. Equation 3 shows a polynomial kernel

$$P(x, y) = (x^T . y + c)^m \qquad \text{Equation 3}$$

Here, m denotes the degree to which the linear solution has been raised to obtain a better separation. Once a certain degree yields a satisfactory separation, the model uses the corresponding hyperplane for estimation. Polynomial kernels are frequently used in Natural Language Processing [21] and can yield significantly good results than linear kernels in remote sensing when multiple types of independent attributes are used such as different spectral bands in remote sensing [22].

The radial basis kernel shown in equation 4 is also a commonly used kernel that uses the Euclidean distance between values to determine the hyperplanes. Here x and y are two samples and $-2\sigma^2$ is the inverse of kernel width. The kernel width and regularization parameter that is chosen varies from one study to the other.

$$P(x, y) = exp(\frac{\|x-y\|^2}{2\sigma^2})^m \qquad \text{Equation 4}$$

Study Area

Landsat 5 multispectral images were used for this study. The data was obtained on September 19, 2009 as shown in Figure 2 (left) and August 21, 2010 as shown in Figure 2 (right). The Landsat WRS_PATH was 29 and WRS_ROW was 28. The dataset resolution is 30m and it covered Richland county in North Dakota and Roberts county in South Dakota. For Minnesota the counties covered include Otter Tail, Grant, Douglas, Stevens, Pope, Big Stone, Swift and parts of Clay, Wadena and Todd. The image consisted of seven bands which were used as attributes for classification. Normalized Difference Vegetation Index (NDVI) obtained from band 3 and 4 was also used for classification.

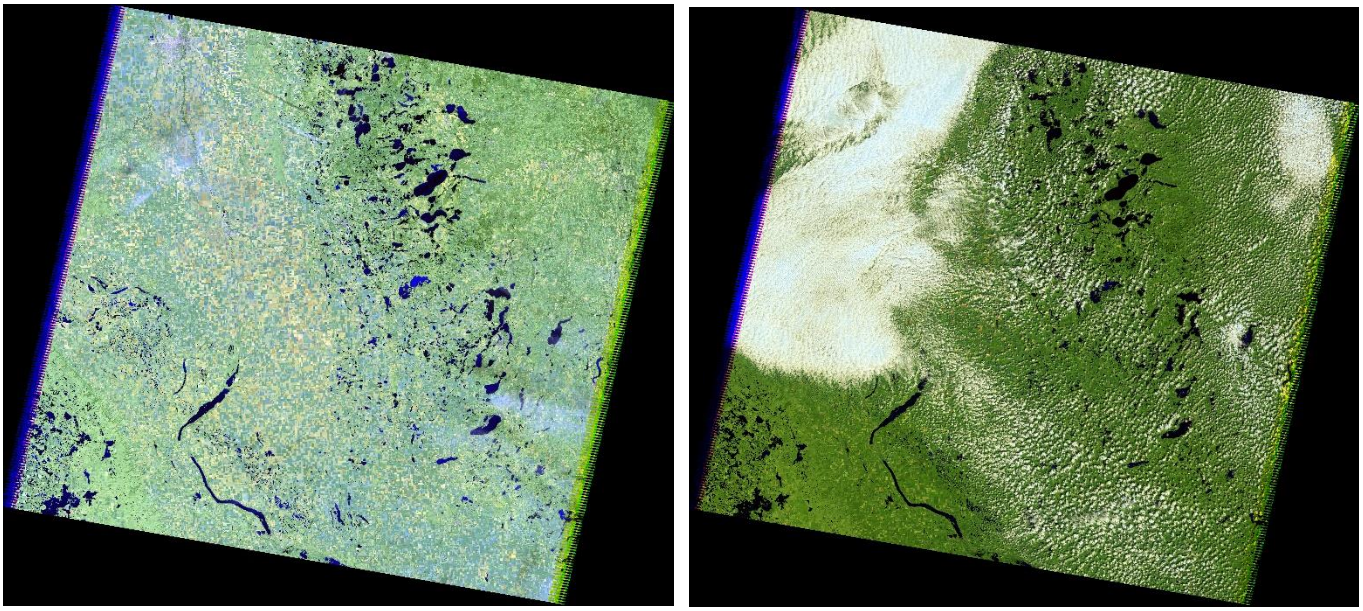

Figure 2: Raster images used for data points 2009 (left), 2010 (right).

