2003

Adaptive feature extraction: Exploring the search space with change detection using inductive learners and image processing

Joseph D. Zeiler
The University of Montana

Permission is granted by the author to reproduce this material in its entirety, provided that this material is used for scholarly purposes and is properly cited in published works and reports.

Please check "Yes" or "No" and provide signature

Yes, I grant permission ✓

No, I do not grant permission ____

Author's Signature: Joseph D. Zeiler

Date: October 10, 2003

Any copying for commercial purposes or financial gain may be undertaken only with the author's explicit consent.

Adaptive Feature Extraction:

Exploring the Search Space with Change Detection using Inductive Learners and Image Processing

By

Joseph D. Zeiler

BS, Computer Science, The University of Montana

Presented in partial fulfillment of the requirements for the Degree of Master of Science
The University of Montana
2003

Approved by:

David W. Opitz

Chairman, Board of Examiners

Dean of the Graduate School

10-10-03

Date

Zeiler, Joseph D., M.S., October 2003 Computer Science

Adaptive Feature Extraction, Exploring the Search Space with Change Detection using Inductive Learners and Image Processing

Director: David W. Opitz DWO

Remotely sensed digital imagery provides snapshots of the earth at given time frames giving scientists information necessary to analyze changes to the earth. Due to the steadily increasing number of images available and the dynamic nature of the earth scientist need tools to help them identify the features that are changing as well as the nature of the changes. Current techniques for feature extraction suffer from being unable to classify complex features and similarly do not generalize well to other datasets with similar features. Machine learning techniques using inductive learners have shown great promise in identifying the desired features in information with minimized user interaction and have greatly reduced the time needed to correctly classify objects in imagery over traditional techniques. We present a technique for adapting existing feature extraction efforts using inductive learners and image processing techniques to improve the quality of identifying features in remotely sensed digital imagery that change over time. The technique compares snapshots of an area taken at different times creating a difference image which is incorporated into the learning process as additional information. Additionally, multiple features are incorporated into the learning task along with the addition of multiple sets of classifiers forming ensembles. The basic idea is that individual classifiers are used to learn each feature. The classifiers are successively combined to produce the predicted output. Multiple "bootstrapped" ensembles of these classifiers then predict the final output for the image classification improving the accuracy and generalization of the classification task. The adaptive feature extraction technique successfully adapts the imagery to the learner, refines the classification, and classifies temporal features in a variety of situations.

Table of Contents

List of Figures

1. INTRODUCTION

The universe is a complex dynamical system. Scientists attempt to understand the behavior of the universe by gathering information at varied intervals of time and analyze the information to determine the state of a given system, its history, and even to predict the future events of a system. Satellite technology provides a way by which to collect information. The process of extracting the information from space over time via satellite and converting the information from the various sensors into digital form is often referred to in literature as remotely sensed digital imagery. This information is analyzed to look at the state of the system at a given time as well as to look at the changes in the state of the system over time. Analysis frequently looks for specific features that are of interest in information. Feature Extraction is an area of research that studies the process of identifying features of interest in information. *This paper presents a technique for adapting existing feature extraction efforts using inductive learners and image processing techniques to improve the quality of identifying features in remotely sensed digital imagery that change over time.*

Current machine learning techniques using inductive learners have shown great promise in identifying the desired features in information with minimized user interaction and have greatly reduced the time needed to correctly classify objects in imagery over traditional techniques (Bain, 2000; Burl et al., 1998; Maloof et al., 1998; Mangrich, 2001). Currently, there is need to improve the current feature extraction techniques that use inductive learners since there are many complex features that are difficult to identify with a single model. Current techniques also suffer from not being able to generalize well to other data sets with similar features.

There is potential promise in the use of incorporating multiple models by using multiple images of the same area of space taken over time. Changes are of primary interest to scientists since they provide information necessary to understand complex dynamical systems. Current research has worked with change detection and has found the problem to be very difficult due to the different sampling parameters of the various sensors in the satellites as well as environmental conditions altering the sampling values. Even though a direct comparison is difficult, change detection has shown potential promise in incorporating the additional data as another model for the inductive learners to use by finding change amongst images and incorporating it into the learning task.

Particularly, this process will help to find features that change dramatically from image-to-image and focus the learner on these features. For example, scientists are often interested in changes that occur after extreme natural phenomena such as earthquakes or fires. Scientists need to be able to ascertain damage after such an event and could use before and after satellite images to detect change and then learn the various levels of change amongst the damage. This could greatly benefit planners working to assess clean-up efforts and prioritize damaged areas.

Increasing the generalization ability of an inductive learner is also a difficult task. Currently various additive models have been used to increase the complexity of the learner until a certain degree of generalization is obtained. Other techniques start with a complex learner and subtract elements of the model until an acceptable level of generalization is reached. The other common approach is to modify the search space of the dataset so that it is more compatible with the learner. Ultimately all of the techniques work to increase the level of generalization while maintaining as much accuracy as possible for the given task at hand.

Together, change detection and image processing techniques incorporated into feature extraction contain the potential to increase the ability of the learner to find more complex features. Incorporating change detection is an important addition to the feature extraction process and will provide analysts the techniques needed to detect change in features and allow them to adapt past feature extraction efforts to new images. This is an important step that will begin to provide the foundation to build predictors for future events. Adaptation of the learning process to other images through image compression techniques will provide a means to increase the generalization ability of the learners to classify other images.

2. BACKGROUND

Earth is a complex dynamical system within the universe that is of primary interest to scientists. Satellites extract information about the earth giving a picture of the state of a given area of space based on the sensors of the device creating an image of that space at a given time at a defined resolution. Scientists interpret images to recognize important features that will help them analyze the region of space the images cover. Satellites sample information over varied intervals of time and hence create many images. Image interpretation can quickly become difficult for scientists since thousands of images can be taken of a given area of space in minutes. This makes traditional feature extraction efforts overly time consuming for scientists. Thus computers are often used as a tool to facilitate the interpretation of images.

2.1. Computer Images

A computer image is a numerical representation of an image (or picture). The computer image is represented in two-dimensions consisting of a set of l bands each of which consists of a $m \times n$ dimension array where m represents the rows and n represents the columns of the image, respectively. Each element of the band, or pixel, contains z bits used to represent the intensity of a specified area of a band in the image. (Watt, 1999)

Images created using satellites contains sensors that sample the given space at a particular time for each sensor producing a finite representation of the space. The information is limited by the sensors' capabilities to extract information from the given area of space and further limited by the precisions of the computer used to store the information (bit depth of each pixel and the number of pixels used to represent that area of space). The process of extracting information with the sensors is also further limited by the state of the system at the time of sampling due to interference from environmental conditions that prevent optimal sampling.

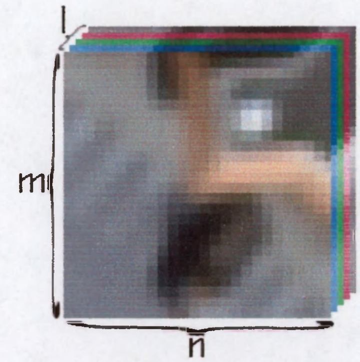


Figure 1: Sample Digital Image

2.2. Image Processing

The goal of image processing is to enhance an image in some way so that it is more easily interpretable. This often means producing an image that is more easily viewable or to reduce noise in an image. Image processing techniques are operations that transform an image either spectrally or spatially in the spatial domain or transform an image based on the spatial frequency in the Fourier domain. Wavelet transforms simultaneously transform the spatial and frequency domains. Image transformations fall into two broad categories global image transformations and local image transformations.

Global image transformations look at the global characteristics of the image. These transformations include intensity, image combining, geometric, and color transformations. Image intensity transformations deal with the histogram of the color intensities of an image. Common operations on the histogram alter the visible contrast of an intensity range about an object of interest. Image combining transformations select bands from images and compare them. Often logical operations such as AND, OR, and XOR are used to perform these operations. Geometric image transformations such as scaling, rotation, and shearing are common operations performed on images. Color image transformations transform the color bands of an image into alternate representations such as hue, intensity, and saturation components versus tradition red, green and blue components.

Local image transformations look at local characteristics of an image and alter them in a local context. Most local image transformations evolve around spatial domain convolution filters. The two most common filters are low-pass filters (smoothing) and high-pass filters (edge enhancement). The operations work in the spatial context by computing values using neighboring pixels to compute the new pixel values based on spatial context.

Convolution filters work on each pixel or a group of pixels of an image. A convolution kernel **H** represented as a matrix containing the coefficients for each pixel in the neighborhood of a defined pixel is used to transform the given point, P, of an image at that pixel. The basic process starts by choosing the point, P, to apply the convolution filter. Multiplication with each convolution kernel coefficient and the respective pixel in the image then occurs. Each product is then summarized, normalized, and placed in the output image at P. This can be represented by $P = \sum_{i,j} I[i,j] * H[i,j]$ as illustrated in *Figure 2* below (Seul, 2000).

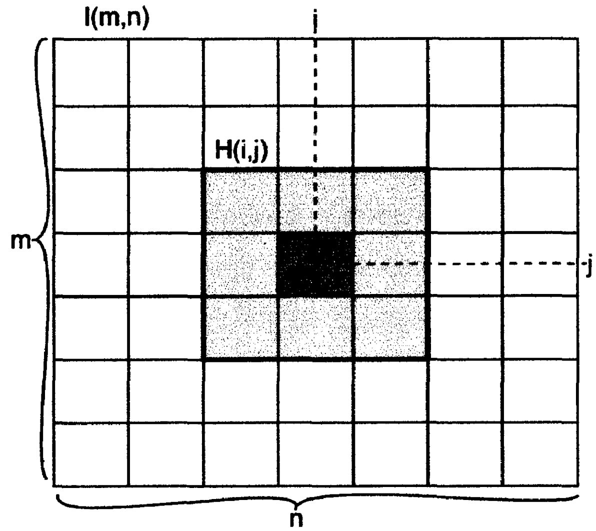


Figure 2: Sample Convolution Filter

2.3. Feature Extraction

Feature extraction is the process of identifying features of interest in a dataset. In remotely sensed digital imagery, satellites take snapshots of a given area of space at a given time producing a representation of that area in digital form at a specific resolution. Feature extraction for digital images looks at the data and identifies areas of an image that represent the specified feature of interest. The term image classification is used to refer to the process of classifying image features into classes or themes. In image classification, algorithms known as image classifiers analyze images and produce classes of image features, which are often represented pictorially as thematic maps.

Extracting information from remotely sensed digital imagery can be a difficult task. Feature extraction is not only limited by image resolution and physical sampling but also the features themselves can be limiting. Features often relate spectrally and spatially to other objects in an image. The term used to refer to the degrees of similarity or separation of objects is disjunctive. For instance, given an aerial view of a city, one can easily identify features like buildings, trees, roads, etc. with similar color and shapes.

It is often difficult to distinguish between different types of similar features since their spectral signatures are similar. If the task at hand were to find all fir trees in an image, this task would be quite difficult if many different types of evergreens also existed in the image. Other information would be needed to help identify the fir trees in the image. This is also true with shapes of similar features.

