

A Machine Learning Approach to Automated Construction of Knowledge Bases for Expert Systems for Remote Sensing Image Analysis With GIS Data (U)

by

H. E. Mackey

Westinghouse Savannah River Company

Savannah River Site

Aiken, South Carolina 29808

X. Huang

University of South Carolina

SC USA

J. R. Jensen

University of South Carolina

SC USA

A document prepared for ANNUAL MEETING OF THE AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING at Baltimore from 04/21/96 - 04/26/96.

DOE Contract No. DE-AC09-89SR18035

This paper was prepared in connection with work done under the above contract number with the U. S. Department of Energy. By acceptance of this paper, the publisher and/or recipient acknowledges the U. S. Government's right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper, along with the right to reproduce and to authorize others to reproduce all or part of the copyrighted paper.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED **MASTER**

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

WSRC-MS-95-0

**A MACHINE LEARNING APPROACH TO AUTOMATED CONSTRUCTION
OF KNOWLEDGE BASES FOR EXPERT SYSTEMS FOR
REMOTE SENSING IMAGE ANALYSIS WITH GIS DATA**

by

Xueqiao Huang and John R. Jensen
Department of Geography
University of South Carolina
Columbia, SC 29208

and

Halkard E. Mackey, Jr.
Westinghouse Savannah River Company
Savannah River Technology Center
Environmental Sciences Section
Aiken, SC 29802

UNCLASSIFIED

DOES NOT CONTAIN
UNCLASSIFIED CONTROLLED
NUCLEAR INFORMATION

ADC 2.

Reviewing

Official:

DeMone - Shelton

(Name and Title)

Date:

May 11/21/95

For publication in the Proceedings of the Annual Meeting of the American Society for
Photogrammetry and Remote Sensing, April 1996, Baltimore, Maryland

This information contained in this manuscript was developed during the course of work
under Contract No. DE-AC09-89SR18035 with the US Department of Energy.

WSRC-MS-95-0

**A MACHINE LEARNING APPROACH TO AUTOMATED CONSTRUCTION
OF KNOWLEDGE BASES FOR EXPERT SYSTEMS FOR
REMOTE SENSING IMAGE ANALYSIS WITH GIS DATA**

by

Xueqiao Huang and John R. Jensen
Department of Geography
University of South Carolina
Columbia, SC 29208

and

Halkard E. Mackey, Jr.
Westinghouse Savannah River Company
Savannah River Technology Center
Environmental Sciences Section
Aiken, SC 29802

For publication in the Proceedings of the Annual Meeting of the American Society for
Photogrammetry and Remote Sensing, April 1996, Baltimore, Maryland

This information contained in this manuscript was developed during the course of work
under Contract No. DE-AC09-89SR18035 with the US Department of Energy.

A MACHINE LEARNING APPROACH TO AUTOMATED CONSTRUCTION OF KNOWLEDGE BASES FOR EXPERT SYSTEMS FOR REMOTE SENSING IMAGE ANALYSIS WITH GIS DATA

Xueqiao Huang
John R. Jensen

Department of Geography, University of South Carolina
Columbia, SC 29208

Halkard E. Mackey, Jr.

Savannah River Technology Center, Westinghouse Savannah River Company
Aiken, SC 29808

ABSTRACT

Knowledge-based remote sensing image analysis with GIS data is acknowledged as a promising technique. However, the difficulty in knowledge acquisition, a well-known bottleneck in building knowledge-based systems, impedes the adoption of this technique. Automating knowledge acquisition is therefore in demand. This paper presents a machine learning approach to automated construction of knowledge bases for image analysis expert systems integrating remotely sensed and GIS data. The methodology applied in the study is based on inductive learning techniques in machine learning, a subarea of artificial intelligence. It involves training with examples from remote sensing and GIS data, learning using the inductive principles, decision tree generating, rule generating from the decision tree, and knowledge base building for an image analysis expert system. This method was used to construct a knowledge base for wetland classification of Par Pond on the Savannah River Site, SC, using SPOT image data and GIS data. The preliminary results show that this method can provide an effective approach to integration of remotely sensed and GIS data in geographic information processing.

I. INTRODUCTION

Remote sensing can in certain instances provide up-to-date land cover information. Unfortunately, the classification accuracy of the resultant maps derived from traditional digital image processing methods may be insufficient for some GIS applications (Trotter, 1991).

