⋆Miner: A Suit of Classifiers for Spatial, Temporal, Ancillary, and Remote Sensing Data Mining

Ranga Raju Vatsavai[1,2], Shashi Shekhar[1]
[1]Department of Computer Science and Engineering, University of Minnesota
EE/CS 4-192, 200 Union Street. SE., Minneapolis, MN 55455.
{vatsavai|shekhar}@cs.umn.edu

Thomas E. Burk[2]
[2] Remote Sensing and Geospatial Analysis Laboratory, Department of Forest Resources,
University of Minnesota. 115, Green Hall, 1530 N. Cleveland Ave, St. Paul 55108.
teb@mallit.fr.umn.edu

Abstract

Thematic classification of multi-spectral remotely sensed imagery for large geographic regions requires complex algorithms and feature selection techniques. Traditional statistical classifiers rely exclusively on spectral characteristics, but thematic classes are often spectrally overlapping. The spectral response distributions of thematic classes are dependent on many factors including terrain, slope, aspect, soil type, and atmospheric conditions present during the image acquisition. With the availability of geo-spatial databases, it is possible to exploit the knowledge derived from these ancillary geo-spatial databases to improve the classification accuracies. However, it is not easy to incorporate this additional knowledge into traditional statistical classification methods. On the other hand, knowledge-based and neural network classifiers can readily incorporate these spatial databases, but these systems are often complex to train and their accuracy is only slightly better than statistical classifiers. In this paper we present a new suit of classifiers developed through NASA funding, which addresses many of these problems and provide a framework for mining multi-spectral and temporal remote sensing images guided by geo-spatial databases.

1 Introduction

Land management organizations and the public have a need for more current regional land cover information to manage resources and monitor land cover change. Remote sensing, which provides inexpensive, synoptic-scale data with multi-temporal coverage, has proven to be very useful in land cover mapping, environmental monitoring, and forest and crop inventory. Several classification algorithms have been proposed in the literature for analysis of remote sensing imagery. These algorithms can be broadly grouped into two categories, supervised and unsupervised, based on the learning scheme used. Among supervised classification methods, the maximum likelihood classifier is the most extensively studied and utilized for classification of multi-spectral images. Unsupervised methods include various clustering algorithms such as ISODATA and k-means. Clustering algorithms are generally used for initial processing to understand natural groupings in the data and to aid in supervised learning. Most of these methods work well if the land cover classes are spectrally separable and classification model assumptions are true. But in reality, the classes under investigation are often spectrally overlapping, and many times the classification model assumptions, such as, samples are drawn from multivariate normal distributions and that they are independent and identically distributed (i.i.d.) are not valid. The spectral response distribution of classes are dependent on many factors including terrain, slope, aspect, soil type and moisture content, and atmospheric conditions. As a result, any classification method based on spectral data alone will fail to capture the full essence of the problem. Similarly, often the neighboring pixels are spatially correlated, which invalidates the i.i.d. assumption. Under the NASA funded TerraSIP [15] project, we have investigated and developed a suit of new classification algorithms, collectively named as, *Miner. We have evaluated these algorithms on various study sites across Upper Great Lakes States, and observed that many of these new classifiers shows an 8 to 10% improvement in overall classification accuracy.

The rest of this paper is organized as follows: Section 2 provides the overall architecture of the *Miner; Section 3 provides knowledge based classification system; Section 4 provides an hybrid classification system; Section 5 provides a brief description of two contextual classifiers; and Section 6 provides brief conclusions and suggests directions for future research.

2 ⋆Miner Architecture

The proposed system consists of several components, built from commercial, public domain and in-house developed software. Major modules were shown in Figure 1. Preprocessing module consists of geometric correction and various band ratioing functions like NDVI (normalized density vegetation index), PCA (principle component analysis etc. Geometric correction module is used to correct geometric errors and to establish a one-to-one correspondence between images and geo-spatial databases. That is once images are geometrically corrected and reprojected to a common projection system, we can easily compute the coordinates of one dataset given the coordinates of other dataset. This is essential for hybrid classification system, as we are using the geo-spatial databases to guide the remote sensing classification. Band ratios are helpful to enhance certain structures in the image and PCA is used to reduce dimensionality of multi-spectral images from seven channels to 2 or 3 channels. The ⋆Miner system is also integrated with the popular machine learning system, Weka [20], which provides support for standard classifiers, such as, decision trees and neural networks.