Most feature extraction efforts gather as many bands as possible incorporating different sensor ranges with as much resolution as possible. This information is then analyzed and processed to select the information that will help to classify the desired features. The image classifiers attempt to exploit spectral differences in the various bands as well as spatial similarities. Even with all of this information, features can be complex and their representation is often nonlinear in the feature space making detection difficult. Thus there has always been a need for improved techniques.

2.4. Current Techniques

Originally, feature extraction for images evolved from data processing techniques already used to analyze data. These techniques evolved with the advent of the computer helping scientists create techniques to handle larger amounts of data producing more complex techniques. Digital images began to emerge and refined the techniques for computer images creating the field of image processing. The evolution of feature extraction started in the late 1950s and continues to the present.

Feature extraction techniques are heavily rooted in artificial intelligence. A branch of artificial intelligence known as machine vision began by exploring techniques to reproduce human vision for robotics. These algorithms started analyzing just spectral information and quickly found need for spatial context and shape.

Many existing image classification techniques currently exist. The two primary ways of classification are supervised and unsupervised. Supervised techniques require user interaction where a user iteratively refines the features in the image until an acceptable level of classification for the image is reached. Unsupervised techniques attempt to automate the entire process. This thesis only looks at supervised techniques that adapt images to the existing feature space to augment feature extraction.

Classic feature extraction techniques include template matching, adaptable pattern recognizers, statistical classifiers, and trainable pattern recognizers. The earliest form of pattern matching became know as template matching. Template matching consists of storing a prototype or template of features for each class and then checking each image for the features based on their similarities to the prototypes. This early method suffered from being able to only correctly classify well-defined features with little change and relied on just image intensity information.

Adaptable pattern recognizers attempt to reduce some of the limitations imposed by the early template matching schemes by making the prototypes deformable. The techniques allowed variations in the prototypes to exist, which allowed for more robust classification. For instance, if an aerial image was taken of a city, and another image was taken of this same area at a different angle, the classic template matching technique would not produce a good match since the templates would differ greatly. These techniques often take an existing template and rotate it at various angles to see if a given image would match the template more closely and perform other transformations. Altering the image intensities to account for different daylight conditions has also extended this technique. In addition to intensify information, spatial context became an important component of this technique. Machine learning researchers have built genetic algorithms that perform these transformations on the templates to perform feature extraction

Statistical classifier feature extraction techniques look at features of an image and attempt to build a representative set of image elements for each classification represented as a multivariate probability distribution function. Each new pattern could then be run through the functions to determine the class of the given feature. Decision tree algorithms, K-Nearest Neighbor and Naïve Bayesian algorithms are current machine learning techniques that have evolved from statistical classifier techniques for image classification.

Trainable pattern classifier feature extraction techniques are inductive learners that accept patterns of different classes. These patterns are given to the algorithm and the algorithm attempts to build a predictor that when given new patterns it will classify them according to its experience gained in training.

2.5. Machine Learning

A branch of artificial intelligence known as machine learning has emerged as an effective discipline for feature extraction. This field evolved from the realm of mathematics particularly in statistics, computation complexity, and information theory as well as from many fields of science such as biology, chemistry, and physics (Mitchell, 1997). Many types of algorithms have evolved from this heavily researched field providing powerful tools to analyze information (Mitchell, 1997). This thesis focuses on using inductive learners for feature extraction.

The general form of an inductive learner accepts a set of examples depicting the features used for training from the hypothesis space. The hypothesis space for a learner is the set of all possible hypotheses (predictions). Each example consists of input and/or output values. The learner analyzes the training examples and builds a target function that models the training examples. The target function accepts an example, or set of examples, as input and produces an output based on its construction inferred from the existing knowledge gained by the training examples. In other words, the learner builds a model from the given set of inputs mapped to the given set of outputs and when given an unseen instance can infer the output. The inductive learning hypothesis states that given a sufficiently large set of training examples that map well to the target function of the hypothesis space the learner will approximate the target function well over unseen instances (Mitchell, 1997).

2.5.1. Problem Representation

For feature extraction of images using inductive learners, inputs are constructed from the images spectral intensity values from selected bands of the image. Given an image, examples can be constructed by selecting a point in an image and retrieving the intensity value of each pixel at that location from the image and assigning the set of input values an output value according to the assigned classification for that pixel. In a simple two-class image classification problem, the sample could be assigned as true or false with respect to the given class. This would form a training example. More elaborate training examples can be constructed by taking a subimage (window) from the existing image. This technique also accounts for spatial information by including neighboring pixels as part of a training example. Spatial context provides important information about the features shape and context. Together, spectral pattern recognition and spatial pattern recognition create the structure of the examples for the given learner. Image preprocessing is often performed on the image before (while) gathering test examples. These processing techniques exploit characteristics of the image that will facilitate the learner in realizing an ideal target function.

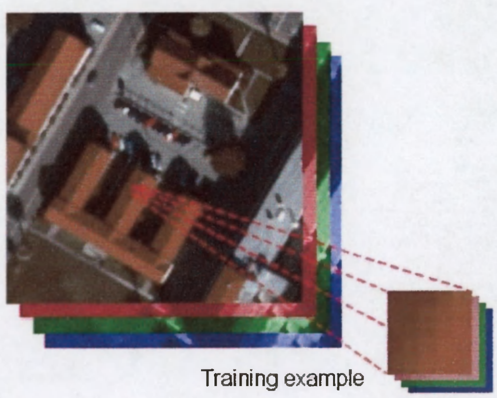


Figure 3: Sample 5x5 roof top training example

2.5.2. Artificial Neural Network

The artificial neural network (ANN) works well as a trainable pattern recognizer. The ANN is capable of learning pattern classes that are not linearly separable which is true of most pattern recognition classes in the real world. They are also robust to noisy training data and converge quickly to the target function. The ANN is a network of interconnected nodes. Each node contains a weight. Training examples are input into the network and produce an output classification. Back-propagation is the type of ANN used in this thesis. Back-propagation alters the weights by comparing the output of training examples with the output of the network at a given time and produces a vector of error terms that are then propagated back into the network to alter the weight in such a way that when the examples are sent into the network again the error term will be smaller. As the examples continue to evolve the network, the error term of the network approaches zero. The error term going to zero is limited by the size and structure of the network as well as the quality of the data. Examples are continuously fed into the network until a fixed number of iterations (epochs) or until convergence criteria is met. (Haykin, 1999)

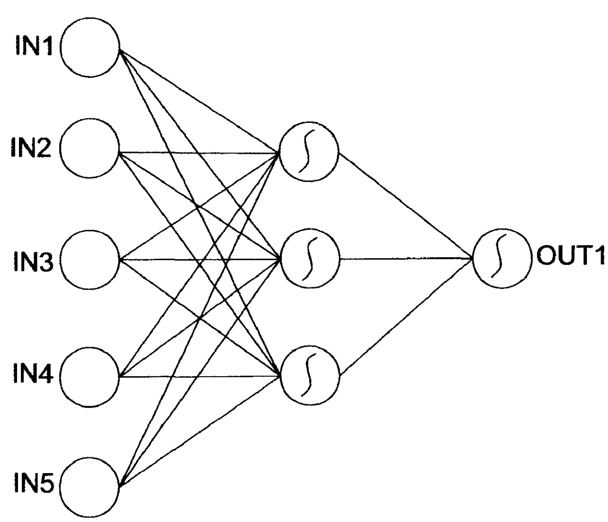


Figure 4: A neural network with 3 layers, 1 hidden layer

2.5.3. K-Nearest Neighbor

The k-nearest neighbor (KNN) also works well as a pattern recognizer. The KNN is an instanced-based learning algorithm. The algorithm accepts an unseen instance and compares it to the existing set of training examples using a distance metric such as the Euclidean distance. The k-nearest training examples is then used to classify the unseen instance as the most common classification. Since KNN does not build an explicit model each new unseen instance must be compared to the set of training examples available. For large sets of training examples, this can be very slow. KNN works well with small sets of training examples and has been found to work well on pruned sets of training examples using a hierarchical learning process where a neural network refines the initial features and then the KNN algorithm can be used to refine the results (Mangrich, 2001). KNN is robust to noisy training data and also is able to handle classes that are not linearly separable (Friedman et al, 1977; Mitchell, 1997).

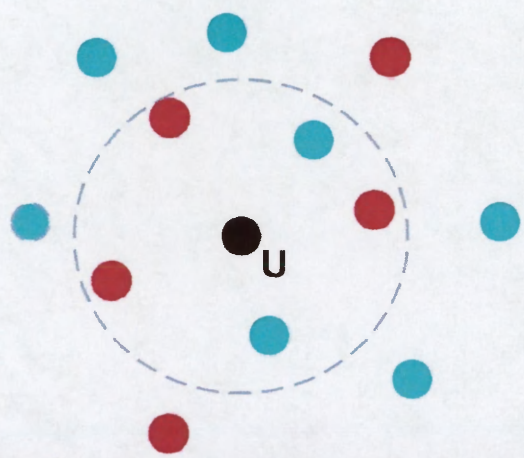


Figure 5: A 2-dimensional instance of k-nearest neighbor, k=5, where U is assigned the classification of red

2.5.4. Naïve Bayes

The Naïve Bayes (NB) classifier is the third type of pattern classifier. The NB classifier is a probabilistic learning algorithm. The NB classifier calculates probabilities of a set of classifications using a simplified version of Bayes theorem that naively assumes that attributes are probabilistically independent. The classifier computes the probabilities of each attribute value/class conjunction and multiplies them with the prior probabilities of the attribute values to produce an estimation of the entire set of attributes. The learner infers the most probable outcome using this simplified set of rules to classify unseen instances. This set of assumptions leads to a computationally efficient algorithm that can quite effectively classify a variety of problems especially where the attributes are independent. Even though this is an unrealistic assumption, the algorithm still produces acceptable results on a wide variety of problems including digital imagery. (Theodoridis, 1999)

2.5.5. Ensemble

An ensemble is an extension to traditional learning techniques. An ensemble combines multiple learners known as predictors. Each predictor is trained using a set of training examples. The predictors in the ensemble are each fed an unseen instance to produce predicted output. The output for each predictor is then compared and analyzed to determine the final output of the ensemble. To produce an effective ensemble, each predictor needs to produce as much disagreement as possible. This can be accomplished by varying the types of predictors or learners, varying the distribution of training examples, and/or modifying the training examples. This in effect produces bias so as to favor different properties in the learner. If the desired feature to learn was buildings, learners could be constructed that favor each roof type. These learners could then effectively find each roof type and then use a weighted average favoring the specialized learner to specify buildings. (Opitz, 1999a)