It is generally acknowledged that the results of statistical image processing techniques are often rather crude when compared with those of a skilled photo-interpreter (Philipson, 1986). Statistical approaches usually rely solely on the spectral information, while image interpretation by human photo-interpreters involves the consideration of spectral, spatial and contextual information in the image (i.e., tone, color, size, texture, shape, pattern, height, shadow, site, association) (Nagao *et al.*, 1980; Jensen, 1996). In addition, the photo-interpreter uses heuristic rules of thumb and ancillary information (Argialas, *et al.*, 1990).

The above observation suggests that it is necessary to incorporate supplemental information and knowledge when performing digital image processing of remote sensing data. This justifies two important research areas: (1) incorporating ancillary GIS in the image analysis process, and (2) employing knowledge-based system or expert system techniques. Enslin *et al.* (1987) point out that geographers should examine how GIS can be used to improve image classification through application of the logic and techniques of artificial intelligence. On the other hand, when incorporating spatial information from GIS data with spectral information from remotely sensed

data, traditional statistical image processing methods are not capable of handling these two kinds of information which usually have different statistical natures. Thus, adopting expert system or knowledge-based approaches to image analysis with remote sensing and GIS data is a natural development in geographic information processing.

Argialas and Harlow (1990) observe that image interpretation techniques have shifted from spectral classification to contextual, spatial syntactic analysis, and to knowledge-based interpretation. Therefore, a gradual fusion has taken place between knowledge-based systems, pattern recognition, image analysis, and GIS. In recent years a number of studies have employed knowledge-based systems or expert systems and incorporated GIS data in the analysis of remotely sensed data (McKeown, 1987; Argialas *et al.*, 1990; Newkirk *et al.*, 1990; Janssen *et al.*, 1992; Knotoes *et al.*, 1993; Johnsson, 1994). However, such methods are complex and not trivial to develop (Knotoes *et al.*, 1993). Most image analysis expert systems are not operational. No commercial software package for knowledge-based image analysis is available. It has been acknowledged in the artificial intelligence community that the problem is due to the "knowledge acquisition bottleneck". Acquiring human knowledge and transferring it to a computer-usable form to build a knowledge base for an expert system is complex, time-consuming and expensive (Feigenbaum, 1981). To solve this problem, great effort has been exerted in the artificial intelligence community to automate knowledge acquisition for obtaining low-cost and high-quality knowledge bases. The studies on automated knowledge acquisition belong to one subfield of artificial intelligence (AI) known as machine learning (Carbonell *et al.*, 1983).

The knowledge acquisition bottleneck does exist in knowledge-based image analysis. Researchers in this area have reported difficulty in building knowledge bases for expert image analysis systems (Kontoes *et al.*, 1993). In fact, geographic knowledge is usually fuzzy and uncertain. Thus, it is even more difficult to elicit. In addition, many current knowledge-based systems for image analysis are ad hoc systems. It is difficult to transfer their knowledge base to other projects with different objectives or geographic conditions. This makes the cost of building an expert system even higher. However, while machine learning techniques have been employed in automatic construction of knowledge bases for expert systems in many areas, its application in remote sensing and GIS is rare. This research fills this gap by developing a methodology using techniques of machine learning to automatically construct a knowledge base for an integrated image analysis expert system that incorporates remotely sensed and GIS data. Because production rules are the most common method of representing the knowledge in a knowledge base, automatically generating rules is the major goal of this research.

I. KNOWLEDGE-BASED SYSTEMS AND MACHINE LEARNING

2.1 Expert Systems and the Knowledge Acquisition Bottleneck

An expert system is a computer program that represents and reasons with knowledge of some specialist subject with a view to solving problems or giving advice (Jackson, 1990). It is based on an extensive body of knowledge about a specific problem domain. Such knowledge is stored in a knowledge base separately from the inference engine, both of which compose the core of an expert system. A knowledge base contains the knowledge usually in the form of facts and rules. Such a system is also called a "rule-based system". Most current expert systems are of this type. It is the extent and quality of its knowledge base that determines the success of an expert system (Forsyth, 1984). Because of the importance of the knowledge base in an expert system, the terms "knowledge-based systems" and "expert systems" are often used interchangeably in the literature.