3 SSTKC: Spectral, Spatial, and Temporal Knowledge-based Classifier

Several recent studies have focused on incorporating ancillary information into the classification process. The most notable approaches are neural networks, expert (knowledge based, rule based) systems, and the maximum likelihood classifiers (MLC) with *a priori* knowledge. Ancillary layers can be directly incorporated into neural network learning, as opposed to maximum likelihood classification, since neural networks do not assume any underlying probability distribution of data. In general neural networks perform as good as MLC or even better in some cases. However, neural network training and the establishment of suitable parameters are difficult and the neural network approach does not offer any significant advantages over conventional classification schemes at the forest type level [13]. Knowledge based systems (KBS) offer a flexible framework for incorporating ancillary spatial data into the classification process. The main issue associated with KB systems is the development of rules.

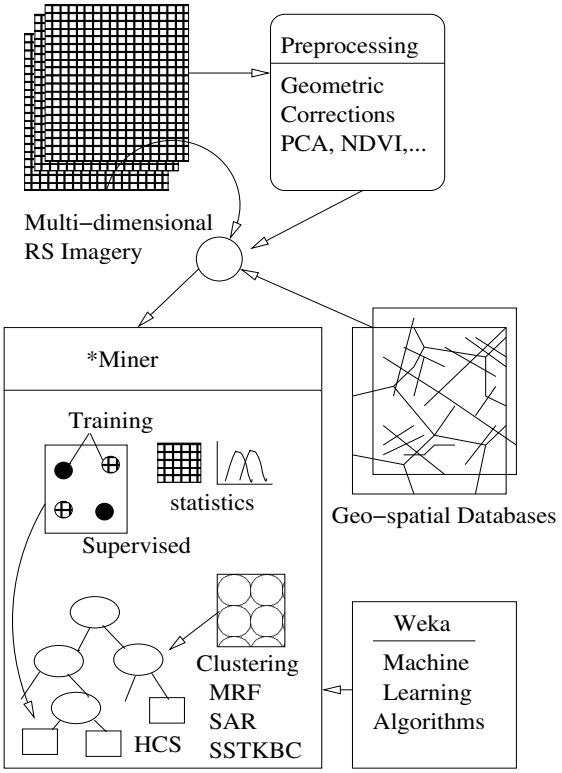


Figure 1. ⋆Miner Architecture

These limitations have led us to investigate a new approach - a fusion of KBS and MLC for classification of multi-spectral remote sensing imagery utilizing knowledge derived from ancillary spatial databases. This approach minimizes the limitation of KB by simplifying the rule-base. In this simplified approach, the rule-base is used to stratify the image into homogeneous regions rather than classifying individual pixels. The stratified regions minimize the overlap among the classes and thus provide a robust environment for MLC. This paper presents the key features of the new classification approach and the initial results. Our classification system consists of four major modules: spectral knowledge, spatial knowledge, semi-automated learning, and classification, each of which is described below.

Spectral Knowledge: Object extraction from spectral relationships only is almost impossible, nonetheless it is interesting and useful to find simple spectral rules, like: $\forall Pixel(p),\ IF(band1(p) > band2(p) > band3(p))\ THEN\ Output(p) =$ 'WATER'. Even though finding such rules is difficult, the main contribution of spectral knowledge is in finding inherent data structures within the image. Often transformations, like normalized density vegetation index (NDVI) and Tasseled Cap (TC),

will yield more insights into the structure of the data. The Tasseled Cap concept involves identifying the existing data structures for a particular sensor and application and changing the viewing perspective such that those data structures can be viewed directly [6]. We have extracted spectral knowledge derived from greenness index channel of the TC transformation for stratifying the TM image. The rules used are summarized in Table 1.

