International Journal of
Intelligent Systems and
Applications in Engineering
ISSN:2147-6799 www.atscience.org/IJISAE
Advanced Technology and Science
Original Research Paper

The Impact of Feature Selection on Urban Land Cover Classification

Turgut Dogan*[1], Alper Kursat Uysal[1]

Accepted : 06/03/2018 Published: 30/03/2018

Abstract: Many of the studies in the literature about land cover classification are focused on the feature extraction and classification rather than feature selection. In this paper, the impact of feature selection on urban land cover classification is extensively analyzed. Three types of features namely spectral, texture, and size/shape features are used for this analysis. This analysis is carried out using three variations of a filter based feature selection method and three widely-known classification algorithms. The feature selection method used for the comparison is a multivariate filter method namely correlation-based feature subset selection where a feature subset evaluator and a search method are integrated. Best first search, genetic search, and greedy stepwise search are three different search methods used for this integration. The classification algorithms employed are Bayesian network, random forest, and support vector machine. The experimental results explicitly indicate that feature selection improves classification accuracy in all cases. Besides, according to the experimental results, random forest classifier is the most successful one among these three classifiers while both feature selection is applied and not applied. Largest improvement in the classification performance is obtained when greedy stepwise search based feature selection method and support vector machine classifier is applied together. Also, the contribution of spectral features to the performance of classification is more than size/shape and texture features.

Keywords: *Classification, Feature selection, Land cover, Remote Sensing.*

1. Introduction

The dynamic development of urban areas caused an increasing need for automatic identification of urban land cover types [1]. This makes land cover information a vital input to various processes such as developmental, environmental, and resource planning applications [2]. Hence, extraction of urban land cover information from high resolution images and automatic classification of this information to various land cover types is a common research topic in remote sensing [3-5]. There are two main approaches for extracting land cover information from high resolution data. These approaches are pixel based image classification [6] and geospatial object-based image analysis [7]. Pixel-based approaches analyze the spectral properties of every pixel within the area of interest. While making this analysis, spatial or contextual information related to these pixels are not taken into account [8]. In geospatial object-based approaches, image is segmented into relatively homogeneous regions generally referred as segment or image objects [6, 9-11]

In the literature, there exist numerous studies dealing with classification of urban land cover and various approaches are utilized in these studies. As an example, QuickBird image data is used over a central region in the city of Phoenix, Arizona in order to examine the performance of object-based classifiers for identifying urban classes [10]. According to the experimental results, the object-based classifier achieves a high overall accuracy while one of the most commonly used pixel-based classification method, namely maximum likelihood classifier, produces a lower overall accuracy. It is stated that the object-based classifier is significantly better than the classical per-pixel classifiers. Besides, the performances of pixel-based and object-based image analysis approaches for classifying broad land cover classes over agricultural landscapes are evaluated using three separate supervised machine learning algorithms namely decision tree, random forest, and support vector machine classifiers [12]. It is reported that the difference between the overall classification accuracies of pixel-based and object-based classifications are not statistically significant when the same machine learning algorithms are applied. Landsat ETM+ images of a mountainous area in Mexico are used to analyze the performance of combined object-based and pixel-based land cover classification [13]. It is stated that the combination method produces the best results among all in terms of overall accuracy. Besides, the performance of different classification methods for land cover mapping in the vicinity of the Alto Ribeira Tourist State Park, a Brazilian Atlantic rainforest area is investigated in a more recent study [14]. The classification results show that the object-based classification explicitly outperforms the pixel-based classification in terms of accuracy where the accuracies are 89.7% and 57.8% for object-based and pixel-based classification, respectively. A multi-scale approach is used for classifying land cover in a high resolution image of an urban area [15]. In this work, super-object information is used as additional input data for image classification. The accuracies of classifications including super-object variables are compared with the classification accuracies of image segmentations not including super-object information. The experimental results show that the accuracy of the system that uses super-object information is higher than the one not including super-object information. Furthermore, three types of classification methods which can be referred as pixel and object-based versions of support vector machine, and pixel-based version of decision tree, are used to classify SPOT 5 satellite image into land cover types [16]. According to the experimental results, the object-based version of support vector machine classifier is the

[1] *Computer Engineering Department, Anadolu University, Iki Eylul Campus, Faculty of Engineering, 26555, Eskisehir, TURKEY*
** Corresponding Author: Email: turgutdogan@anadolu.edu.tr*

best performer among all in terms of accuracy.