Accuracy Assessment

Based on the pixel granularity, five distinct classes were selected for the study, namely grass, forest, farm, urban and water. Since a 30m resolution is too coarse for classification of other features such as wetlands, they were not included in the study. A total of 1400 points were used for training the model and a separate set of 1200 data points were used for testing. The total of 16 attributes were used for the study with bands 1 to 7 for each of the two years long with their NDVI values. Out of the 1400 data points used for training, 20% of the dataset consisted of cloud cover. This variation was introduced to check if SMOTE results are affected by their inclusion. The testing dataset was completely free of cloud cover. The accuracy results are shown in table 2 for the three kernels used for the study.

SMOTE was applied to the training dataset to create a balanced dataset. A comparison of the number of class labels before and after SMOTE application is shown in table 1. The highest-class label, farm was 5.67 times the lowest, urban. After running the model on the new training dataset, evaluation was done on the same testing dataset to evaluate any increase in accuracy. The results are shown in table 3.

Label	**Farm**	**Forest**	**Grass**	**Urban**	**Water**
Before SMOTE	425	300	400	75	200
After SMOTE	425	425	425	425	425

Table 1: Class labels with and without SMOTE

a) *Radial without SMOTE*

Prediction	Reference						
	Farm	***Forest***	***Grass***	***Urban***	***Water***	**Total**	**Specificity**
Farm	360	0	284	16	41	701	0.51
Forest	0	250	0	0	0	250	1.00
Grass	36	0	114	4	0	154	0.74
Urban	4	0	2	30	0	36	0.83
Water	0	0	0	0	59	59	1.00
Total	400	250	400	50	100	1200	
Sensitivity	0.90	1.00	0.29	0.60	0.59		

Accuracy 67.75 **Kappa** 0.55

b) Linear without SMOTE

Prediction	Reference						
	Farm	***Forest***	***Grass***	***Urban***	***Water***	**Total**	**Specificity**
Farm	298	0	268	12	0	578	0.52
Forest	0	250	5	0	0	255	0.98
Grass	88	0	127	1	0	216	0.59
Urban	14	0	0	37	0	51	0.73
Water	0	0	0	0	100	100	1.00
Total	400	250	400	50	100	1200	
Sensitivity	0.75	1.00	0.32	0.74	1.00		

Accuracy 67.67 **Kappa** 0.56

c) Polynomial without SMOTE

Prediction	Reference						
	Farm	***Forest***	***Grass***	***Urban***	***Water***	**Total**	**Specificity**
Farm	331	0	323	10	0	664	0.50
Forest	0	248	0	0	0	248	1.00
Grass	68	2	75	1	0	146	0.51
Urban	1	0	2	39	0	42	0.93
Water	0	0	0	0	100	100	1.00
Total	400	250	400	50	100	1200	
Sensitivity	0.83	0.99	0.19	0.78	1.00		

Accuracy 66.08 **Kappa** 0.53

Table 2: Results for SVM classification without SMOTE

a) Radial with SMOTE

Prediction	Reference						
	Farm	***Forest***	***Grass***	***Urban***	***Water***	**Total**	**Specificity**
Farm	389	0	300	26	41	756	0.51
Forest	0	249	0	0	0	249	1.00
Grass	10	1	98	4	0	113	0.87
Urban	1	0	2	20	0	23	0.87
Water	0	0	0	0	59	59	1.00
Total	400	250	400	50	100	1200	
Sensitivity	0.97	1.00	0.25	0.40	0.59		

Accuracy 67.92 **Kappa** 0.55

b) Linear with SMOTE

Prediction	Reference						
	Farm	***Forest***	***Grass***	***Urban***	***Water***	**Total**	**Specificity**
Farm	297	0	241	12	0	550	0.54
Forest	0	250	3	0	0	253	0.99
Grass	70	0	154	0	0	224	0.69
Urban	33	0	2	38	0	73	0.52
Water	0	0	0	0	100	100	1.00
Total	400	250	400	50	100	1200	
Sensitivity	0.74	1.00	0.39	0.76	1.00		