Knowledge Base	Class/Region
TM B1 > B2 > B3 > B4	Water
$(TassledCap.Greenness \leq 15)$ && $(Pop.Density > 5000)$ || $(Road.Density > 0.0145)$	High density developed
$(15 > TassledCap.Greenness \leq 25)$ && $(1000 \leq Pop.Density\, leq 5000)$ || $(0.0078 \leq Road.Density \leq 0.0145$	Low density developed

Table 1. Spectral and Spatial Knowledge Base.

Spatial Knowledge: The purpose of the spatial knowledge base is to stratify the TM image into homogeneous regions with the following properties:

Let R be any given image.
The purpose of image stratification is to find a finite set of regions $R_1, R_2, ..., R_q$, such that
$R = \bigcup_{i=1}^{q} R_i, \quad R_i \cap R_j = \emptyset$
and k classes $C_{ik}\ \varepsilon\ R_i$ are spectrally separable (i.e. inter-class variation is minimum and intra-class variation is maximum).

Our objective is to find regions in such a way that signature continuity holds within any region R_i and for any class: if $C_{rk} = C_{jk}$, then r = j. But in practice we may not find such regions, so there may be some common classes among the regions. In the training phase we have to collect sufficient samples for overlapping classes to avoid artificial contours in the final classified image. Careful study of the TM image shows that we can find two distinct regions called 'developed lowlands' and 'undeveloped uplands'. The flow chart for extracting spectral and spatial knowledge to derive these two regions is shown in Figure 2. The knowledge base is summarized in Table 1.

Semi-automated Learning: Sample plots were collected for the required classes 'to be used as seed points'. A region growing algorithm was applied at each of the seed points to populate polygons with homogeneous characteristics. Approximately 25 aerial photographs and additional ground truth observations were utilized in collecting sample plots.