Although not all of the technical problems in developing an expert system have been overcome, the purely technical issues involved in the system's role, such as selecting an appropriate inference engine, are no longer major obstacles. Instead, many human factors have been identified as major

obstacles (Suh *et al.*, 1993). The time-consuming and expensive process of knowledge acquisition is widely considered the major bottleneck in building expert systems (Feigenbaum, 1981; Forsyth, 1984; Suh *et al.*, 1993). There are two major reasons for this bottleneck: (1) the process requires the engagement of the domain expert and the knowledge engineer over a long period of time; and (2) although experts are capable of using their knowledge in their decision making, they are usually not capable of formulating their knowledge explicitly in a form sufficiently systematic, correct and complete to form a computer application (Bratko *et al.*, 1989). Therefore, there is a need to automate the knowledge acquisition process. Machine learning may be used to accomplish this task.

2.2 Machine Learning: a Possible Solution to the Problem

Machine learning is a subfield of artificial intelligence. It is the science of computer modeling of human learning processes. One of its major objectives is to automate the process of knowledge acquisition for other AI applications, including expert systems. Machine learning techniques enable a computer to acquire knowledge from existing data, theories or knowledge with certain inference strategies, such as induction or deduction. Using computer-acquired knowledge for automated construction of knowledge bases for expert systems is one of the major applications of machine learning techniques, and also one of the possible solutions to the knowledge acquisition bottleneck in building expert systems (Bratko, 1990, Maniezzo *et al.*, 1993).

III. RESEARCH METHODOLOGY

Over the years, research in machine learning has been pursued with varying degrees of intensity using different approaches and placing emphases on different aspects and goals. Based on the underlying learning strategy, machine learning can be classified into rote learning, learning from instruction (or learning by being told), learning by analogy, learning from examples (supervised learning), and learning from observation and discovery (unsupervised learning) (Carbonell *et al.*, 1983).

In this study, learning from examples (supervised learning), which is a form of inductive learning was used. This technique is suitable to a problem domain with (1) descriptions of data attribute values, (2) a predefined classification scheme, (3) discrete classes, and (4) sufficient data (Quinlan, 1993). Remote sensing image analysis with GIS data has these features. Supervised learning tries to induce a classification rule from a set of pre-classified training examples. Usually, the input to the program in such a learning process is a set of training data that can be viewed as a list. Each record in this list contains a class attribute indicating the class to which this example belongs and the values of other attributes that are used to classify the raw data. The output is expected to be a method of classifying subsequent instances (Ginsberg, 1993). The resulting classification method can be in the form of production rules (Quinlan 1993), which can be used directly for the automated construction of knowledge bases for expert systems. In this way, supervised inductive learning exploits the known empirical observation that experts find it easier to produce good examples than to provide explicit and complete general theories (Bratko, 1990). This is exactly the case in human image interpretation. Therefore, supervised learning is an appropriate technique for our problem domain.

There are a number of algorithms based on inductive learning, such as ID3 (Quinlan 1979) and AQ15 (Michalski *et al.*, 1986). Induction of decision trees is one of the most widely used approaches to inductive (supervised) learning. ID3 is the most famous tree induction algorithm and tree induction is therefore often referred to as ID3 (Bratko, 1990). Algorithm C4.5 improves and augments ID3 (Quinlan, 1993).

The method of applying inductive learning technique in this research is based on Quinlan's learning algorithms C4.5. In addition to the abovementioned common features of the supervised inductive learning, C4.5 is very flexible. There are no assumptions or requirements for data statistical distribution. Data can be either nominal or numerical. This is very beneficial to the integration of GIS data with remote sensor data as they usually have different statistical distributions and many ancillary GIS data are nominally scaled. The major idea of C4.5 is to induct a decision tree for classification from training examples, then generate decision rules from this decision tree. The learning process with the C4.5 algorithm includes the following stages:

1. Data set definition. This includes selecting appropriate remotely sensed data and GIS data to be used in the learning process and thereafter in the knowledge-based classification system. The value of the data attributes must be well described. The values may be non-numerical or numerical.

2. Class Definition. This is similar to the traditional supervised classification method. A classification scheme must be adopted before the learning takes place.