Accuracy 69.92 **Kappa** 0.59

c) Polynomial with SMOTE

Prediction	Reference						
	Farm	***Forest***	***Grass***	***Urban***	***Water***	**Total**	**Specificity**
Farm	330	0	309	16	0	655	0.50
Forest	0	246	0	0	0	246	1.00
Grass	57	4	90	0	0	151	0.60
Urban	13	0	1	34	0	48	0.71
Water	0	0	0	0	100	100	1.00
Total	400	250	400	50	100	1200	
Sensitivity	0.83	0.98	0.23	0.68	1.00		

Accuracy 66.67 **Kappa** 0.57

Table 3: Results for SVM classification with SMOTE

The results obtained from accuracy assessments show slight increase after SMOTE application. The highest increase of 2.25% has been recorded for linear kernel, while an increase of 0.17% for radial and 0.58% for polynomial has been recorded. The accuracy suffers mostly due to misclassification of grass. The SMOTE can deal with the impact of cloud cover to some extent, but it cannot distinguish between grass and farm as it has similar spectral signatures to some extent. Object based classification can be used to some extent to rectify this problem which is beyond the scope of this paper and will be kept as future work.

Comparative Analysis

Since SMOTE could not provide a significant increase in accuracy, one potential cause could be the low ratio between highest and lowest class labels. To test this hypothesis, another accuracy assessment was conducted on a landcover dataset obtained from UCI Machine Learning Repository [23]. The time-series dataset consists of 28 attributes corresponding to NDVI values obtained across a period of January 6th, 2014 to July 20th, 2015. Each attribute has varying degrees of cloud cover in them. The class labels used for this dataset were farm, forest, grass, impervious, orchard and water. The training dataset consists of 10545 labels and testing dataset has 300 labels. The accuracy assessment before SMOTE application is shown in table 5.

Once these accuracies were obtained, SMOTE analysis was done on the dataset to balance it. The difference between the number of records for each class label before and after SMOTE is shown in table 4. In this dataset the ratio of the highest-class label forest to the lowest class label orchard is 140.21 which is much higher than our previous study. The accuracy assessment conducted after SMOTE implementation is shown in table 6.

Label	**Farm**	**Forest**	**Grass**	**Impervious**	**Orchard**	**Water**
Before SMOTE	1441	7431	446	969	53	205
After SMOTE	7431	7431	7431	7431	7431	7431

Table 4: Class labels with and without SMOTE

Results indicate that the accuracy increases at a greater rate for this dataset. The highest increase has been recorded for the polynomial kernel at 4.33%, while linear kernel records an increase of 2.33 % and radial kernel increases by 1% showing that the extent of imbalance could be a contributing factor to the jump in accuracy.

a) Radial without SMOTE

Prediction	Reference							
	Farm	***Forest***	***Grass***	***Impervious***	***Orchard***	***Water***	**Total**	**Specificity**
Farm	39	7	10	1	9	1	67	0.58
Forest	13	57	13	1	26	20	130	0.44
Grass	0	13	7	1	3	1	25	0.28
Impervious	1	1	5	37	0	3	47	0.79
Orchard	0	0	0	0	9	0	9	1.00
Water	0	0	1	0	0	21	22	0.95
Total	53	78	36	40	47	46	300	
Sensitivity	0.74	0.73	0.19	0.93	0.19	0.46		

Accuracy 56.67 **Kappa** 0.46

b) Linear without SMOTE

Prediction	Reference							
	Farm	***Forest***	***Grass***	***Impervious***	***Orchard***	***Water***	**Total**	**Specificity**
Farm	33	4	5	0	5	1	48	0.69
Forest	19	61	15	1	36	4	136	0.45
Grass	0	12	13	4	1	1	31	0.42
Impervious	1	1	2	35	0	8	47	0.74
Orchard	0	0	0	0	5	0	5	1.00
Water	0	0	1	0	0	32	33	0.97
Total	53	78	36	40	47	46	300	
Sensitivity	0.62	0.78	0.36	0.88	0.11	0.70		