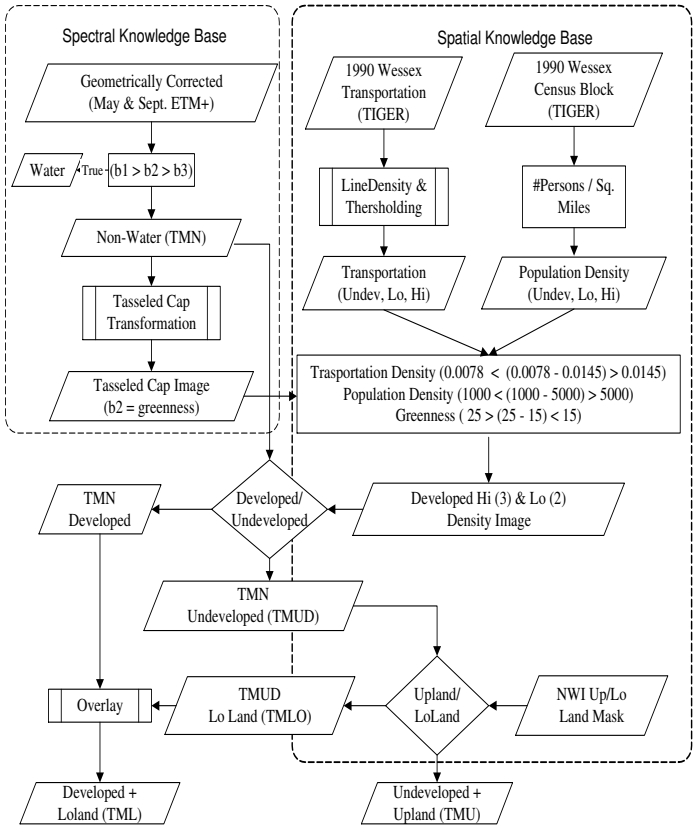


Figure 2. Flow chart showing the use of spectral and spatial knowledge base

The main criterion used in region growing was obtaining minimum region of N+1 pixels (where N is the number of spectral bands) to 25 pixels within a spectral Euclidean distance of 10 pixels. For a N-dimensional multi-spectral space, we need at least N+1 pixels to avoid a singular covariance matrix. We chose the 25-pixel criterion to check that the sample comes from a homogeneous area. We can't compute this threshold beforehand, so we have iteratively varied the thresholds to reach an optimum limit satisfying the above criteria and eliminated some of the seed points during this process. Once the training samples were collected, training statistics were generated and analyzed both visually and quantitatively to check the between-class separability. Co-spectral, ellipsoidal plots in two-dimensional feature space provide first-hand visual information about between-class separability. In all our experimental studies, the training samples with low TD values were carefully studied and either merged or deleted based on ground truth verification. The flow chart for supervised learning is shown in Figure 3.

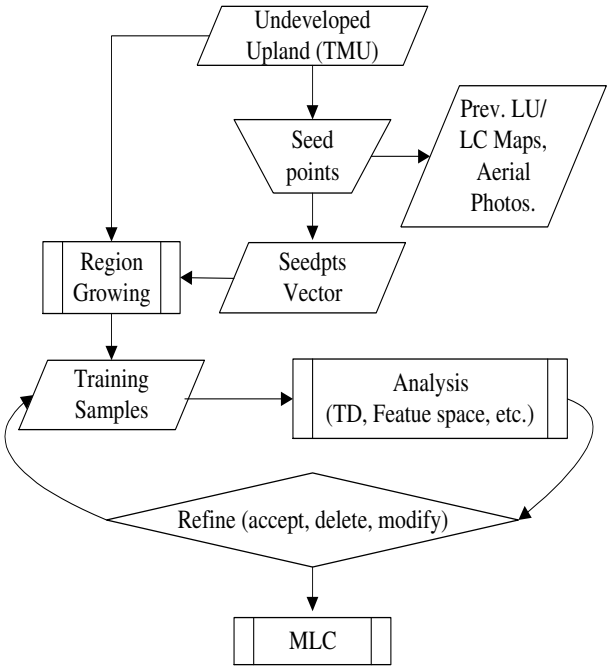


Figure 3. Semi-automated supervised learning scheme

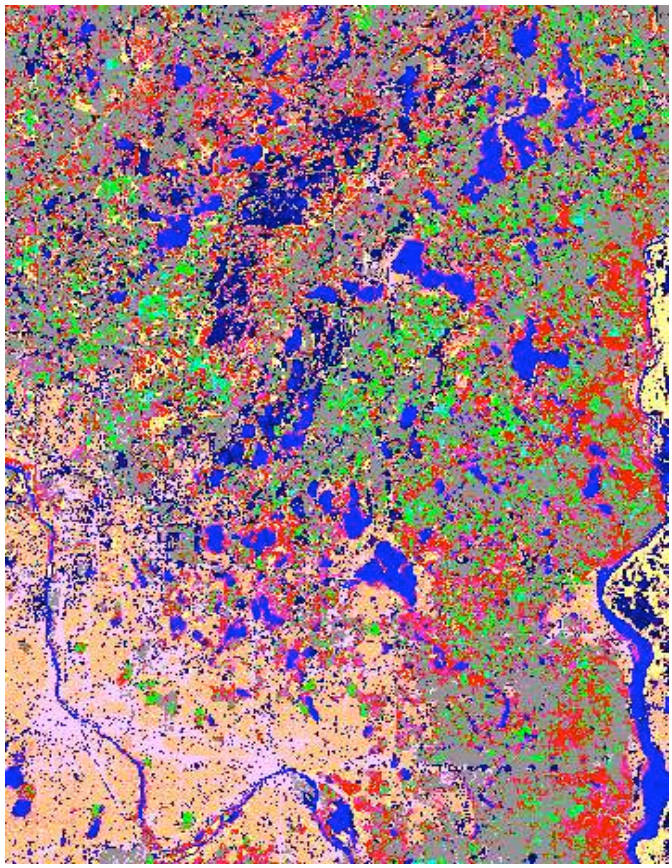


Figure 4. Final Classified Image

Classification: Classification is performed using MLC with the following discriminant function:

$$g_i(x) = -ln|\Sigma_i| - (x - m_i)^t \Sigma_i^{-1}(x - m_i), \qquad (1)$$

where m_i and Σ_i are the mean vector and covariance matrix of the training data for class ω_i. Any given pixel vector x is assigned to ω_i if $g_i(x) > g_j(x) \ \ \forall j \neq i$. More details on transformed divergence and maximum likelihood classification methods can be found in [7], [11]. All classified regions are merged to obtain a final classified image as shown in Figure 4.