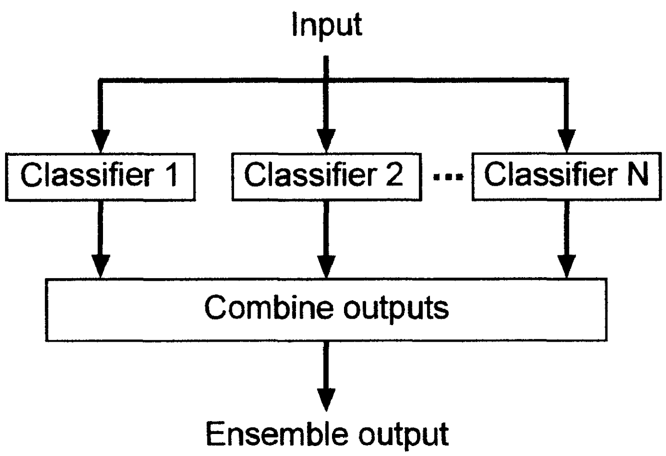


Figure 6: Sample ensemble

Bagging (Breiman, 1996) is the ensemble technique used in this thesis where each of the *N* classifiers is trained on a random resample (with replacement) of the original training set and then combined to produce a prediction of the output. Bagging has often been shown to reduce errors in the overall classification task better than a single classifier since the distribution of the resampled training sets produces sufficient diversity among the classifiers to reduce the impact of increased error rates among the individual classifiers. (Opitz, 1999c)

2.6. Change Detection

Change detection is an important concept for future feature extraction efforts. Dynamical systems and chaos theory are integrally connected with change detection. Computer imaging especially video compression algorithms exploit changes in a scene and perform difference operations to find the changes in images. This operation exploits the fact that there is usually little change from one image to the next. The differences can be stored instead of each image and hence reduce the size of the video greatly. Similarly, difference operations can illustrate the changes of images taken over time and represent additional knowledge for a learner to learn a disjunctive concept. These difference operations compare bands taken over periods of time. Difference operations can be taken from each set of images to produce a difference image that represents a phase change between images. Successive phase change images can then be used to illustrate changes between each phase. This knowledge can greatly help a learner find disjunctive concepts that change over time. (Lunetta, 1999)

2.6.1. History

Change detection for satellite images first analysis efforts attempted to detect changes in the luminosity and position of stars from Landsat MSS images taken at different time periods. The technique was further refined to visually analyze the various locations of space by filming the changes in position of the stars using color addition/subtraction on the changes. This process evolved from a purely visual interpretation process to an analytical process. Two primary change detection analysis methods emerged: post-classification and pre-classification methods.

2.6.2. Post-Classification Change Detection

Post-classification techniques perform categorization on each dataset independently and then analyze the differences between the classifications. Classification categorization techniques do not require any data normalization between multitemporal datasets, but do often require additional time to separately classify each set. Classification can also be different between the multitemporal data due to changes in the satellites making it difficult to compare the separate classifications. The separate classifications of each image when combined also propagate errors due to the errors inherent in the separate classifications.

2.6.3. Pre-Classification Change Detection

Pre-classification techniques work with multitemporal data directly performing the necessary analysis and transformations to the data and then produce classifications based on the multitemporal data. Five basic pre-classification methods currently exist: composite analysis, image differencing, principle component analysis, change vector analysis, and spectral mixing analysis.

Composite analysis performs a single analysis of a multitemporal dataset where the dataset is collected under similar conditions from different years. Composite analysis uses standard pattern recognition and spectral classification.

Image differencing performs a difference operation between two multitemporal datasets producing a map of the degrees of change between two time periods. Image differencing is often accompanied by a pre-processing step that normalizes two images to correct different sampling parameters.

Principle component analysis is a data redundancy technique that explores multitemporal data and locates the principle components in the data that accounts for the greatest amount of variance in the data. Subsequently, additional components in the data are accounted for to provide additional information to help with classification.

Change vector analysis looks at the changes in multitemporal data as it changes over time storing the changes from one time step to another in change vectors. The lengths of the change vectors are analyzed to determine the magnitude of change and the direction of the given vector that indicate the nature of change.

Spectral mixing analysis looks at high spectral resolution multitemporal data combining multiple spectral signatures of the data that contribute to the overall reflectance of the image that can be analyzed to classify subtle land cover changes.

This thesis focuses on image differencing using satellite imagery taken over extended periods of time with the primary intent to extract changes that have occurred between the time periods.

3. ADAPTIVE FEATURE EXTRACTION

Satellites have been creating images of the earth for many years. Millions of images have been taken of different locations at different times with varying environmental conditions using widely varying equipment and techniques. This imagery has been used to analyze the earth to help scientists to understand the evolving environmental conditions of earth. Due to the dynamics of the earth, previous feature extraction efforts quickly become outdated as new imagery becomes available. New technologies also provide higher resolution imagery with additional sensors providing the capabilities to analyze the image much more thoroughly than was possible in the past. Thus there is great need to refine techniques that will allow existing feature extraction efforts to facilitate enhanced feature extraction efforts.

This thesis presents an adaptive feature extraction approach that uses change detection techniques to incorporate changes into the learning process to iteratively refine the feature extraction process. Additionally, image compression is used to simplify the learning process and to adapt existing feature extraction efforts at lower resolutions to higher resolution images. This in effect improves the classification of the images at the higher resolution by refining the coarser approximations made at lower resolutions.

Image processing is the first component of the adaptive feature extraction approach. The images are first projected into the same coordinate system and referenced so that each point in space of the image coincides to the same point of space in the other image. The images are then normalized to minimize differences in sampling parameters. Any additional image preprocessing techniques are then used to enhance the bias of the features being extracted for the learner from the images.

Change detection is the second component of the adaptive feature extraction approach. Change detection first analyzes two images taken at different periods of time and produces a phase change image, or difference image, that represents the changes over time. These changes direct the learner to features that change at various levels. Typically, not all of the features in an image change at the same rate. For instance, in a small city that has had recent development. The buildings will contain large differences where new development occurs and help a learner learn the new development. The resulting phase change image can then be incorporated into the learning process as additional information for the learner.

Image resolutions continue to increase as satellite technology evolves providing much more information. Adapting existing feature extraction efforts requires adjustments to existing images and feature extraction efforts. Image compression allows images to be compressed to fit the existing images. Learners can then transfer existing knowledge gained from previous efforts to the new images. The new images can then be uncompressed progressively along with the new features extracted from previous efforts.

Multiple predictors will be needed in each phase to classify each feature. Separate learners are used to classify each feature. The learners will then combine their output forming an ensemble to determine the final image classifications for the various classes. Ensembles will be used at each phase to classify the individual features. Bagging will also be used on combinations of the classifiers to produce a unique classification for each pixel in the image. Similarly, multiple ensembles can also be combined to produce a unique classification for each pixel in the image.

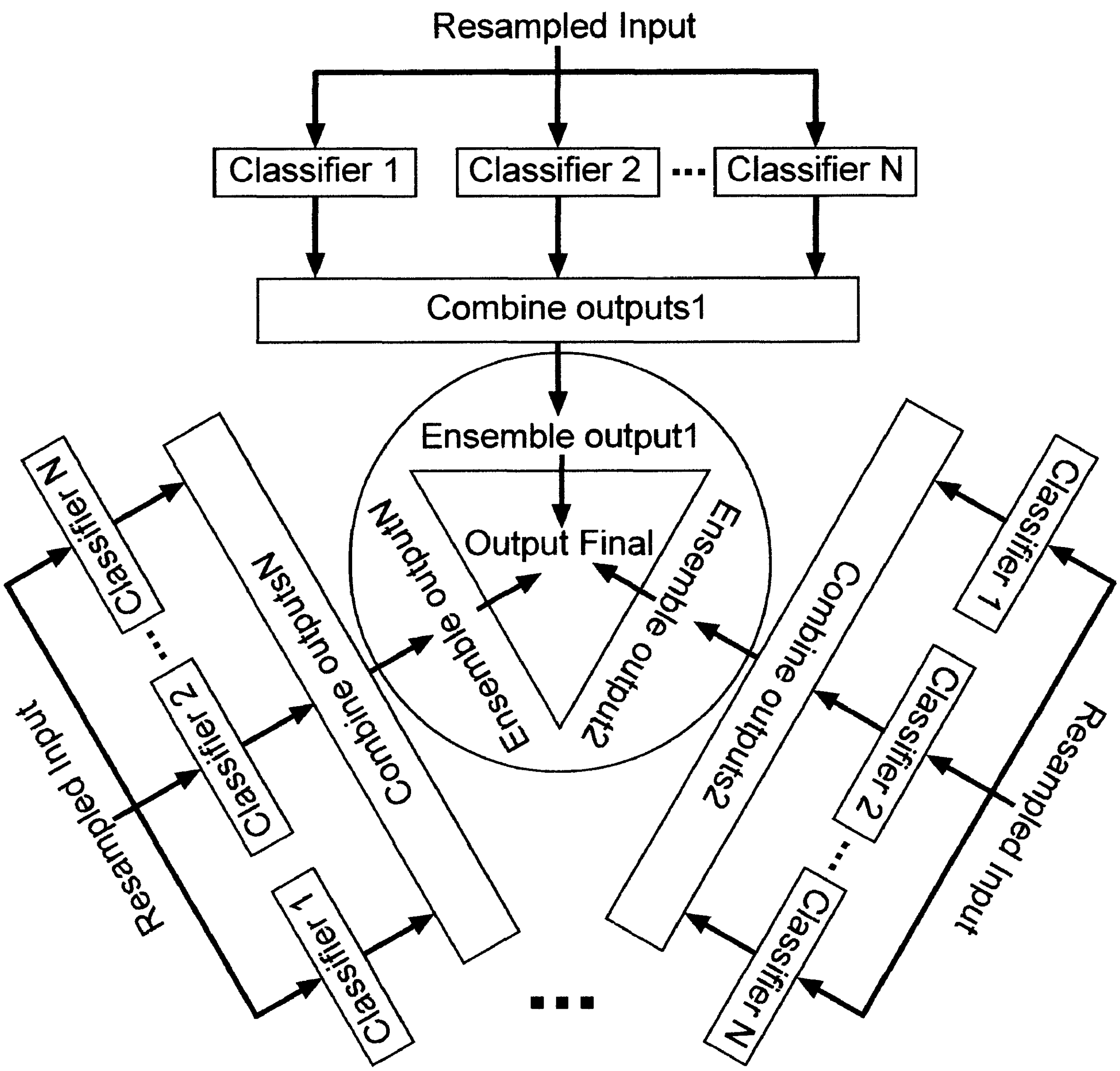


Figure 7: Collection of Ensembles Combining Predictions

4. METHODOLOGY

Three study areas are presented for the adaptive feature extraction techniques: (1) and (2) classifying the severity of wildfire to vegetation, and (3) classifying built-up area in two urban scenes at low resolution. The experiments illustrate learning to classify changes in features using temporal satellite imagery and the utility of adapting the images to the learners to improve the classification.

4.1. General Methodology

This section describes the general parameters for all three experiments performed in this thesis. The 30m LandSat images for each experiment were geo-rectified to previously terrain-corrected images using Erdas Imagine software and then clipped to the scene of interest. The before-and-after images were corrected for atmospheric scattering using histogram equalization (Seul, 2000). Difference images were constructed from these corrected images by taking the absolute difference of each band from the before-and-after images.