In the literature, feature selection techniques are commonly applied for various types of image classification tasks. Since the land cover information extracted from images can be high-dimensional, selecting smaller subsets of features consisting of informative ones may be useful to classification as it is valid for various pattern recognition applications. However, there exist limited numbers of studies dealing with feature selection for tasks related to land cover classification. As an example, the impact of feature selection is analyzed for classification of urban structure types with maximum likelihood classifier [17]. In this study, 50 spatial feature types are calculated initially and sequential forward feature selection method is subsequently employed. As another example, the impact of feature selection on support vector machine classification of two hyperspectral sensor data sets is analyzed with four separate feature selection method [18]. According to the experimental results, the accuracy of classification may decline significantly with the addition of features when a small training sample is used. However, it is reported that feature selection may be useful when a large training sample is available and this makes is a valuable process for support vector machine classification. Besides, a class-based feature selection method based on fast constrained search algorithm is employed for classification of hyperspectral data [19]. In this study, a new scheme for feature selection employing Bayesian classifier is proposed and the experimental results show that the proposed method increase effectiveness of classifier in terms of accuracy. Furthermore, classification tree analysis feature selection method is employed for object-based land cover classification and it is evaluated with support vector machine, maximum likelihood, and neural network classification algorithms [20]. The total numbers of object features extracted are 47 and the numbers of selected features vary between 1 and 22. The experimental results show that both support vector machine and neural network classifier produces stable results as the feature dimension increases towards 22. On the other hand, the performance of maximum likelihood classifier decreases considerably in terms of accuracy when the feature dimension increases towards 22. It is reported that the eigen-space projection-based parameter selection method provides better classification accuracy than other feature selection methods. Furthermore, land cover classification frameworks including feature selection stage and support vector machine classifier are analyzed [21]. In this study, a third-order class-dependent mutual information based feature selection method is proposed and compared with three separate feature selection methods namely maximum mutual information, maximum-relevance minimum-redundancy, and conditional mutual information maximization. It is stated that the proposed method gives a comparatively better ranking than the rests.

Many of the studies in the literature are focused on the feature extraction and classification part of the land cover classification task. However, feature selection is less focused part in these studies and also the analysis of feature selection on urban land cover classification is given as a future work in one of the previous recent studies [15]. For this purpose, in this study, an extensive analysis is realized for measuring the impact of feature selection on object-based urban land cover classification. In order to make this analysis, a recently published public dataset including features extracted from a high resolution image of an urban area is used. This dataset includes spectral, texture, size/shape features obtained at the end of the feature extraction process for object-based land cover classification. In order to evaluate feature selection process, the performances of three variations of a multivariate filter-based feature selection method and three widely-used classifiers are compared. This comparison is realized with the well-known F-Measure and 10-fold cross validation is used in the experiments for fair evaluation. Besides, profiles of the feature sets obtaining higher scores are analyzed for detecting common informative features and their corresponding categories.

Rest of the paper is organized as follows: Section 2 describes the features used in the experiments. Feature selection methods employed in the experiments are clarified in Section 3. Section 4 explains classifiers utilized in the experiments. The experimental work is presented in Section 5. Finally, some concluding remarks are given in Section 6.

2. Feature Extraction

The features used in this study can be categorized as spectral, texture, and size/shape features. Detailed information about these features is given below.

2.1. Spectral Features

Mean values for each band (green, red, near infrared), brightness and normalized difference vegetation index (NDVI) can be listed as spectral features.

2.2. Texture Features

Texture features consist of three types of grey-level co-occurrence matrix (GLCM) and standard deviation values of three spectral bands (green, red, near infrared). GLCM texture calculations are calculated with NIR band of source image.

2.3. Size/Shape Features

The size/shape features used in this study can be listed as border index, shape index, area, round, compactness, length/width, rectangularity, density, asymmetry, and border length.

3. Feature Selection

In order to determine best performing feature subsets, three variations of a multivariate filter-based feature selection method are utilized in this study. The feature selection method used for the comparison is widely-known correlation-based feature subset selection method [22] consisting of a feature subset evaluator and a search method. The three variations of correlation-based feature subset selection method include best first search, genetic search, and greedy stepwise search. WEKA [23] software is used to utilize feature selection processes. These three methods are explained in the next subsections.