4 HCS: Hybrid Classification System

It is found in several previous studies [3, 10], that the classification accuracies can be improved by combining various classifier, the approach commonly being known as, MCS or Ensemble learning. The essence of MCS is to combine classifications from the Ensemble learned on the same training data set (or various subsets generated using Bagging or Boosting algorithms) to produce the final classification. In our HCS approach, the objective is to combine statistical pattern recognition methods and knowledge based classification systems by reducing the complexity of both systems. We used unsupervised clustering algorithms to overcome the need for accurate training data in the first stage. Several clustering algorithms have proven to be very useful in remote sensing analysis. However, the main problem with the clustering algorithms is to map these spectral clusters into ground classes. There is no simple way to do this, so the clustering approach is mainly used to find natural cluster in the data and then take the training samples from these clusters. However, we observe that ancillary geo-spatial data can be efficiently used to map these spectral clusters into ground classes. So in the second stage a decision tree is used to guide the classification process using ancillary geo-spatial data. This framework offers greater flexibility than knowledge based classification approaches, because we do not need a complex knowledge base. Here the requirement is to map the entire cluster rather than individual pixel as in traditional approach. So, HCS is essentially a classification sequence, that is, $HCS = f_n(f_{n-1}(f_{n-2}(\ldots(f_1))))$, where f_i is some classifier from the selected ensemble. It should be noted that automatically finding such an HCS is very difficult, one can use domain knowledge (or heuristics) to come up with an efficient sequence.

We now briefly explain the overall classification system

that we have prototyped and tested. The major components of the system were shown in Figure 5. Each of the components were described in the following sub-sections.

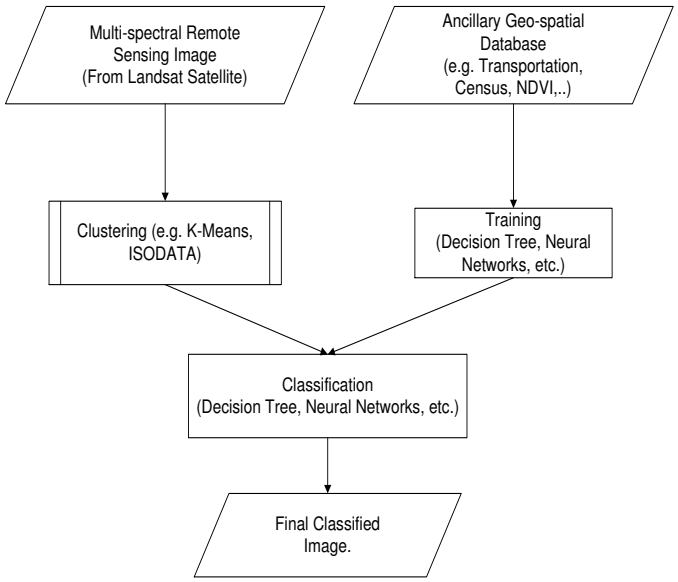


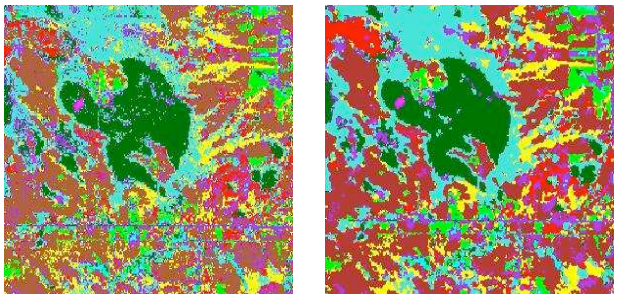
Figure 5. Flow diagram for new hybrid classification system

Clustering Using the ISODATA clustering algorithm we have obtained 20 spectral clusters on one of the Twin-cities Landsat image. Small clusters (with size < 5 pixels) were further eliminated using the "clump analysis" and the resulting clustered image was converted into polygonal vector layer.