3. Training. This is the process where the domain image analysis experts become involved. They select a subset of the data that should be representative of the remaining data in the data set and classify this data subset according to the classification scheme using their domain expertise. They are not required to explain their expertise explicitly. The computer will perform this task in the following stages. This is the major difference between machine learning and the conventional method of building a knowledge base.

4. Generating decision trees. Based on the training example, the algorithm tries to employ an induction strategy to generalize a series of decision rules for classification, which can be viewed as a decision tree for the classification. This is the core of C4.5. One of the unique features of C4.5 is its use of information theory to guide the selection of the attribute used at each node of the decision tree. It maximizes the information gain at each node by selecting the appropriate attribute used to decide the division of example data to produce the most effective decision tree.

5. Pruning decision trees. The output from the last process is often a very complex tree and can be simplified. This is done by discarding one or more subtrees and replacing them with leaves. Such replacement is based on predicting the error rate of a tree and its subtrees. If replacement of this subtree with a leaf, or with its most frequently used branch, would lead to a lower predicted error rate then the tree is pruned accordingly.

6. Generating rules. This is the most important step in the process as it automatically constructs a knowledge base for the image analysis expert system. First, initial rules from each path of the decision tree are generated. Each rule is then simplified by removing conditions that do not seem helpful for discriminating the nominated class from other classes. For each class in turn, rules are further simplified according to their contribution to the accuracy of the set of rules as a whole.

7. Building the knowledge base. Using rules generated during the previous stage the knowledge base is built. It can be used in an expert system to conduct final classification of the entire data set with an inference engine.

IV. A CASE STUDY: PAR POND WETLAND LAND COVER CLASSIFICATION

4.1 Study Area and Data

Par Pond is a 1000 ha. reservoir on the Savannah River Site, SC. It was constructed in 1958. Natural invasion of wetland has occurred over its 37-year history with much of the shoreline having developed extensive beds of persistent and non-persistent aquatic macrophytes. Par Pond

has been the object of numerous studies of wetland ecology using remote sensing and GIS technologies (Jensen *et al.*, 1992, 1993). The data and expertise from the previous and ongoing projects on Par Pond are beneficial to this study.

A SPOT multispectral image of Par Pond obtained on February 2, 1994, was used. The previous study has shown that spatial information and the domain expertise is very essential to the identification of some wetland vegetation. For instance, it has been confirmed that four biophysical variables that can be obtained from GIS, including the water depth or elevation, slope, fetch (exposure pattern) and soils affect aquatic macrophyte growth (Jensen *et al.*, 1992). These data were obtained from a GIS database and used as the ancillary data along with the spectral data to build the knowledge base for image analysis.

DEM. The DEM of Par Pond was developed by digitizing large scale (1:1,200) photogrammetrically derived topographic maps for the lower half prior to its construction and 1:24,000 USGS 7.5' topographic quadrangles for the upper half. A triangulated irregular network (TIN) model was built from the digitized contour lines. It was then converted to a 5x5 m grid DEM.

Slope. The percentage slope data were generated from the DEM with the same grid.

Fetch. Fetch is the unobstructed distance that wind can blow over water in a specified direction. It is an important factor affecting the growth of aquatic macrophytes. Fetch used for Par Pond was computed according to the distance from a pixel to the shore at a specific angle, which can range from 0 to 360 degree, and the distance from the pixel to the shore in the direction of dominant wind (Jensen *et al.*, 1992). The fetch data are also in the format of a 5x5 m grid.

Soil. The soil data for Par Pond was classified into five categories according to its suitability for aquatic macrophytes, including worst, poor, moderate, good and best. The ARC/INFO coverage of the soils was also converted into the format of a 5x5 m grid.

A classification scheme was developed for the abovementioned ongoing project. This scheme was adopted in the training and for the final classification. The classes identified for the specific study date according to this scheme are: water (w), bare soil (b), spikerush (s), bullrush (br), dead vegetation (d), old field (o) and pine/hardwoods (p). These six classes were used for the training data and final classification.

Training data were selected by ecology experts using large scale color infrared aerial photography (NAPP photography) and other ground truth data, including ground survey, ground photography and video records. A total of 131 points with relatively even distribution were selected. The NAPP photography obtained on January 18, 1994, was scanned and geometrically registered to the SPOT image which was rectified to a UTM projection. The training points were digitized on the screen with a background of rectified NAPP photography. The UTM coordinates of the training points were used to extract the values of the seven variables (attributes) from the GIS database.