3.1. Best first search based feature selection (BFSFS)

BFSFS involves best first search (BFS) as a part of correlation-based feature subset selection method. BFS searches the space of attribute subsets by greedy hill climbing method with a backtracking facility [24, 25]. BFS may both search forward or backward depending on starting with the empty set of attributes or the full set of attributes. In the experiments, BFS is executed as forward search, in which input to the algorithm is an empty attribute subset. Moreover, the search termination criteria is set to 5, i.e. BFS will terminate the search after 5 consecutive backtracks.

3.2. Genetic search based feature selection (GSFS)

GSFS includes genetic search (GS) as a part of correlation-based feature subset selection method. GS is a suboptimal search method inspired from biological evolution process [26]. The main idea behind GS is the survival of the fittest solutions among potential

solutions for a specific problem. In the initialization phase of GS, it generally begins with a random sample of candidate attribute subsets which is also known as a population. New generations are obtained by applying the genetic operators, namely crossover and mutation, on these attribute subsets which can also be referred as chromosomes. The chromosomes are encoded with binary (0, 1) alphabet. While the indices in a chromosome represented with "1" indicate the selected attributes, the ones represented with "0" indicates attributes which are not selected. As an example, the chromosome {0 1 0 1 0 0 0 1} specifies that the 2nd, 4th, and 8th attributes are used while the others are disregarded. In the evaluation phase of GS, each chromosome in the population is evaluated using a fitness function which is a kind of success measure. A proportion of the existing population is selected to create new generations in each step in GS. This process is repeated until reaching a termination condition [27]. In the experiments, default parameters of Weka [23] software are used. Thus, the population size, crossover probability, mutation probability, maximum number of generations as termination condition was set as 20, 0.6, 0.033, and 20, respectively.

3.3. Greedy stepwise search based feature selection (GSSFS)

GSSFS involves greedy stepwise search (GSS) as a part of correlation-based feature subset selection method. GSS performs a greedy search either forward or backward through the space of attribute subsets [28]. The search can be both initialized with the empty set of attributes or the full set of attributes. Unlike BFS, GSS does not perform backtracking on the search space of attribute subsets. GSS stops when the addition or deletion of any attributes that remains results in a decrease in evaluation.

4. Classifiers

In order to investigate contributions of the selected features to the performance of classification, three different classification algorithms were employed. The first classifier is Random Forest [29] which is a non-linear classifier. The second one is linear support vector machine classifier [30]. The third and last classifier is a Bayesian network classifier [31].

4.1. Random Forest (RF)

Ensemble classification algorithms consist of multiple classifiers and they have an increased interest because of being more accurate than an individual classifier. RF is an ensemble classification method involving a combination of decision tree classifiers where each classifier is constructed using a random vector sampled independently from the training set [32]. For classification, each decision tree in a RF contributes with a single vote for determining the class label of an input vector. Then, the output of the RF classifier is determined by a kind of majority voting technique. RF can handle high dimensional data.

4.2. Support Vector Machines (SVMs)

SVM aims to find a hyperplane that successfully separates the samples into two classes. In order to make an effective separation, it is necessary to find a decision boundary that minimizes misclassifications [33]. The essential point of SVM is the margin concept [29] where margin is the distance of closest samples from the decision boundary. The main objective of SVM is to find the appropriate hyperplane which maximizes the margin. For this purpose, it is necessary to detect support vectors, which are the data points that lie at the border between the two classes. SVM can be either linear or nonlinear classifier according to its kernel type and the widely-known kernel functions are linear, polynomial, radial basis function and sigmoid kernels [34]. In this study, SVM with linear kernel is employed in the experiments.

4.3. Bayesian network (BN)

BN is a directed acyclic graph having a conditional probability distribution for each node [35]. While each node in this graph represents an attribute, each arc between these nodes represents a probabilistic dependency. A BN can be used to compute the conditional probability of an attribute using the values assigned to the other attributes. Once a directed acyclic graph has been constructed, the joint probability of any particular instantiation of all n variables in a BN can be calculated as follows:

$$P(x_1,...,x_n) = \prod_{i=1}^{n} p(x_1 \mid x_n) \quad (1)$$

where X_i represents the instantiation of the variable X_i and π_i represents the instantiation of the parents of X_i [36]

5. Experimental Work

In the experimental work, an in-depth investigation is carried out to analyze the impact of three different variations of correlation based feature selection method for urban land cover classification. Experimental settings including the utilized dataset and classification algorithms are first briefly described. Then, the profile of selected features by three different feature selection methods and their corresponding accuracy scores are provided.