Decision Tree Construction About 180 training samples were collected through random sampling. At each plot center the training features were collected from each of the geospatial data layer (road density, National Wetland Inventory map, Census data, ...) and the clustered image. We used standard C4.5 decision tree learning algorithm for mapping spectral clusters into thematic classes. We have trained the decision tree classifier with 90% accuracy. The final pruned tree had 29 leaves with a maximum depth of 9. This tree was then used to classify the clustered image into ten thematic classes with an overall accuracy of 86%.

5 Spatial (contextual) Classifiers

Traditional data mining algorithms [1] often make assumptions (e.g. independent, identical distributions) which violate Tobler's first law of Geography: everything is related to everything else but nearby things are more related than distant things [16]. In other words, the values of attributes of nearby spatial objects tend to systematically affect each other. In spatial statistics, an area within statistics devoted to the analysis of spatial data, this is called spatial autocorrelation [5]. Knowledge discovery techniques which ignore spatial autocorrelation typically perform poorly in the presence of spatial data. The simplest way to model spatial dependence is through spatial covariance. Often the spatial dependencies arise due to the inherent characteristics of the phenomena under study, but in particular they arise due to the fact that imaging sensors have better resolution than object size. For example, remote sensing satellites have resolutions ranging from 30 meters (e.g., Enhanced Thematic Mapper of Landsat 7 satellite of NASA) to one meter (e.g., IKONOS satellite from SpaceImaging), while the objects under study (e.g., Urban, Forest, Water) are much bigger than 30 meters. As a result, the per-pixel-based classifiers, which do not take spatial context into account, often produce classified images with *salt and pepper* noise. These classifiers also suffer in terms of classification accuracy.



(a) MLC (b) MRF

Figure 6. Small portion from the NW corner of the Carleton image. (a) MLC, (b) MRF

There are two major approaches for incorporating spatial dependence into classification/prediction problems. They are spatial auto-regression models (SAR) [2, 9], and Markov Random Field models (MRF) [4, 14]. The computational complexity of SAR is very high and its almost impossible to apply the exact solution to large data sets such as satellite images. We have implemented efficient approximate solutions of SAR [8] and provided both theoretical and experimental comparisons [12]. Figure 6 shows the differences between standard maximum likelihood classifier (MLC) and MRF. Our experimental results show that the contextual classifier improves the overall classification accuracy by more than 5% compared to base classifier (e.g., MLC), but more importantly the contextual classifiers remove "salt and pepper" noise which is hard to quantify through the overall classification accuracy measure.

6 Conclusions and Future Directions

We have developed and tested a suit of new classification algorithms that exploit additional knowledge (spectral, geo-spatial, temporal, contextual) for efficient mining of remote sensing imagery. We have developed a simplified knowledge base derived from multi-spectral images and ancillary spatial databases. Such knowledge base is useful in stratifying the image into non-overlapping (spectral) regions, so that classes in each region are more easily separable. Similarly, we developed several heuristics to build hybrid classification systems, which have computational advantages as compared to standard MCS or ensemble methods. These new algorithms not only showed an improved (classification) performance on various study sites, but have potential to scale to large scale remote sensing data mining. More details of these classification methods can be found in [17, 18, 8, 12]. In addition, our ⋆Miner is integrated with Weka, which provides a plethora of standard machine learning algorithms. We are planning to release ⋆Miner as an open source system so that large community of remote sensing users can be benefited. In addition, we are investigating semi-supervised learning methods which works with partially labeled training data sources [19].