Training examples were selected by visually interpreting the image. Existing lifeform (tree, grass, shrub, etc.) layers and expert analysts assisted in the interpretation of the images to come up with a set of training examples and a truth image. Examples were constructed by defining point and polygonal representations (shapes) of the features using ESRI ArcMap GIS software and were then converted to Boolean mask bands for each feature. Inputs of the selected examples consisted of each band from the after satellite image and optionally each band of the difference image. The inputs along with the feature classification represented the training example for each feature.

The ANN classifier settings consisted of a learning rate of 0.1, momentum of 0.9, and a single hidden layer topology with a real-valued output between 0 and 1 (Mitchell, 1997). The KNN classifier used 5 neighbors with inverse squared real-valued distance weighting. The NB classifier computed outputs based on the real valued inputs merged into 100 values. The outputs from the inductive classifiers are real valued numbers normalized between 0 and 1. Percent error was computed by comparing the predicted output image to the truth image for each feature of interest.

For each individual classifier type (ANN, KNN, NB) experiment the final classification of a given pixel was determined by individually training an instance of the classifier type for each feature of interest. Examples were constructed such that the given feature examples were positive and all other features of interest were negative for the classifier designated to learn the given feature. Each learner was trained on the assigned feature and all outputs of the classifier were compared. The dominant learner (learner with the highest predicted output) then classified the pixel with its output type. This was done for each pixel in the scene and represents a simple ensemble classifier.

The ensembles experiments each used 9 classifiers composed of 3 ANN, 3 KNN, and 3 NB ensemble classifiers arranged as depicted in Figure 7. Each ensemble classifier's input consisted of random samples with replacement equal to the size of the original training set. Each classifier type formed another ensemble where the majority output classification for each of the ensemble classifiers represented the vote for the 3 learner ensemble. The outputs of each of the classifier types were then compared and a dominant type was selected. For the final output from the ensemble each classifier type ensemble (3 ANN, 3 KNN, 3 NB) voted on the predicted output type and the majority output type was selected.

4.1.1. Receiver Operating Characteristic Curves

A receiver operating characteristic (ROC) curve is an effective representation of the quality of the discriminatory power of a classification algorithm when looking at the positive and negative results of the learning algorithm on a given set of data under various evaluations of the output of the classifier. To construct the ROC curve the real valued output is tested at different thresholds and then plotted with the true positives on the y-axis and the false positives on the x-axis. The area under the ROC curve is approximated using the trapezoid rule on the true/false positive points and helps determine the effectiveness of the leaner at learning the given task and will be used to compare the effectiveness of the different learners to one another. ROC analysis removes the inductive bias of each of the learners and costs of unknown and unequal classification error. (Maloof, 2002)

		Truth: Yes	Truth: No
Prediction	Yes	True Positive	False Positive
	No	False Negative	True Negative

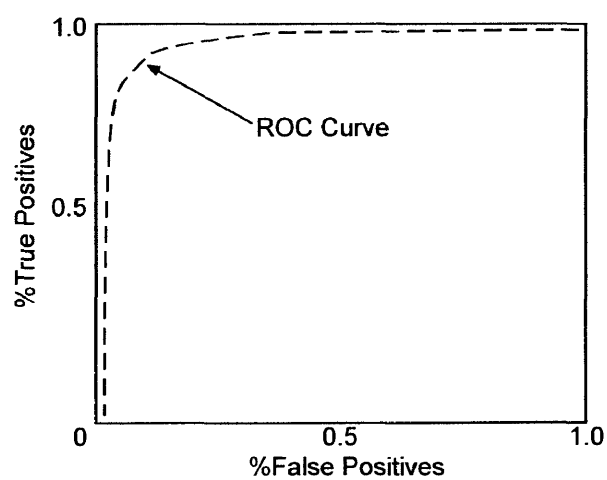


Figure 8: A sample confusion matrix and ROC curve

4.2. Ashland Fire Experiment

This experiment presents a scenario that explores the utility of including a phase change image as additional input for the learner. Given a set of Landsat-7 TM images of an area near Ashland Montana now known as the Fort Howes fire complex due to a 15,000-acre fire that burned in late July 2000 classify the fire burn characteristics. The images are snapshots of the area before the fire and after the fire. The area consists of two different Landsat-7 TM scenes, Path 35/Row 28 and Path 35/Row 29.

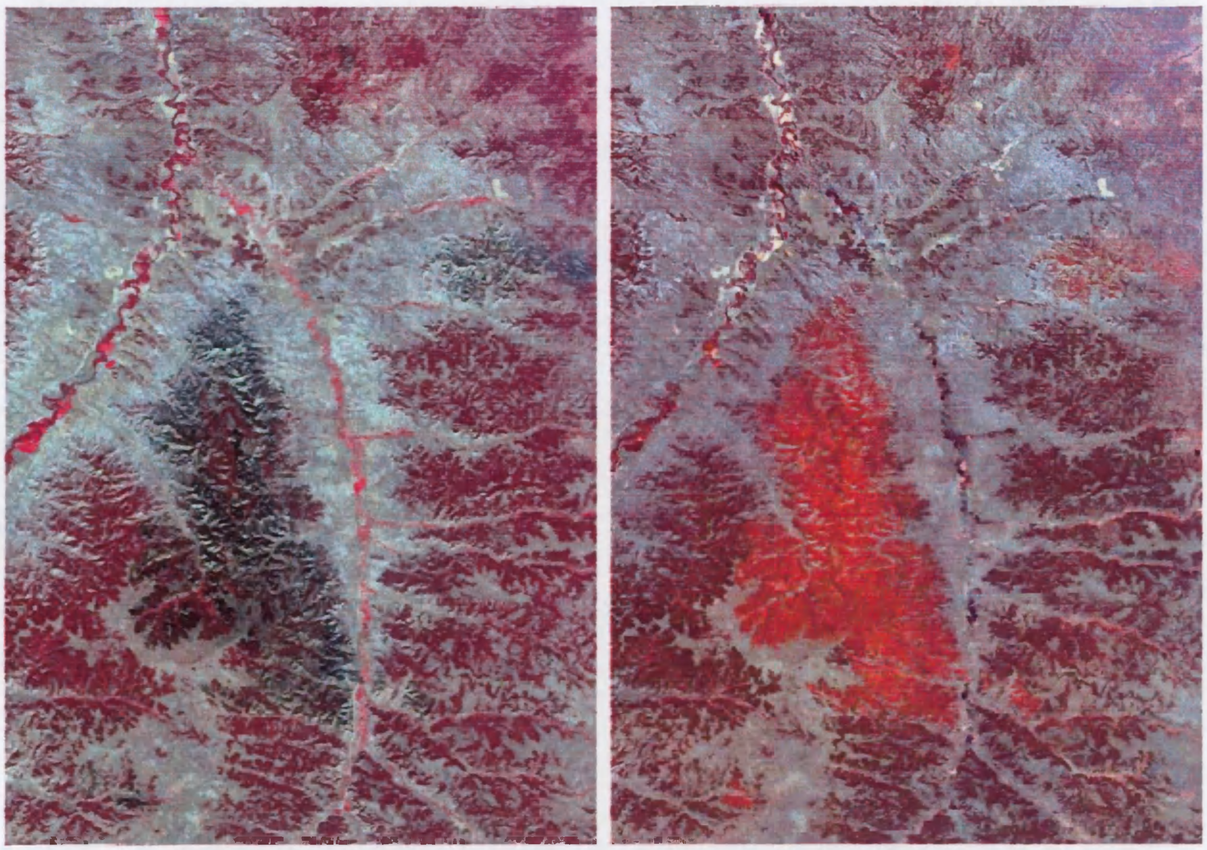


Figure 9: Ashland after fire image and difference image using bands 4, 3 and 2 as R, G and B

4.2.1. Task

The task is to classify the after fire image into eight classes:

1. Unburned Tree
2. Mixed Burn Tree (mosaic burn)
3. Burned Tree
4. Burned Grassland
5. Unburned Grassland
6. Burned Shrub Land
7. Unburned Shrub Land
8. Barren

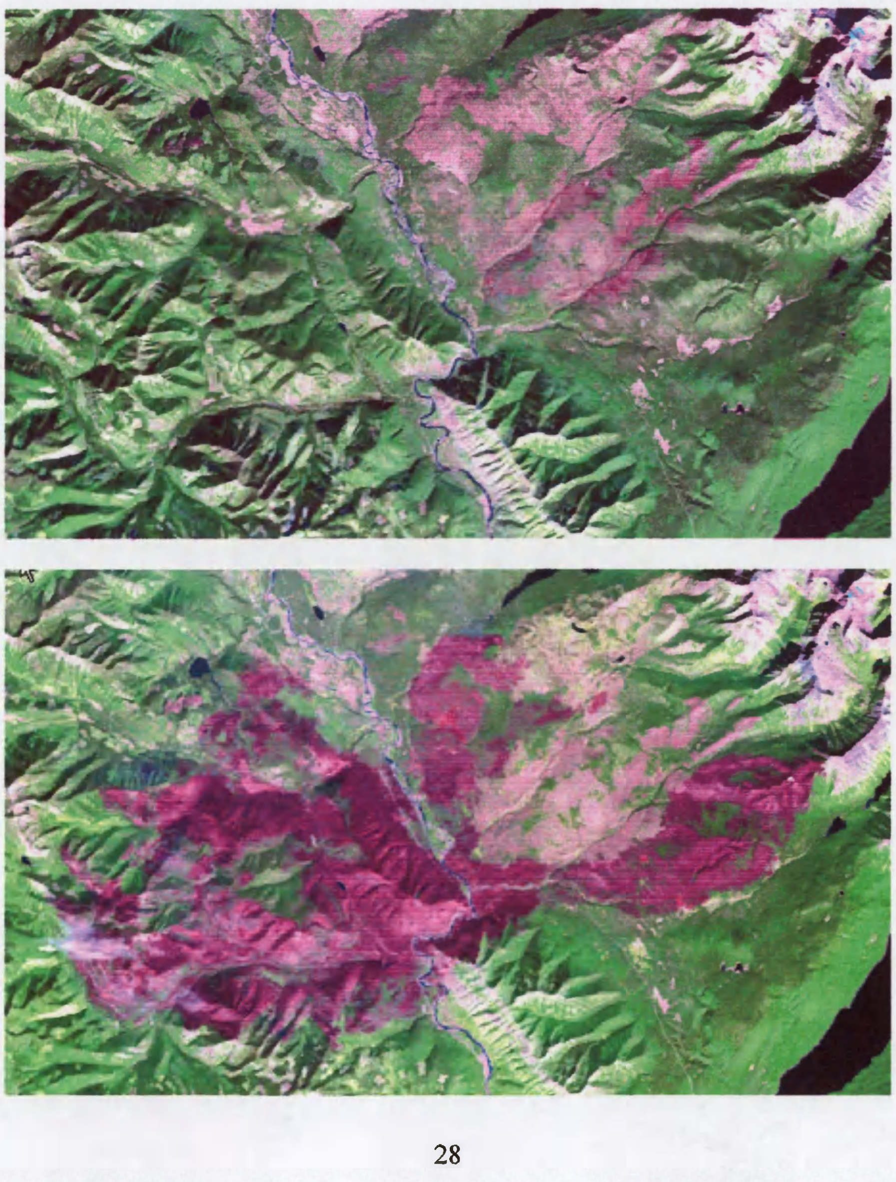
This training set consists of a set of examples (approximately 1,000 pixels from each layer for each class) selected by an analyst who derived them using air photo interpretation and an existing life form layer created before the fire.