Accuracy 59.67 **Kappa** 0.5

c) Polynomial without SMOTE

Prediction	Reference							
	Farm	***Forest***	***Grass***	***Impervious***	***Orchard***	***Water***	**Total**	**Specificity**
Farm	38	7	6	5	11	1	68	0.56
Forest	15	57	15	2	23	3	115	0.50
Grass	0	14	12	0	3	2	31	0.39
Impervious	0	0	2	33	0	8	43	0.77
Orchard	0	0	0	0	10	0	0	1.00
Water	0	0	1	0	0	32	33	0.97
Total	53	78	36	40	47	46	300	
Sensitivity	0.72	0.73	0.33	0.83	0.21	0.70		

Accuracy 60.67 **Kappa** 0.51

Table 5: Results for SVM classification without SMOTE

a) Radial with SMOTE

Prediction	Reference							
	Farm	***Forest***	***Grass***	***Impervious***	***Orchard***	***Water***	**Total**	**Specificity**
Farm	42	8	7	4	12	1	74	0.57
Forest	10	52	8	0	19	25	114	0.46
Grass	0	16	13	1	0	1	31	0.42
Impervious	1	1	8	35	0	4	49	0.71
Orchard	0	0	0	0	16	0	16	1.00
Water	0	1	0	0	0	15	16	0.94
Total	53	78	36	40	47	46	300	
Sensitivity	0.79	0.67	0.36	0.88	0.34	0.33		

Accuracy 57.67 **Kappa** 0.48

b) Linear with SMOTE

Prediction	Reference							
	Farm	***Forest***	***Grass***	***Impervious***	***Orchard***	***Water***	**Total**	**Specificity**
Farm	35	7	13	0	6	1	62	0.56
Forest	8	44	5	1	12	1	71	0.62
Grass	1	21	16	5	0	3	46	0.35
Impervious	2	1	2	29	0	7	41	0.71
Orchard	6	3	0	1	28	0	38	0.74
Water	1	2	0	4	1	34	42	0.81
Total	53	78	36	40	47	46	300	
Sensitivity	0.66	0.56	0.44	0.73	0.60	0.74		

Accuracy 62.00 **Kappa** 0.54

c) Polynomial with SMOTE

Prediction	Reference							
	Farm	***Forest***	***Grass***	***Impervious***	***Orchard***	***Water***	**Total**	**Specificity**
Farm	48	11	8	1	13	3	84	0.57
Forest	3	43	3	0	12	1	62	0.69
Grass	0	19	18	0	3	4	44	0.41
Impervious	1	3	6	38	0	9	57	0.67
Orchard	1	2	0	0	19	0	22	0.86
Water	0	0	1	1	0	29	31	0.94
Total	53	78	36	40	47	46	300	
Sensitivity	0.91	0.55	0.50	0.95	0.40	0.63		

Accuracy 65.00 **Kappa** 0.57

Table 6: Results for SVM classification with SMOTE

Conclusions and Future Work.

A study has been conducted on how SMOTE can affect remote sensing dataset and impact accuracy. Two different datasets have been used, one with an imbalance ratio of 140.21 and the other is 5.67. SVM classification using three kernels namely radial, polynomial and linear kernels have been used on the datasets. Results indicate that the accuracy increases with SMOTE application. However, the rate of increase is not the same for both datasets. Using the three kernels, average increase in accuracy for the dataset with imbalance ratio of 5.67 is 1.003%, whereas for dataset with 140.21 imbalance ratio, it is 2.553%. Since there is no significant increase in accuracy, we can derive that SMOTE application is not so efficient in differentiating multispectral attributes. Grass, farms and forests which share similar spectral signatures often gets misclassified when using multispectral data alone for classification. SMOTE can deal with cloud cover to some extent, but there is not sufficient help it can provide to deal with this class boundary problem. Further work would be done to establish how SMOTE behaves when object-based classification is attempted. Object based classification in remote sensing uses attributes such as shape, density and elevation data to derive class labels. It is expected to produce higher accuracy and complement multispectral analysis. The extent of SMOTE's behavior to such a dataset would be verified.

Acknowledgement

This material is based upon a work supported by the National Science Foundation through grant IIA-1355466. We thank SatShot for valuable discussions.

References

[1] L. T.- ICML and undefined 1999, "Feature selection as a preprocessing step for hierarchical clustering," *Citeseer*.

[2] S. E. Franklin and M. A. Wulder, "Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas," *Prog. Phys. Geogr.*, vol. 26, no. 2, pp. 173–205, Jun. 2002.