5.1. Settings

The dataset used in the experiments is obtained from a recent remote sensing study whose aim is to classify a high resolution aerial image into 9 types of urban land cover [15]. This high resolution image data is collected from an area around the city of Deerfield Beach, Florida. Urban land cover types in the dataset are concrete, shadows, trees, asphalt, buildings, grass, pools, cars, and soil. The dataset originally contains separate train and test splits. However, in this study, these splits are explicitly combined in order to make a fair evaluation by applying 10-fold cross-validation technique. The dataset currently consist of 675 samples and 147 features. The class distribution of the dataset is indicated in Table 1.

Table 1. Class distributions of the dataset

No	Class label	Distribution of samples
1	Asphalt	8.7 %
2	Building	18.1 %
3	Car	5.3 %
4	Concrete	17.2 %
5	Grass	16.6 %
6	Pool	4.3 %
7	Shadow	9 %
8	Soil	5 %
9	Tree	15.7 %

21 feature types used in the experiments are explained in Section 2. However, it is also necessary to note that these features are calculated according to seven different scale parameters changing between 20 and 140 at an interval of 20 on segmentation process. Consequently, the numbers of features used in the study were 147 in spite of the number of feature types is 21.

Weka software is used in the experiments in order to evaluate feature selection methods and classification algorithms. The experiments consist of two parts. In the first part, RF, SVM, and

BN classifiers are built using this dataset. These classifiers consider all features in the dataset. The first part is performed in order to see the performance of the classifiers without feature selection. In the second part, feature selection was performed on the urban land cover dataset using three different methods described in Section 3. The resulting feature subsets were used to build BN, RF and SVM classifiers. Traditional F-Score metric was used for evaluation in all of the experiments. Then, a comparison was carried out between all cases in order to see the best performing setting and the impact of feature selection methods on urban land cover classification.

5.2. Profiles of reduced feature sets

In this subsection, the profiles of reduced feature sets are analyzed for three feature selection methods. Table 2 shows the ratio of the selected features among all. In this table, the column FS refers the feature selection method employed, SRF refers the size of reduced feature subset and TNF refers the total number of features. The bold ids in the table indicate the common features that are selected by all of the three methods.

Table 2. Selected features for BFSFSS, GSFS and GSSFS methods

FS	SRF / TNF	Indices of the selected features
BFSFS	29 / 147	**3 5 7 8 9 10 11 20 22** 26 **29 30 31 41** 42 **43** 49 59 **62 70 83 84 92** 94 95 **104 112 125 134**
GSFS	57 / 147	**3 5 7 8 9 10 11** 18 19 **20** 21 **22** 23 25 26 **29 30 31** 33 34 36 **41 43** 44 45 47 50 51 52 53 56 **62** 68 69 **70** 71 72 73 76 80 **83 84** 89 **92** 93 103 **104** 110 **112** 113 118 **125** 132 **134** 136 146 147
GSSFS	28 / 147	**3 5 7 8 9 10 11 20 22 29 30 31 41** 42 **43** 49 59 **62 70 83 84 92** 94 95 **104 112 125 134**

As shown in Section 2, features in the dataset can be grouped into size/shape, texture and spectral. In this subsection, features listed in Table 2 were also analyzed according to their membership to these groups. As an example, for BFSFS, the ratios of size/shape, texture and spectral features is 24%, 17% and 59%, respectively. For GSFS, the ratios of size/shape, texture and spectral features is 30%, 18%, and 52%, respectively. Similarly in GSSFS, the ratios of those features are 25%, 18% and 57% respectively. In the light of this information, it can be said that the contribution of spectral features is more than size/shape and texture features. Also, the contribution of size/shape features are more than texture features. It is necessary to note that these findings are valid for all of the three feature selection methods. Apart from these, the ratio of common features for BFSFS, GSFS, and GSSFS are 79%, 40%, and 82%, respectively.

5.3. Accuracy Analysis

In this subsection, individual performances of three classifiers and contribution of feature selection methods to the classification performance were extensively analyzed. Table 3-5 shows the corresponding F-Scores for RF, SVM, and BN classifiers, respectively. In the tables, the first column represents the performance scores when feature selection is not applied. Then the other columns show the performance scores when BFSFS, GSFS, and GSSFS methods were utilized for feature selection, respectively. In the experiments, as mentioned before, 10 fold cross validation was used for fair evaluation.