4.2.2. Specific Methodology

Two experiments were conducted for this image set. The first experiment looked at only the after fire image. The second experiment compared the two images and produced a phase change image. This phase change image was then added as additional bands to the after fire image. The features were relatively small so inputs consisted of subimages of size 3 pixels by 3 pixels from each layer. Each classifier was trained using the standard bands and the standard bands with the phase change image. The results of the learner's classifications were combined as an ensemble and then classified each of the pixels into one of the eight classes giving the classification to the learner with the highest prediction.

4.3. Moose Fire Experiment

This experiment presents a scenario that also explores the utility of including a phase change image as additional input for the learner over the existing technique of principle component analysis. Given a set of Landsat-7 TM images of an area near Columbia Falls Montana now known as the Moose fire due to a 71,000-acre fire that burned in summer of 2001 classify the fire burn characteristics. The images are snapshots of the area before the fire and after the fire. The area consists of two different Landsat-7 TM scenes.



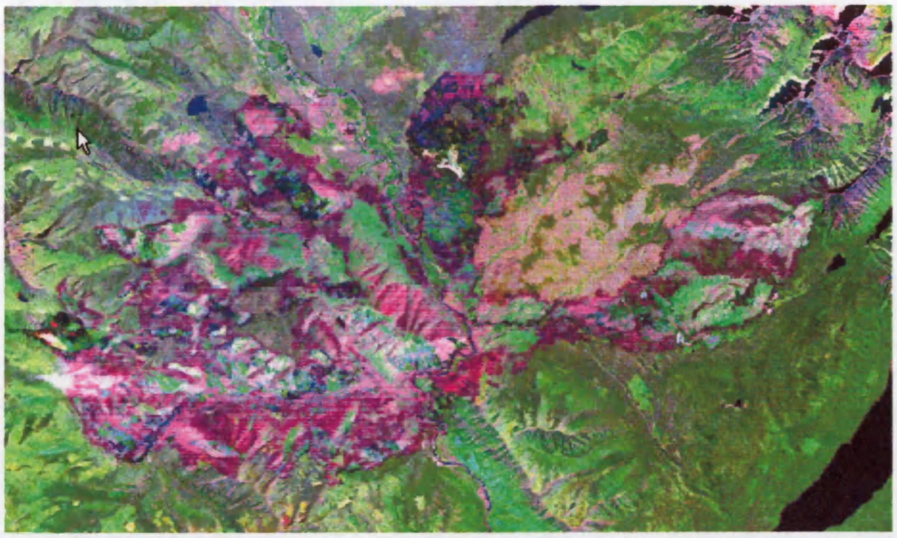


Figure 10: Moose before fire, post fire and difference images

4.3.1. Task

The task is to classify the after fire image into nine classes:

1. Unburned Tree
2. Mixed Burn Tree (mosaic burn)
3. Burned Tree
4. Burned Grassland
5. Unburned Grassland
6. Burned Shrub Land
7. Unburned Shrub Land
8. Barren
9. Water

This training set consists of a set of examples (approximately 10,000 pixels from each layer for each class) selected by an analyst who derived them using air photo interpretation and a life form layer created before the fire.

4.3.2. Specific Methodology

Two experiments were conducted for this image set. The first experiment compared the two images and produced a phase change image that was combined with the after fire image. The second experiment used principle component analysis. Each classification was trained using the standard bands with the phase change image. The principle component analysis used bands 4 and 5 to derive a fire perimeter which was then combined with a lifeform layer to separate the burned and unburned vegetation types out for each lifeform (tree, grass, and shrub). The tree class was further divided into mixed and burned. The results of the learner's classifications were compared to the results of the principle component analysis to show the utility of using the learners to learn change detection versus the more traditional principle component analysis.

4.4. Land Development Experiment

This experiment presents a scenario that explores the utility of including phase change images as additional input into the learner for land development. Given two LANDSAT-5 TM images taken over the city of Missoula spanning a 12-year period classify the changes in land development. The first image was taken in 1984 and the second image was taken in 1996.

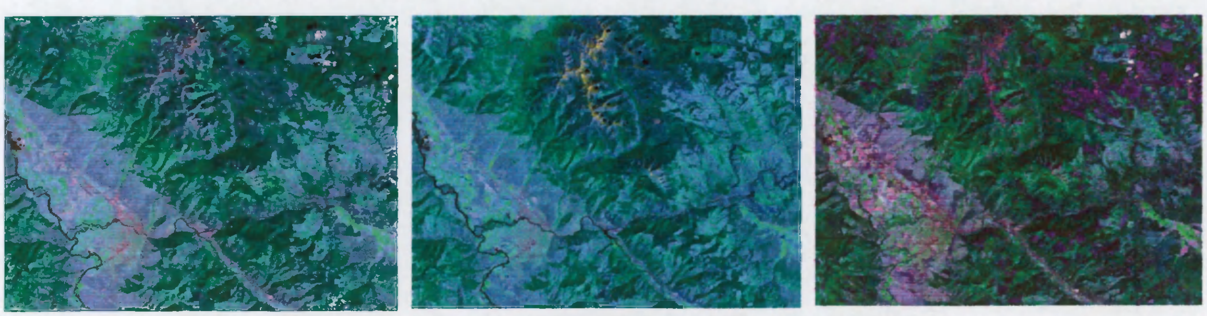


Figure 11: Missoula 1984, 1996 and difference images

4.4.1. Task

The task is to classify the after phase change image and the 1996 image into four classes:

1. No Change
2. Low Change
3. Moderate Change
4. High Change

This training set consists of a set of examples (approximately 15,000 pixels from each layer for each class) selected by an analyst who derived them using air photo interpretation.

4.4.2. Specific Methodology

This experiment compared the two images and produced a phase change image. This phase change image was then added as an additional band to the 1996 image. The features were relatively small so inputs consisted of subimages of size 3 pixels by 3 pixels for each layer. The training examples and truth image were constructed using the lifeform layer as well as city maps of Missoula and the surrounding area. Each classification was trained using the standard bands and the phase change image. These learners were combined as an ensemble and then classified each of the pixels into one of the four classes giving the classification to the learner with the highest prediction.

5. RESULTS AND DISCUSSION

Results for the three study areas are presented to illustrate the utility of the adaptive feature extraction approach in using the difference images and image compression to improve the feature extraction results. The first two experiments illustrate the use of difference images in classifying temporal features of interest for fires while the third experiment presents the utility of classifying temporal features of interest for land development in a scene.

5.1. Ashland Fire Experiment

Figure 12 illustrates the after fire scene as a false color image to highlight the areas that burned where bands 4, 5, and 3 represent the red, green, and blue components of the image. The maroon hues represent the living vegetation and the dark green hues represent the burn areas where the darker the green the more intense the burn. Figure 13 highlights the predicted burn fire perimeter area and was constructed using the hierarchical learning process with the available burn point data.

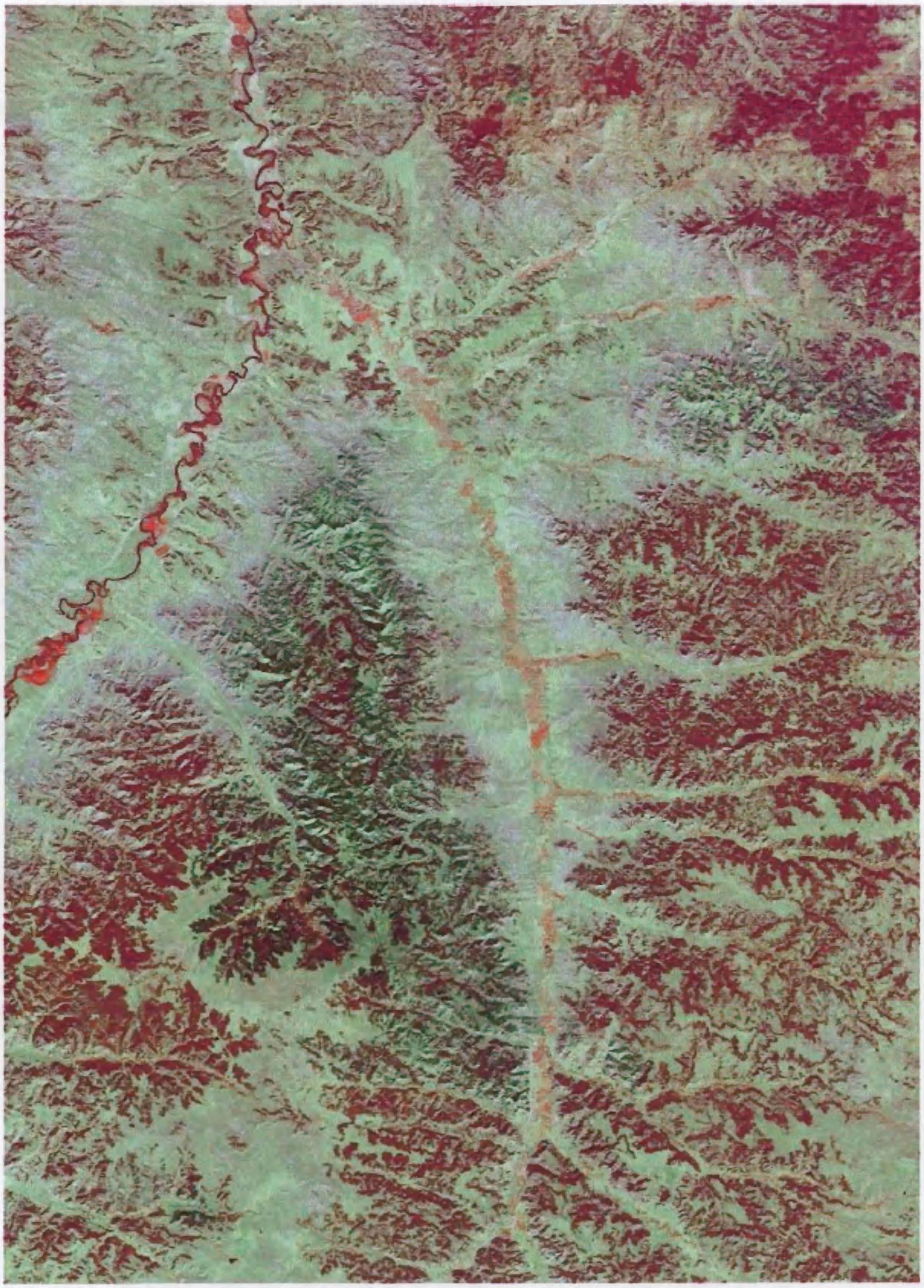
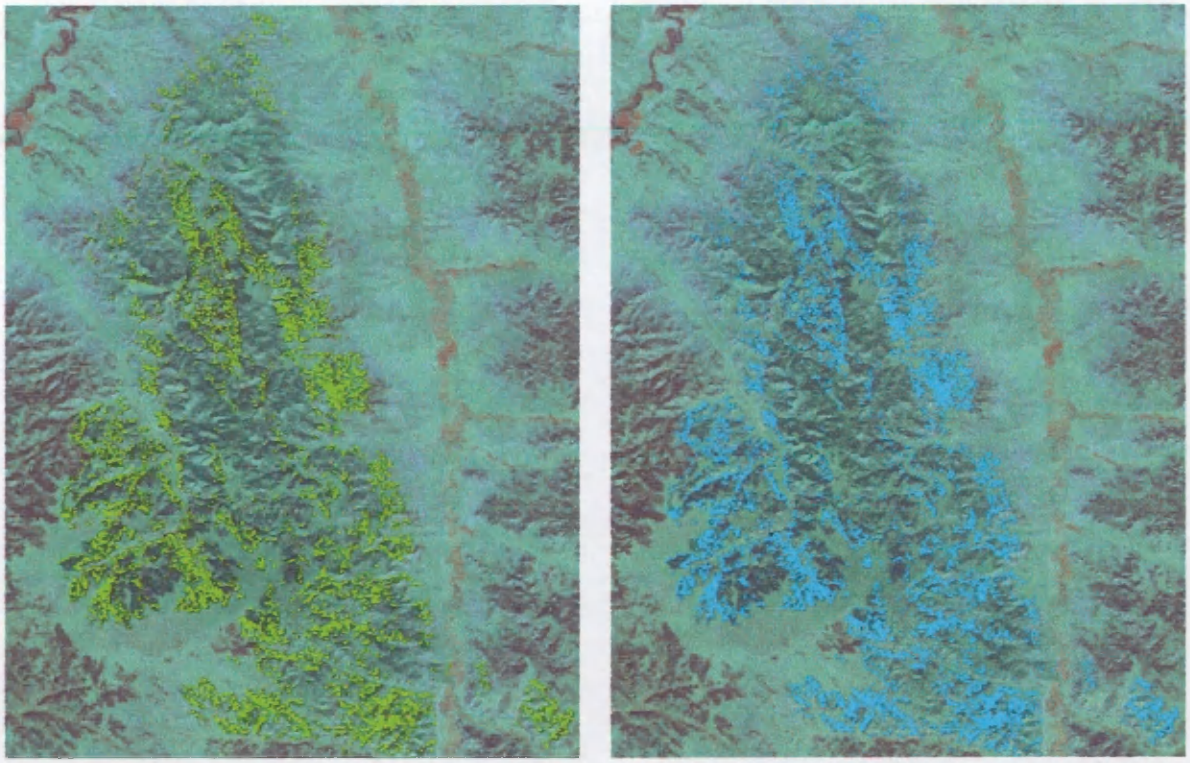