4.2 Preliminary results

Two experiments were conducted to test the learning procedure and evaluate its usefulness in building knowledge bases for image analysis with GIS data. In the first test, the learning procedure was conducted only with SPOT spectral data (Band 1: B1, Band 2: B2, Band 3: B3). The C4.5 decision tree program used the 131 training samples and generated a decision tree with 33 nodes (Figure 1). A knowledge base with 11 rules was constructed (Figure 2). The rules were used to classify the training data and an error of 15.3% resulted (Table 1). Major errors included confusion between old field (o) and spikerush (s) and dead vegetation. Of 20 samples of old field, 9 were

incorrectly classified as other classes. This led to an error of 45%. This is not a surprise because such confusion reflects the overlap of spectral classes representing these landcover classes in the feature space of the SPOT multispectral data and has been noticed in the previous interpretation and conventional classification of SPOT spectral data in Par Pond.

Table 1. The error matrix of the classification of the training data using the knowledge base in the Figure 2

Ground/Classified	w	b	s	br	d	p	o	Total
water	23		3					26
bare soil		23	2					25
spikerush			18	1			1	2
bullrush				8	1		1	10
dead vegetation			1		19			20
pine/hardwood			1			9		10
old field			6	2	1		11	20
Total	23	23	31	11	21	9	13	131

In the second test, the four GIS variables were used in addition to the three spectral variables. This produced a decision tree with the same number of nodes and a knowledge base with the same number of rules as in the first test (Figures 3 and 4). The classification error of the training data using these rules, however, was only 8.4%. The misclassified number of old field was 3 with an error rate of 15%, compared to 9 and 45%, respectively, in the first test (Table 2). The GIS variables played an important role in the rules used to classify the old field (rules 8, 9, 10 and 11), spikerush and dead vegetation (rule 6). This indicates that the GIS variables enhanced the power of classifying these three classes that were confused spectrally. On the other hand, spectral data are important in the rules to classify water and bare soil, which are two spectrally extreme classes and easy to be classified using only spectral data. This suggests that the rules produced by the machine learning method are compatible with human interpretation methods.

Table 2. The error matrix of the classification of the training data using the knowledge base in the Figure 4

Ground/Classified	w	b	s	br	d	p	o	Total
water	24		2					26
bare soil		25						25
spikerush			19				1	20
bullrush			2	7			1	10
dead vegetation			1		19			20
pine/hardwood			1			9		10
old field			3				17	20
Total	24	25	28	7	19	9	19	131

V. CONCLUSIONS

Ancillary data available in a GIS are very valuable in land cover classification of remotely sensed data. Traditional statistical methods are inherently incapable of handling the complexity of a broad spectrum of data in such cases. Knowledge-based methods provide a powerful and flexible approach to dealing with the nonspectral data. However, this requires the expense of building of complex knowledge bases. Machine learning, a subfield of artificial intelligence provides a potential solution to this dilemma. In this study, a supervised inductive machine learning technique generated decision trees and production rules from training data to construct knowledge bases

automatically for a land cover classification in Par Pond on the Savannah River Site, SC using SPOT multispectral image data and GIS data. This method exploits the empirical observation that experts find it easier to produce good examples than to provide explicit and complete general theories by allowing the experts to do the easier part of the job and leaving the difficult part of the job (providing the rules) to the computer. The preliminary results of the study are promising. It generated automatically a usable knowledge base for the land cover classification. It also incorporated the GIS data into the image analysis effectively and efficiently. Further studies should evaluate, (1) using a separate test data set rather than the training data, (2) new techniques, such as windowing (Quinlan, 1993) to generate more accurate trees and knowledge bases, and (3) incorporating GIS data with nonnumerical attribute values.

ACKNOWLEDGMENTS

We would like to thank Dr. David Cowen, Dr. John Rose and Dr. Daniel Wagner for their valuable suggestions.

REFERENCES

Argialas, D., and Harlow, C., 1990. Computational image interpretation models: an overview and perspective, *Photogrammetric Engineering and Remote Sensing* 56(6):871-886.

Bratko, I., Kononenko, I., Lavrac, N., Mozetic, I. and Roskar, E., 1989. Automatic synthesis of knowledge: Ljubljana research, in Kodratoff, Y. and Hutchinson A. (eds.) *Machine and Human Learning*, Columbia, MD: GP Publishing, Inc. pp. 25-33.