7 Acknowledgments

This research has been supported through cooperative agreement with NASA (NCC 5316) and by the University of Minnesota Agriculture Experiment Station project MIN-42-044. Support in part is provided by the Army High Performance Computing Research Center under the auspices of Department of the Army, Army Research Laboratory Cooperative agreement number DAAH04-95-2-0003/contract number DAAH04-95-C-0008.

We would like to thank our collaborators Barry T. Wilson, Perry Nacionales, Jaime Smedsmo, Tim Mac, Uygar Ozesmi, Maria Gini, Weili Wu, Baris Kazar, and RSL and SDBMS research group members for their useful comments and help during this research.

References

[1] R. Agrawal. Tutorial on database mining. In *Thirteenth ACM Symposium on Principles of Databases Systems*, pages 75–76, Minneapolis, MN, 1994.

[2] L. Anselin. *Spatial Econometrics: methods and models*. Kluwer, Dordrecht, Netherlands, 1988.

[3] Bauer and R. Kohavi. The tasseled cap de-mystified. *An empirical comparison of voting classification algorithms: Bagging, boosting, and variants*, 36(1-2), 1999.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast Approximate Energy Minimization via Graph Cuts . *International Conference on Computer Vision*, September 1999.

[5] N. Cressie. *Statistics for Spatial Data (Revised Edition)*. Wiley, New York, 1993.

[6] E. Crist and R. J. Kauth. The tasseled cap de-mystified. *Photogrammetric Engineering & Remote Sensing*, 52(1):81–86, January 1986.

[7] J. R. Jensen. *Introductory Digital Image Processing, A Remote Sensing Perspective*. Prentice Hall, Upper Saddle River, NJ-07458, 1996.

[8] B. M. Kazar, S. Shekhar, D. J. Lilja, R. R. Vatsavai, and R. K. Pace. Comparing exact and approximate spatial auto-regression model solutions for spatial data analysis. In *GI-Science*, pages 140–161, 2004.

[9] J. LeSage. Regression Analysis of Spatial data. *The Journal of Regional Analysis and Policy (Publisher: Mid-Continent Regional Science Association and UNL College of Business Administration)*, 27(2):83–94, 1997.

[10] D. Opitz and R. Maclin. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11:169–198, 1999.

[11] J. A. Richards and X. Jia. *Remote Sensing Digital Image Analysis*. Springer, New York, 1999.

[12] S. Shekhar, P. Schrater, R. Vatsavai, W. Wu, and S. Chawla. Spatial contextual classification and prediction models for mining ge ospatial data. *IEEE Transaction on Multimedia*, 4(2):174–188, 2002.

[13] A. Skidmore, B. Turner, W. Brinkhof, and E. Knowles. Performance of a neural network: Mapping forest using gis and remorely sensed data. *Photogrammetric Engineering & Remote Sensing*, 63(5):501–514, May 1997.

[14] A. H. Solberg, T. Taxt, and A. K. Jain. A Markov Random Field Model for Classification of Multisource Satel lite Imagery. *IEEE Transaction on Geoscience and Remote Sensing*, 34(1):100–113, 1996.

[15] TerraSIP. Terrasip introduction. http://terrasip.gis.umn.edu/.

[16] W. Tobler. *Cellular Geography, Philosophy in Geography*. Gale and Olsson, Eds., Dordrecht, Reidel, 1979.

[17] R. R. Vatsavai, T. E. Burk, P. V. Bolstad, M. E. Bauer, S. K. Hansen, T. Mack, J. Smedsmo, and S. Shekhar. Multi-spectral image classification using spectral and spatial knowledge. In *CISST, 2001*.

[18] R. R. Vatsavai, T. E. Burk, S. Shekhar, and M. Gini. An efficient hybrid classification system for mining multi-spectral remote sensing imagery guided by spatial databases. In *2nd Pattern Recognition of Remote Sensing Workshop*, 2002.

[19] R. R. Vatsavai, S. Shekhar, and T. E. Burk. A semi-supervised learning method for remote sensing data mining. In *ICTAI*, pages 207–211, 2005.

[20] Weka. Weka machine learning project. http://www.cs.waikato.ac.nz/ ml/index.html.