Figure 12: Ashland after fire false positive image, bands 4, 5, 3 (RGB)

Figure 13: Ashland fire burn perimeter inferred using hierarchical learning process.



(A) (B)

Figure 14: Results for Ashland mixed burn classification. 12(A) Represents the after fire experiment and 12(B) represents the difference image experiment.

Table 1 illustrates the mean percent correct for all of the learners for the after fire experiment and the after and difference image experiment. In the Figure 14 image A represents the mixed burn classification for the after fire image and image B represents the difference fire image results. The differenced image improved the classification results by 9 percent.

Classification	After fire image	After and difference image
Unburned Tree	63	70
Mixed Burn Tree	54	57
Burned Tree	83	80
Burned Grassland	50	75
Unburned Grassland	75	85
Burned Shrub Land	80	100
Unburned Shrub Land	85	90
Barren	72	78
Mean	70	79

Table 1: Mean percent correct by learners for each class

The figures below illustrate the effectiveness of predicting the classification task for each type for each of the classifiers. Figure 15, Figure 16, Figure 17, and Figure 18 depicts the ROC curves for the KNN, ANN, NB and ensemble respectively and Figure 19 depicts the approximate area under the ROC curve for the respective classifiers.

The barren classification was particularly difficult for KNN and can likely be attributed to the limited number of example compared with the actual number of true barren points. The ANN and NB classifier were able, however, to learn the concept. The ensemble seemed to average the effect out and improve overall.