Table 3. F-Scores for RF classifier

Class	-	BFSFS	GSFS	GSSFS
Concrete	0.838	0.871	0.869	0.856
Shadow	0.932	0.912	0.894	0.919
Tree	0.851	0.861	0.841	0.879
Asphalt	0.894	0.891	0.86	0.902
Building	0.823	0.863	0.869	0.86
Grass	0.856	0.852	0.83	0.877
Pool	0.846	0.964	0.909	0.926
Car	0.817	0.889	0.889	0.93
Soil	0.556	0.75	0.733	0.765
w.avg.	0.839	0.869	0.856	**0.876**

Table 4. F-Scores for SVM classifier

Class	-	BFSFS	GSFS	GSSFS
Concrete	0.702	0.815	0.784	0.787
Shadow	0.855	0.85	0.85	0.85
Tree	0.731	0.756	0.806	0.81
Asphalt	0.821	0.846	0.86	0.855
Building	0.724	0.774	0.797	0.764
Grass	0.684	0.768	0.765	0.797
Pool	0.755	0.893	0.857	0.893
Car	0.761	0.806	0.806	0.806
Soil	0.413	0.743	0.694	0.743
w.avg.	0.723	0.796	0.799	**0.803**

Table 5. F-Scores for BN classifier

Class	-	BFSFS	GSFS	GSSFS
Concrete	0.834	0.872	0.827	0.877
Shadow	0.866	0.917	0.893	0.909
Tree	0.814	0.829	0.838	0.829
Asphalt	0.903	0.909	0.867	0.917
Building	0.84	0.877	0.825	0.88
Grass	0.766	0.821	0.823	0.821
Pool	0.842	0.966	0.893	0.966
Car	0.88	0.919	0.919	0.919
Soil	0.545	0.73	0.641	0.73
w.avg.	0.818	0.864	0.835	**0.866**

According to Table 3-5, feature selection improved the performance of classification and the best F-Scores were obtained when GSSFS feature selection method is employed for all of the three classifiers. The best performance obtained for RF, SVM, and BN classifiers were 0.876, 0.803, and 0.866, respectively. According to the tables, one can note that the best performing setting was the case that RF classifier is used and GSSFS feature selection method was employed. The runner-up best performing feature selection method is BFSFS except one case that the SVM classifier was employed. The increase of performance in terms of F-scores was 0.037, 0.08, and 0.048 for RF, SVM, and BN, respectively. Thus, the impact of feature selection for SVM classifier was more than the RF and BN classifiers. Besides, according to the class-based F-Sores, the best performance of classes changes for different settings. However, the recognition ratio of soil class is the worst one among all.

6. Conclusion

In this study, the impact of feature selection was extensively analyzed using three different variations of a multivariate filter-based feature selection method and various widely-known classifiers. For this analysis, a recently published dataset including 147 features was employed. Experiments were realized using 10-fold cross validation and the success measure used in the experiments were widely-known F-Score. According to the experimental results, it can be said that feature selection improves the performance of classification for all cases and RF classifier was more successful than BN and SVM classifiers. Also, if the profiles of reduced feature sets are investigated, it can be easily seen than

the contribution of spectral features to the performance of classification is more than size/shape and texture features. As a future work, the impact of feature selection may also analyzed for various types of remote sensing tasks that there is limited number of studies as urban land cover classification.

References

[1] X. Tong, H. Xie, and Q. Weng, "Urban land cover classification with airborne hyperspectral data: What features to use?," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 10, pp. 3998-4009, 2014.

[2] S. Kandrika and P. S. Roy, "Land use land cover classification of Orissa using multi-temporal IRS-P6 awifs data: A decision tree approach," *International Journal of Applied Earth Observation and Geoinformation*, vol. 10, no. 2, pp. 186-193, 2008.

[3] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 277-293, 2017.

[4] P. Zope, T. Eldho, and V. Jothiprakash, "Impacts of land use–land cover change and urbanization on flooding: a case study of Oshiwara River Basin in Mumbai, India," *Catena*, vol. 145, pp. 142-154, 2016.

[5] W. Y. Yan, A. Shaker, and N. El-Ashmawy, "Urban land cover classification using airborne LiDAR data: A review," *Remote Sensing of Environment*, vol. 158, pp. 295-310, 2015.

[6] Q. Yu, P. Gong, N. Clinton, G. Biging, M. Kelly, and D. Schirokauer, "Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 7, pp. 799-811, 2006.

[7] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS journal of photogrammetry and remote sensing*, vol. 65, no. 1, pp. 2-16, 2010.

[8] R. C. Weih and N. D. Riggan, "Object-based classification vs. pixel-based classification: Comparative importance of multi-resolution imagery," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, no. 4, p. C7, 2010.