[3] R. Alvarez *et al.*, "Multitemporal land-cover classification of Mexico using Landsat MSS imagery," *Int. J. Remote Sens.*, vol. 24, no. 12, pp. 2501–2514, Jan. 2003.

[4] P. F. Rozario, P. Oduor, L. Kotchman, and M. Kangas, "Transition Modeling of Land-Use Dynamics in the Pipestem Creek, North Dakota, USA," *J. Geosci. Environ. Prot.*, vol. 5, no. 3, pp. 182–201, 2017.

[5] P. Rozario, P. Oduor, and L. Kotchman, "Uncertainty Analysis of Spatial Autocorrelation of Land-Use and Land-Cover Data within Pipestem Creek in North Dakota," *J. Geosci.*, vol. 5, no. 8, pp. 71–88, 2017.

[6] A. Tran, J. Gardon, S. Weber, and L. Polidori, "Mapping disease incidence in suburban areas using remotely sensed data," *Am. J. Epidemiol.*, vol. 156, pp. 662–668, 2002.

[7] S. Kalluri, “Monitoring ecosystems vulnerable to climate change,” *IEEE Int. Geosci. Remote Sens. Symp.*, vol. 5, no. C, pp. 2802–2804, 2002.
[8] R. G. Congalton, “A Review of Assessing the Accuracy of Classification of Remotely Sensed Data A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data,” *Remote Sens. Environ.*, vol. 4257, no. September, pp. 34–46, 1991.
[9] X. Yang and C. P. Lo, “Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area,” *Int. J. Remote Sens.*, vol. 23, no. 9, pp. 1775–1798, Jan. 2002.
[10] A. M. Denton, M. Ahsan, D. Franzen, and J. Nowatzki, “Multi-scalar analysis of geospatial agricultural data for sustainability,” in *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, 2016, pp. 2139–2146.
[11] H. Bu, L. K. Sharma, A. Denton, and D. W. Franzen, “Comparison of satellite imagery and ground-based active optical sensors as yield predictors in sugar beet, spring wheat, corn, and sunflower,” *Agron. J.*, vol. 109, no. 1, pp. 299–308, 2017.
[12] T. Deepa and M. Punithavalli, “An E-SMOTE technique for feature selection in High-Dimensional Imbalanced Dataset,” in *2011 3rd International Conference on Electronics Computer Technology*, 2011, pp. 322–324.
[13] M. Pal, “Random forest classifier for remote sensing classification,” *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, Jan. 2005.
[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
[15] T. M. Padmaja, N. Dhulipalla, R. S. Bapi, and P. R. Krishna, “Unbalanced data classification using extreme outlier elimination and sampling techniques for fraud detection,” in *15th International Conference on Advanced Computing and Communications (ADCOM 2007)*, 2007, pp. 511–516.
[16] R. Akbani, S. Kwek, and N. Japkowicz, “Applying Support Vector Machines to Imbalanced Datasets,” 2004, pp. 39–50.
[17] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, vol. Addison-We. 1989.
[18] R. Gomes and J. Straub, “Genetic algorithm for flood detection and evacuation route planning,” in *spiedigitallibrary.org*, 2017, p. 1019816.
[19] B. Scholkopf *et al.*, “Comparing support vector machines with Gaussian kernels to radial basis function classifiers,” *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2758–2765, 1997.
[20] Y. Liu and Y. F. Zheng, “One-Against-All Multi-Class SVM Classification Using Reliability Measures,” in *Proceedings of International Joint Conference on Neural Networks*, 2005, pp. 849–854.
[21] Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C.-J. Lin, “Training and Testing Low-degree Polynomial Data Mappings via Linear SVM,” *J. Mach. Learn. Res.*, vol. 11, pp. 1471–1490, 2010.
[22] B. Dixon and N. Candade, “Multispectral landuse classification using neural networks and support vector machines: One or the other, or both?,” *Int. J. Remote Sens.*, vol. 29, no. 4, pp. 1185–1206, Feb. 2008.
[23] B. A. Johnson and K. Iizuka, “Integrating OpenStreetMap crowdsourced data and

Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines," *Appl. Geogr.*, vol. 67, pp. 140–149, 2016.