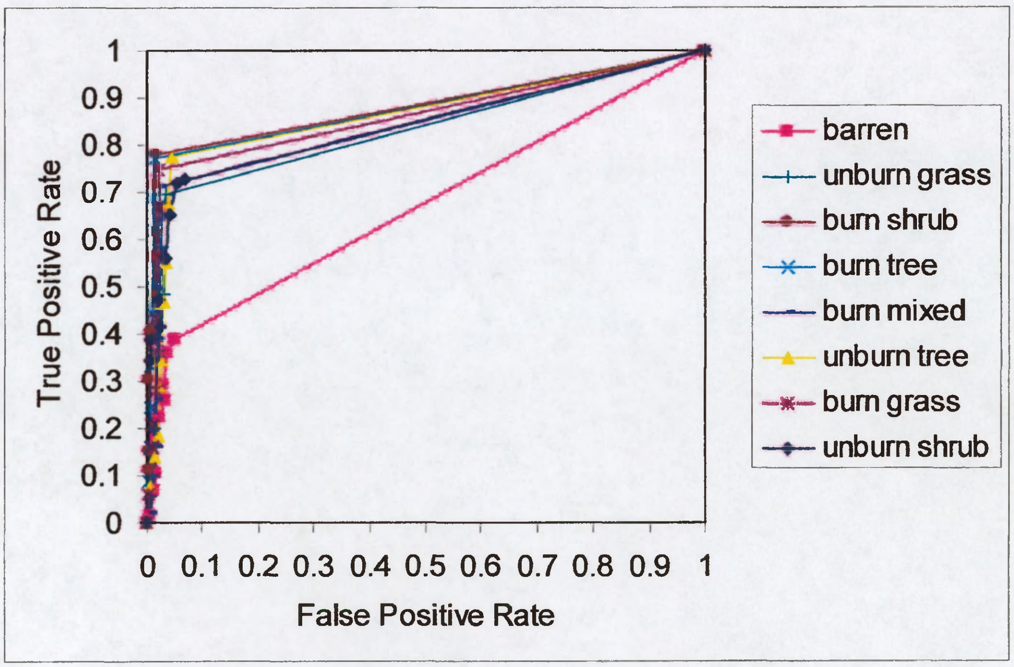


Figure 15: ROC curves for Ashland KNN after fire experiment

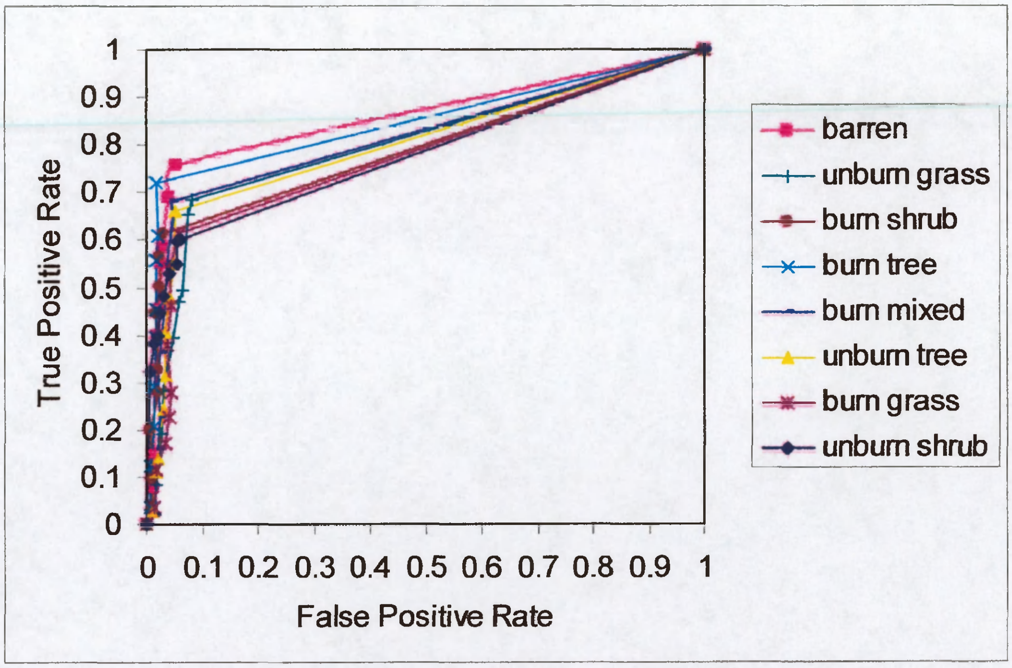


Figure 16: ROC curves for ANN Ashland after fire experiment

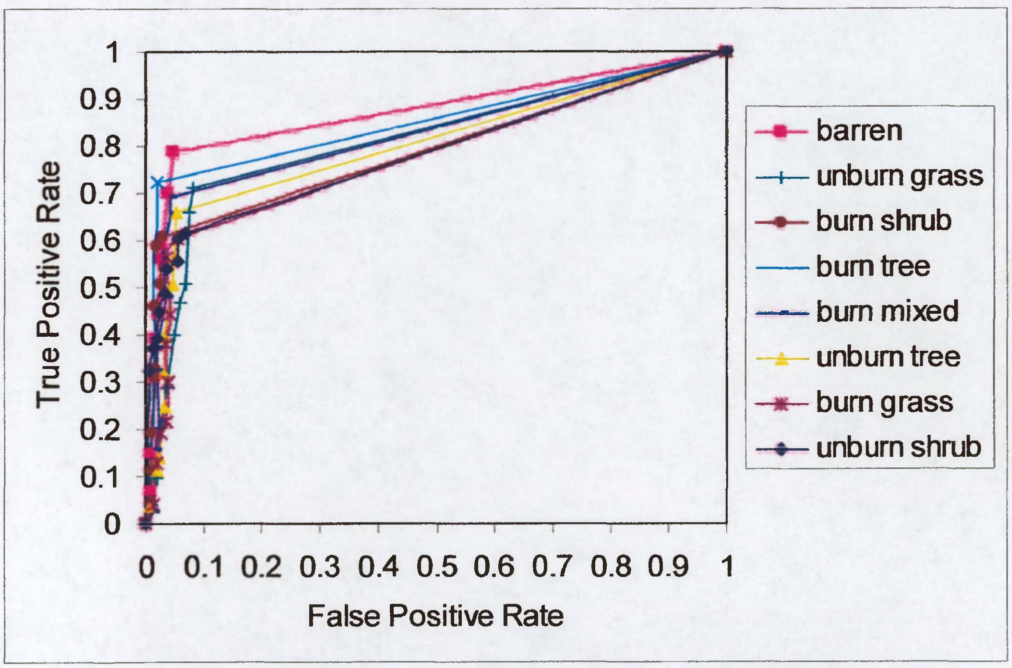


Figure 17: ROC curves for NB Ashland after fire experiment

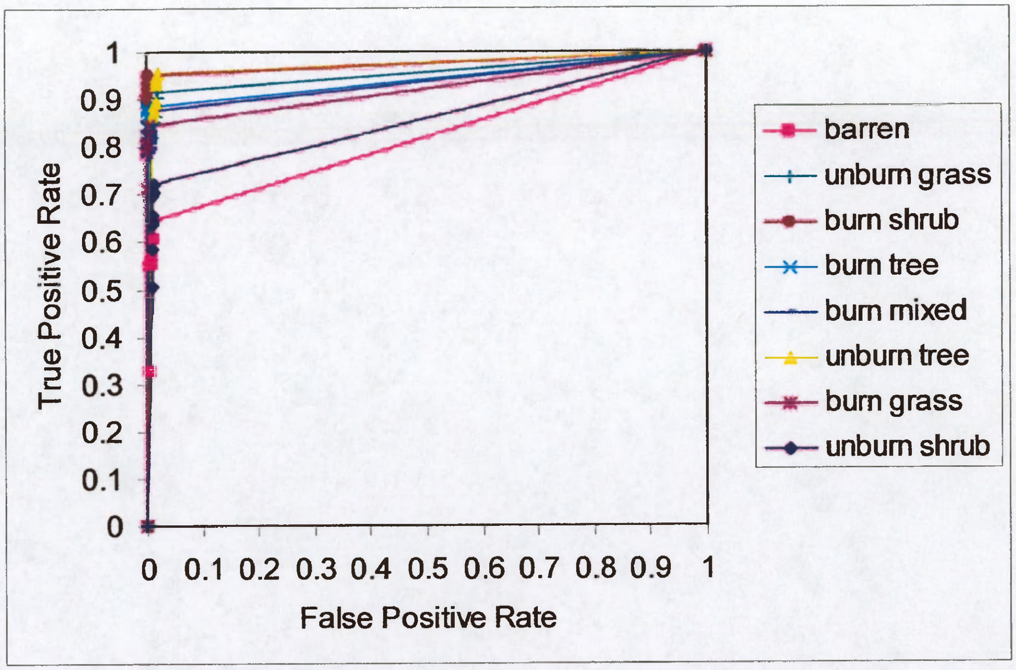


Figure 18: ROC curves for 9-learner ensemble after fire experiment

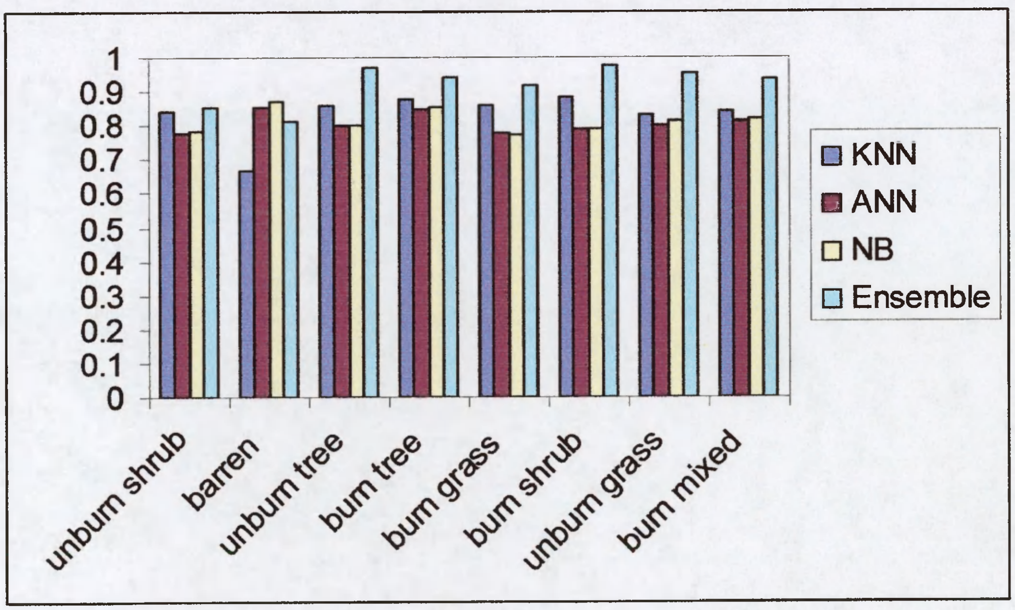


Figure 19: Approximate area under ROC curves for after fire experiment

5.2. Moose Fire Experiment

Figure 20 illustrates the changes that occurred over the moose fire scene from before and after the fire as a false color image using bands 4, 3, and 2 for the red, green and blue bands. The scene shows the changes in the red bands where the darker the purple the more the fire burned a given vegetation class. After one repetition of the hierarchical learning technique the changes in vegetation and the changes in development were classified with a high degree of accuracy.

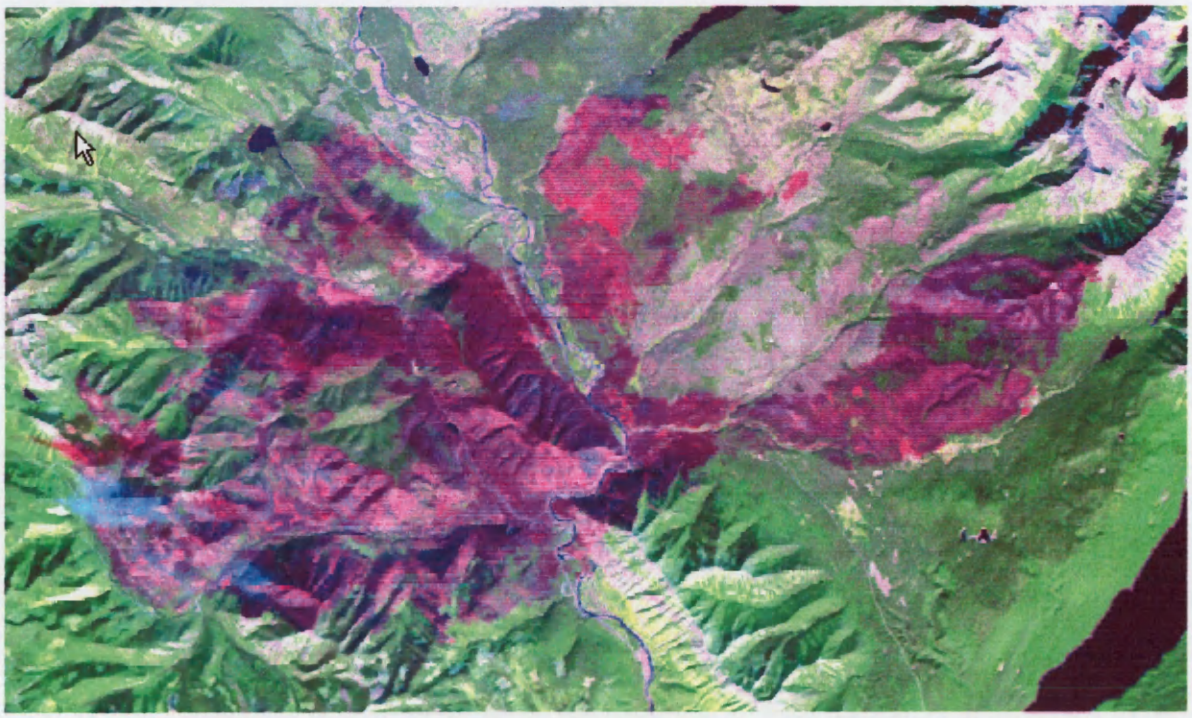


Figure 20: Missoula difference false color image

Table 2 represents the percent correct of each output class of each classifier as well as the percent correct for the principle component analysis of each output class. The figures below represent the classification of the dominant changes in the difference scene. Figure 21, Figure 22, Figure 23, and Figure 24 depicts the ROC curves for the KNN, ANN, NB and ensemble respectively and Figure 25 depicts the approximate area under the ROC curve for the respective classifiers.

The percent correct for each classifier of each output class shows that the ANN and KNN classifiers were effective at classifying the output classes. The PCA classifier was 20 percent less reliable on average from the lowest overall learner, the NB classifier. The ensemble improved the overall accuracy of the learners and illustrated the effectiveness at applying the bagging technique to a variety of classification tasks given all of the classifiers and a random resample of the input data for each classifier.

Classification	ANN	KNN	NB	Ensemble	PCA
Unburned Tree	93	93	85	96	77
Mixed Burn Tree	82	91	74	91	68
Burned Tree	77	78	70	88	82
Burned Grassland	94	94	74	99	54
Unburned Grassland	93	92	75	98	47
Burned Shrub Land	83	85	84	92	74
Unburned Shrub Land	93	92	85	95	66
Barren	96	97	82	98	59
Water	98	98	99	100	57
Mean	90	91	81	95	66

Table 2: Percent Correct for each classifier

All three classifiers did well overall though the mixed burn class was particularly challenging. The learners often misclassified mixed burn as tree burn or shrub burn. Grass output classes were better represented in the truth points and hence did much better this time than the Ashland experiment. Similarly, mixed burn was better represented in the Ashland experiment and thus did much better than the Moose experiment. In both experiments it was difficult to find a good separation in the mixed burn class since the mixed state varied so much.