Bratko, I., 1990. *PROLOG: Programming for Artificial Intelligence*, second edition, Wokingham, England: Addison-Wesley Publishing Company.

Carbonell, J. G., Michalski, R. S. and Mitchell, T. M., 1983. An overview of machine learning, in Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.) *Machine Learning*, Vol. I., San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Enslin, W. R., Ton, J. and Jain, A., 1987. Land cover change detection using a GIS-guided, feature-based classification of Landsat Thematic Mapper data, *Proc. ASPRS*, 6:108-20.

Feigenbaum, E. A., 1981. Expert systems in the 1980s. in Bond, A. (Ed.), *State of the Art Report on Machine Intelligence*, Maidenhead: Pergamon-Infotech.

Forsyth, R., 1984. The architecture of expert systems, in *Experts Systems: Principles and Case Studies*, London: Chapman and Hall

Ginsberg, M., 1993. *Essentials of Artificial Intelligence*, San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Jackson, P., 1990. *Introduction to expert systems*, 2nd edition, Wokingham, England: Addison-Wesley Publishing Company.

Janssen, L. L. F. and Middelkoop, H., 1992. Knowledge-based crop classification of a Landsat thematic mapper image, *Int. J. of Remote Sensing*, 13(15): 2827-2837.

Jensen, J. R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Englewood Cliffs: Prentice-Hall.

- Jensen, J. R., Narumalani, S., Weatherbee, O., Morris, K. S. and Mackey, H., 1992. Predictive modeling of cattail and waterlily distribution in a South Carolina reservoir using GIS, *Photogrammetric Engineering and Remote Sensing*, 58(11):1561-1568.
- Jensen, J. R., Narumalani, S., Weatherbee, O., and Mackey, H., 1993. Measurement of Seasonal and Yearly Cattail and Waterlily changes using multirate SPOT panchromatic data, *Photogrammetric Engineering and Remote Sensing*, 59(4):519-525.
- Johnsson, K., 1994. Segment-based land-use classification from SPOT satellite data, *Photogrammetric Engineering and Remote Sensing*, 60(1): 47-53.
- Kontoes, C., Wilkinson, G. G. and et al, 1993. An experimental system from the integration of GIS data in knowledge-based image analysis for remote sensing of agriculture, *Int. J. Geographical Information Systems*, 7(3): 247-262.
- Maniezzo, V. and Morpurgo, R., 1993. D-KAT: a deep knowledge acquisition tool, *Expert Systems*, 10(3):157-165.
- McKeown, D. M., 1987. The role of artificial intelligence in the integration of remotely sensed data with geographic information systems, *IEEE Transactions on Geoscience and Remote Sensing*, 25(3): 330-348.
- Michalski, R. S., Mozetic, Hong, I. and Lavrac N., 1986. The multi-purpose incremental learning system AQ15 and its testing application to three medical domains, *Pro. of the 5th Annual National Conference on Artificial Intelligence*, Philadelphia, 1041-1045.
- Nagao, M., and Matsuyama, T., 1980. *A Structural Analysis of Complex Aerial Photographs*, New York: Plenum Press
- Newkirk, R. T. and Wang, F., 1990. A common knowledge database for remote sensing and geographic information in a change-detection expert system, *Environment and Planning (B)*, 17(4): 395-404
- Philipson, W., 1986. Problem solving with remote sensing: an update, *Photogrammetric Engineering and Remote Sensing*, 52(1): 109-110.
- Quinlan, J. R., 1979. Discovering rules by induction from large collections of samples, in Michie D., (ed.) *Expert Systems in the Micro Electronic Age*, Edinburgh, UK: Edinburgh University Press.
- Quinlan, J. R., 1993. *C4.5: Programs for Machine Learning*. San Mateo: Morgan Kaufmann Publishers.
- Suh, C. and Suh, E., 1993. Using human factor guidelines for developing expert systems, *Expert Systems*, 10(3): 151-156.
- Trotter, C. M., 1991. Remotely-sensed data as an information source for geographical information systems in natural resource management: a review, *Int. J. Geographical Information Systems*, 5(2): 225-239.