[9] T. Blaschke, C. Burnett, and A. Pekkarinen, "Image segmentation methods for object-based analysis and classification," in *Remote sensing image analysis: Including the spatial domain*: Springer, 2004, pp. 211-236.

[10] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote sensing of environment*, vol. 115, no. 5, pp. 1145-1161, 2011.

[11] N. Thomas, C. Hendrix, and R. G. Congalton, "A comparison of urban mapping methods using high-resolution digital imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 9, pp. 963-972, 2003.

[12] D. C. Duro, S. E. Franklin, and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sensing of Environment*, vol. 118, pp. 259-272, 2012.

[13] J. Aguirre-Gutiérrez, A. C. Seijmonsbergen, and J. F. Duivenvoorden, "Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in Mexico," *Applied Geography*, vol. 34, pp. 29-37, 2012.

[14] T. Rittl, M. Cooper, R. Heck, and M. Ballester, "Object-based method outperforms per-pixel method for land cover classification in a protected area of the Brazilian Atlantic rainforest region," *Pedosphere*, vol. 23, no. 3, pp. 290-297, 2013.

[15] B. Johnson and Z. Xie, "Classifying a high resolution image of an urban area using super-object information," *ISPRS journal of photogrammetry and remote sensing*, vol. 83, pp. 40-49, 2013.

[16] M. N. Jebur, H. Z. Mohd Shafri, B. Pradhan, and M. S. Tehrany, "Per-pixel and object-oriented classification methods for mapping urban land cover extraction using SPOT 5 imagery," *Geocarto International*, vol. 29, no. 7, pp. 792-806, 2014.

[17] M. Bochow, H. Taubenböck, K. Segl, and H. Kaufmann, "An automated and adaptable approach for characterizing and partitioning cities into urban structure types," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, 2010, pp. 1796-1799: IEEE.

[18] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2297-2307, 2010.

[19] Y. Maghsoudi, M. J. Valadan Zoej, and M. Collins, "Using class-based feature selection for the classification of hyperspectral data," *International journal of remote sensing*, vol. 32, no. 15, pp. 4311-4326, 2011.

[20] G. Myburgh and A. Van Niekerk, "Effect of feature dimensionality on object-based land cover classification: A comparison of three classifiers," *South African Journal of Geomatics*, vol. 2, no. 1, pp. 13-27, 2013.

[21] B. Banerjee, A. Bhattacharya, and K. M. Buddhiraju, "A generic land-cover classification framework for polarimetric SAR images using the optimum Touzi decomposition parameter subset—An insight on mutual information-based feature selection techniques," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1167-1176, 2014.

[22] M. A. Hall, "Correlation-based feature selection for machine learning," 1999.

[23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10-18, 2009.

[24] A. K. Pandey, P. Pandey, K. Jaiswal, and A. K. Sen, "DataMining Clustering Techniques in the Prediction of Heart Disease using Attribute Selection Method," *heart disease*, vol. 14, pp. 16-17, 2013.

[25] M. D. Patil and D. S. S. Sane, "Effective Classification after Dimension Reduction: A Comparative Study," *International Journal of Scientific and Research Publications*, vol. 4, no. 7, p. 1, 2014.

[26] A. K. Uysal and S. Gunal, "Text classification using genetic algorithm oriented latent semantic features," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5938-5947, 2014.

[27] D. E. Goldberg, "Genetic algorithms in search, optimization, and machine learning, 1989," *Reading: Addison-Wesley*, 1989.

[28] R. O. Mujalli and J. De ONa, "A method for simplifying the analysis of traffic accidents injury severity on two-lane

highways using Bayesian networks," *Journal of safety research*, vol. 42, no. 5, pp. 317-326, 2011.

[29] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*. Academic Press, 2008, p. 900.

[30] C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of remote sensing*, vol. 23, no. 4, pp. 725-749, 2002.

[31] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine learning*, vol. 29, no. 2-3, pp. 131-163, 1997.

[32] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217-222, 2005.

[33] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247-259, 2011.

[34] T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, no. 5, pp. 352-359, 2009.

[35] J. Cheng and R. Greiner, "Comparing Bayesian network classifiers," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, 1999, pp. 101-108: Morgan Kaufmann Publishers Inc.

[36] P. Larrañaga, M. Poza, Y. Yurramendi, R. H. Murga, and C. M. H. Kuijpers, "Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters," *IEEE transactions on pattern analysis and machine intelligence*, vol. 18, no. 9, pp. 912-926, 1996.