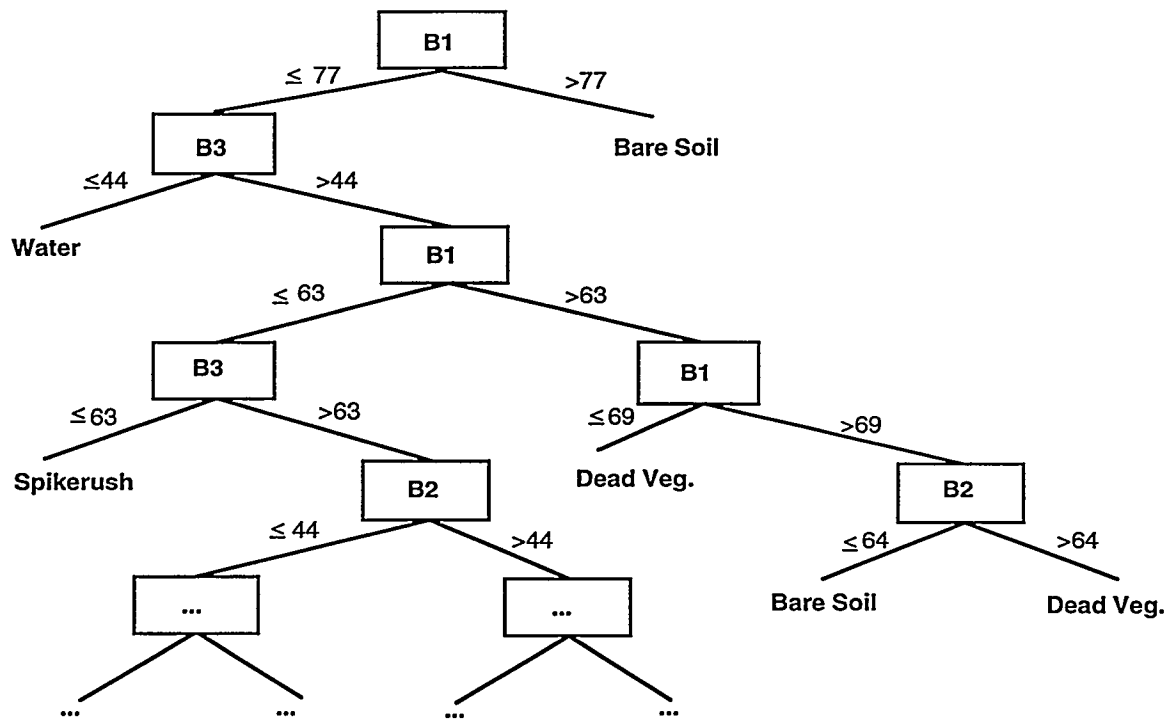


Figure 1. A portion of the decision tree generated only from the SPOT spectral data

- Rule 1: $B3 < 44$ \rightarrow class w*
Rule 2: $B1 > 77$ \rightarrow class b
Rule 3: $B2 < 44$ AND $B3 > 71$ AND $B3 < 79$ \rightarrow class p
Rule 4: $B1 < 52$ AND $B3 > 63$ \rightarrow class p
Rule 5: $B1 < 58$ AND $B2 > 46$ AND $B3 > 76$ \rightarrow class p
Rule 6: $B1 > 52$ AND $B1 < 63$ AND $B3 < 63$ \rightarrow class s
Rule 7: $B1 > 62$ AND $B1 < 63$ AND $B3 < 76$ \rightarrow class s
Rule 8: $B1 > 63$ AND $B1 \leq 69$ \rightarrow class d
Rule 9: $B1 > 58$ AND $B1 \leq 77$ AND $B3 > 78$ \rightarrow class d
Rule 10: $B1 \leq 62$ AND $B2 > 44$ AND $B3 > 63$ AND $B3 \leq 76$ \rightarrow class o
Rule 11: $B2 > 44$ AND $B2 \leq 49$ AND $B3 > 76$ \rightarrow class br

Figure 2. The Knowledge base generated only from the SPOT spectral data

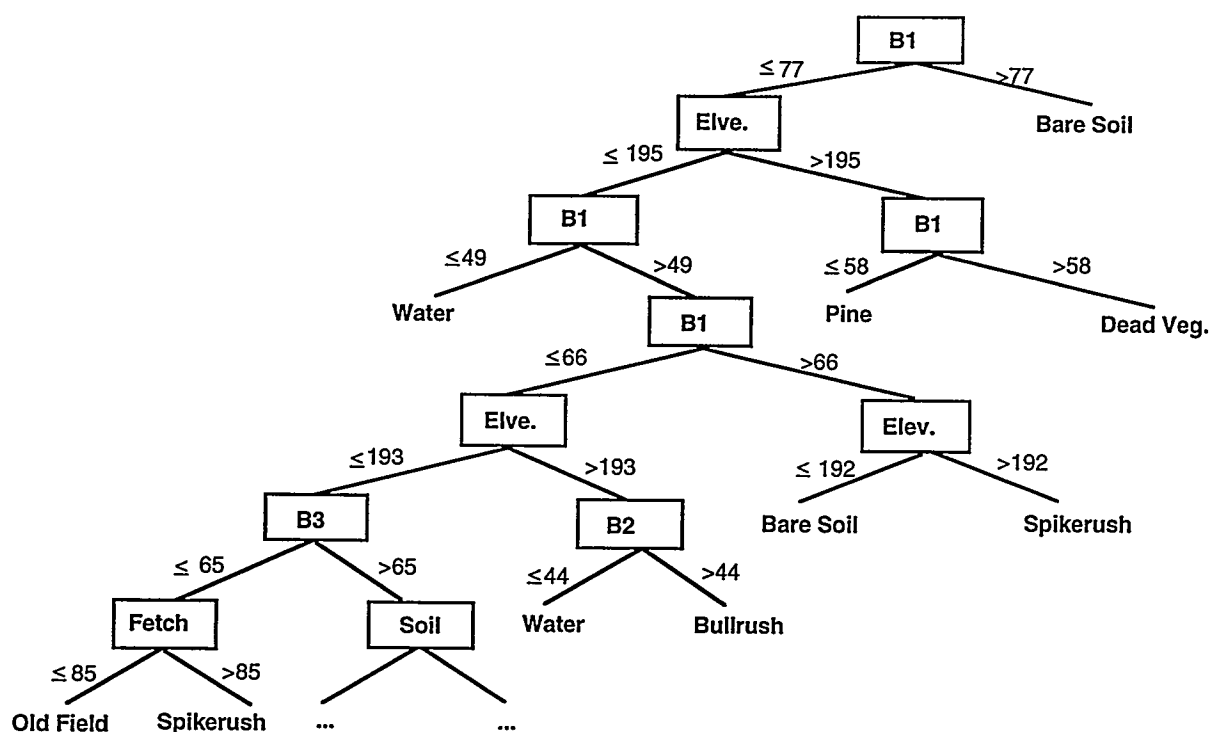


Figure 3. A portion of the decision tree generated from the SPOT spectral data and GIS data

- Rule 1: $B1 \leq 49$ AND elevation ≤ 195 \rightarrow class w
- Rule 2: $B1 > 77$ \rightarrow class b
- Rule 3: $B1 > 66$ AND elevation ≤ 192 \rightarrow class b
- Rule 4: $B1 > 49$ AND $B3 \leq 65$ AND elevation ≤ 195 AND fetch > 85 \rightarrow class s
- Rule 5: $B1 > 55$ AND $B2 \leq 49$ AND $B3 > 65$ AND slope > 2 AND elevation ≤ 193 AND soil > 0 and soil ≤ 4 \rightarrow class br
- Rule 6: $B1 > 58$ AND $B1 \leq 77$ AND elevation > 195 \rightarrow class p
- Rule 7: $B1 \leq 58$ AND elevation > 195 \rightarrow class p
- Rule 8: $B1 \leq 66$ AND $B3 > 65$ AND slope ≤ 2 AND elevation ≤ 195 AND soil > 0 \rightarrow class o
- Rule 9: $B1 \leq 66$ AND $B3 > 65$ AND elevation ≤ 195 AND soil > 4 \rightarrow class o
- Rule 10: $B1 > 49$ AND $B3 \leq 65$ AND fetch ≤ 85 \rightarrow class o
- Rule 11: $B1 \leq 77$ AND $B3 > 70$ AND elevation ≤ 193 AND soil ≤ 0 \rightarrow class o

Figure 4. The Knowledge base generated from the SPOT spectral data and GIS data