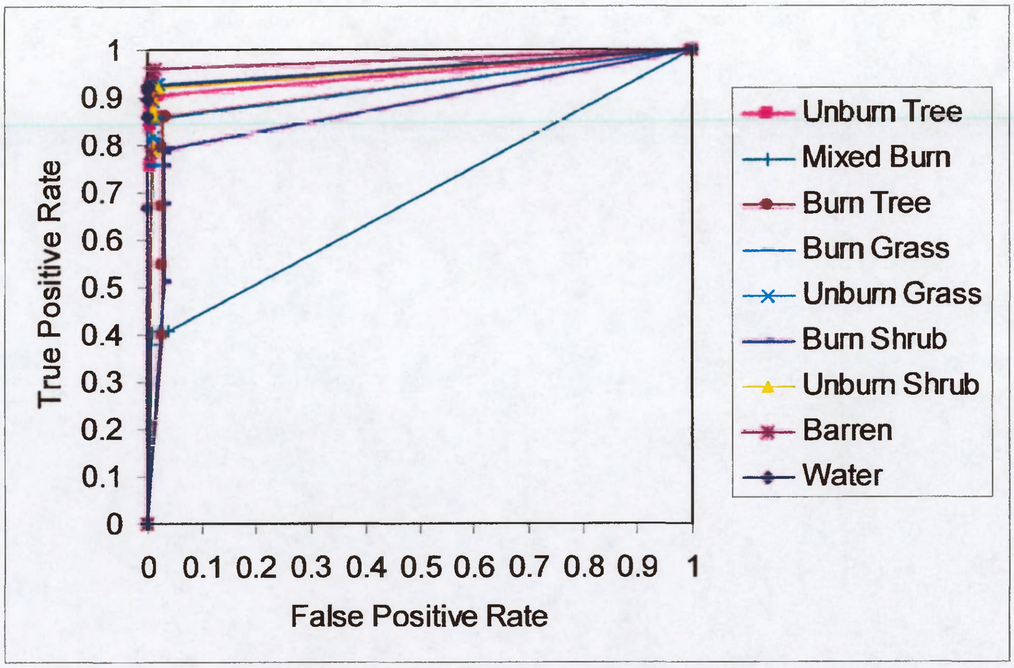


Figure 21: ROC curves for KNN Moose fire experiment

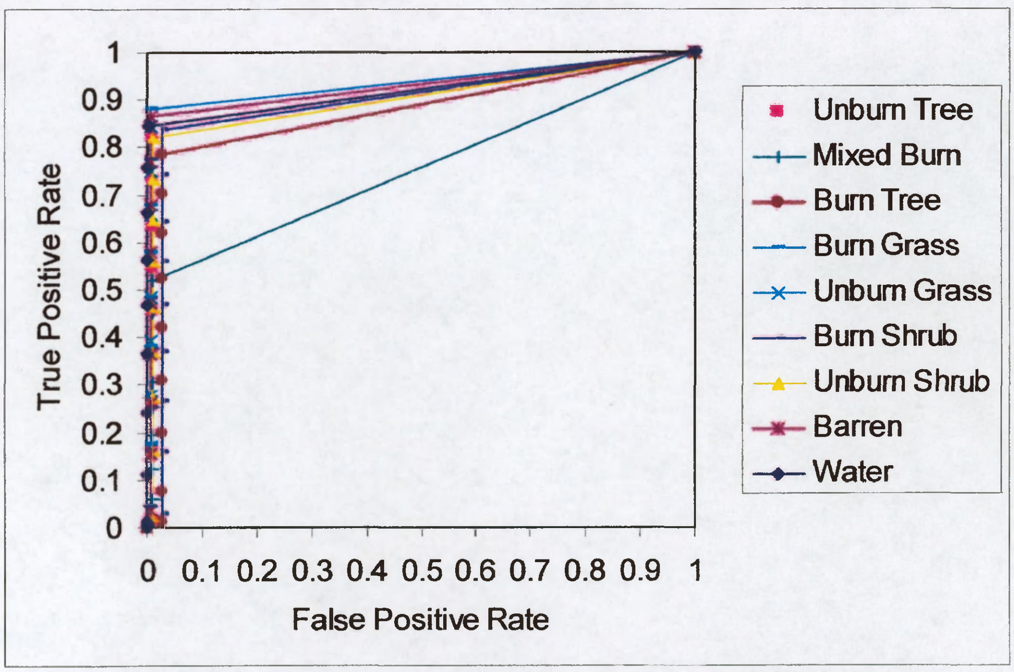


Figure 22: ROC curves for ANN Moose fire experiment

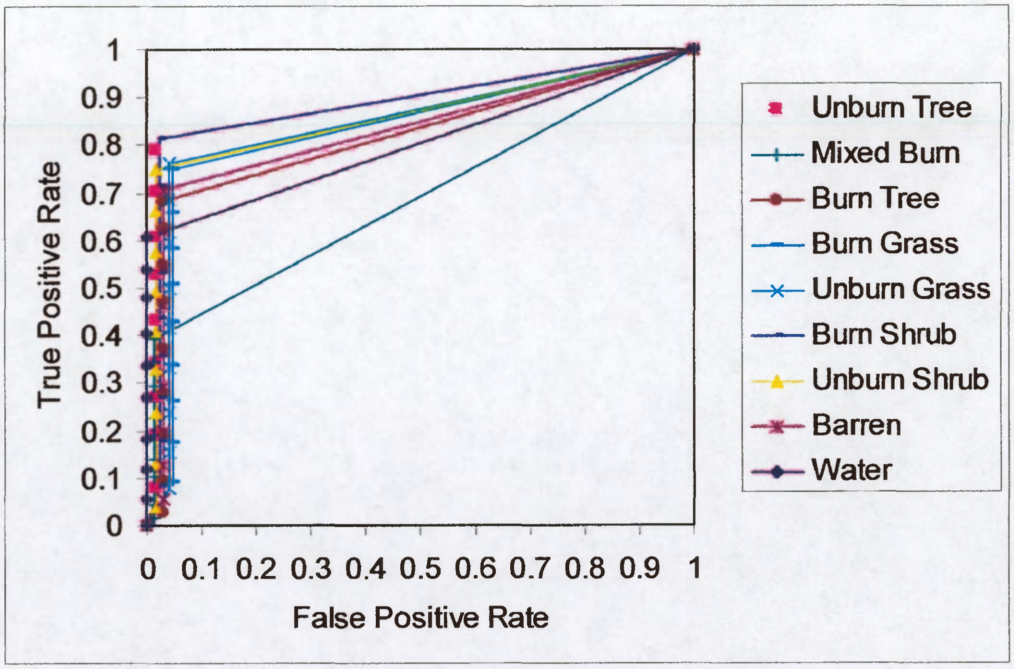


Figure 23: ROC curves for NB Moose fire experiment

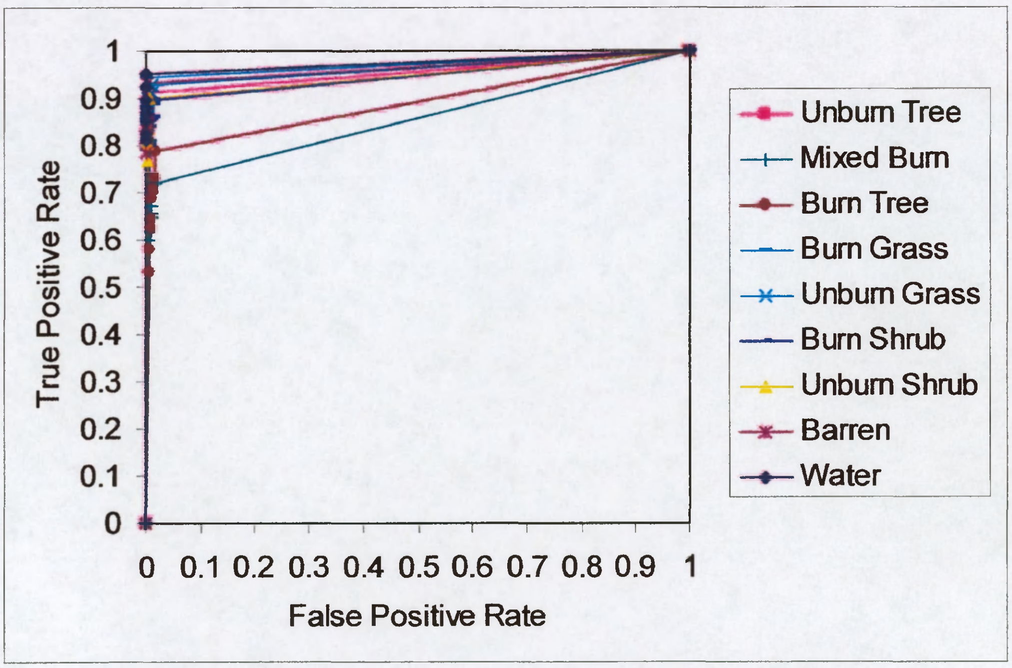


Figure 24: ROC curves for 9-learner ensemble Moose fire experiment

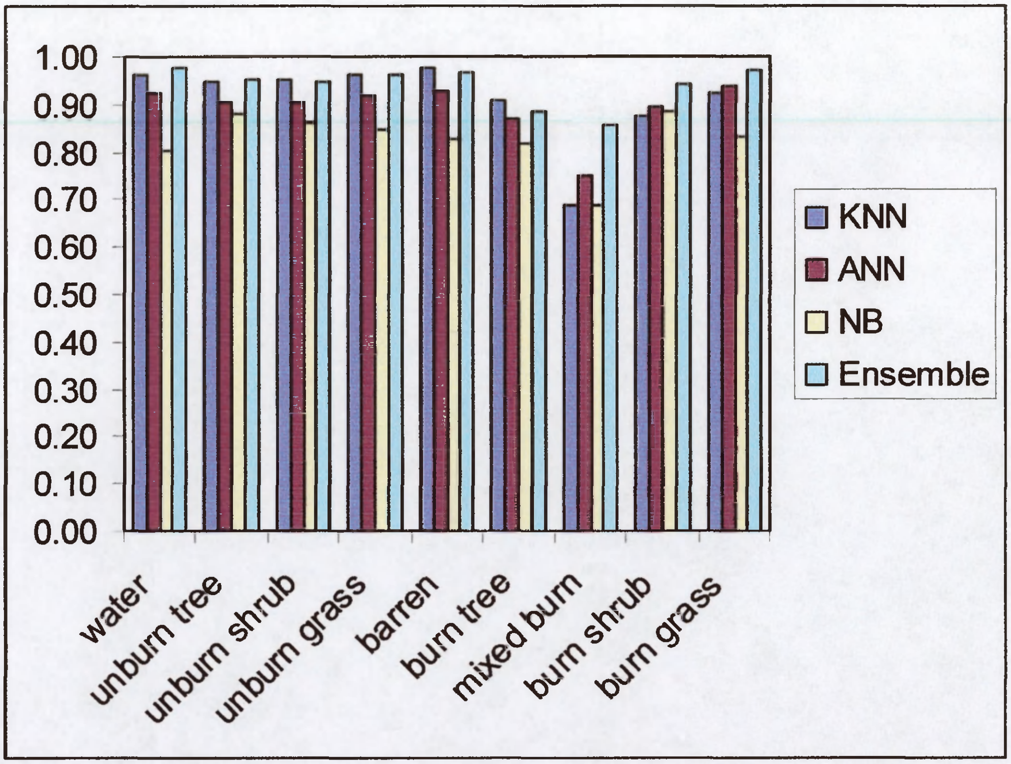


Figure 25: Approximate area under ROC curves for Moose fire experiment

5.3. Land Development Experiment

Figure 26 illustrates the changes that occurred over the Missoula scene from 1984 to 1996 as a false color image using bands 4, 3, and 2 for the red, green and blue bands. The scene shows the changes in the red bands where the darker the red the more the change that occurred particularly with vegetation changes and land development. The green band also depicted changes in development. After one repetition of the hierarchical learning technique the changes in vegetation and the changes in development were classified with a high degree of accuracy.

Figure 26: Missoula difference false color image

The figures below represent the classification of the dominant changes in the difference scene for each learner. Figure 27, Figure 28, Figure 29, and Figure 30 depicts the ROC curves for the KNN, ANN, NB and ensemble respectively and Figure 31 depicts the approximate area under the ROC curve for the respective classifiers.

The KNN performed the best of the three classifiers overall and each learner had the greatest difficulty discriminating between the no development and high development due to the inherent complexity of selecting a good breaking point from no development to low development and moderate development to high development. In the case of no development there was a great deal of mixture of features from roads to streams to vegetation to development to merge together. Breaking the vegetation, roads, streams, and existing development classes out would likely help these results. The ensemble was able to generalize much better while the KNN outputs illustrated a very large separation between no and low development and moderate and high development. Polygons were used to outline the boundaries of change for the truth set and included small portions of another class increasing the error of the training set.

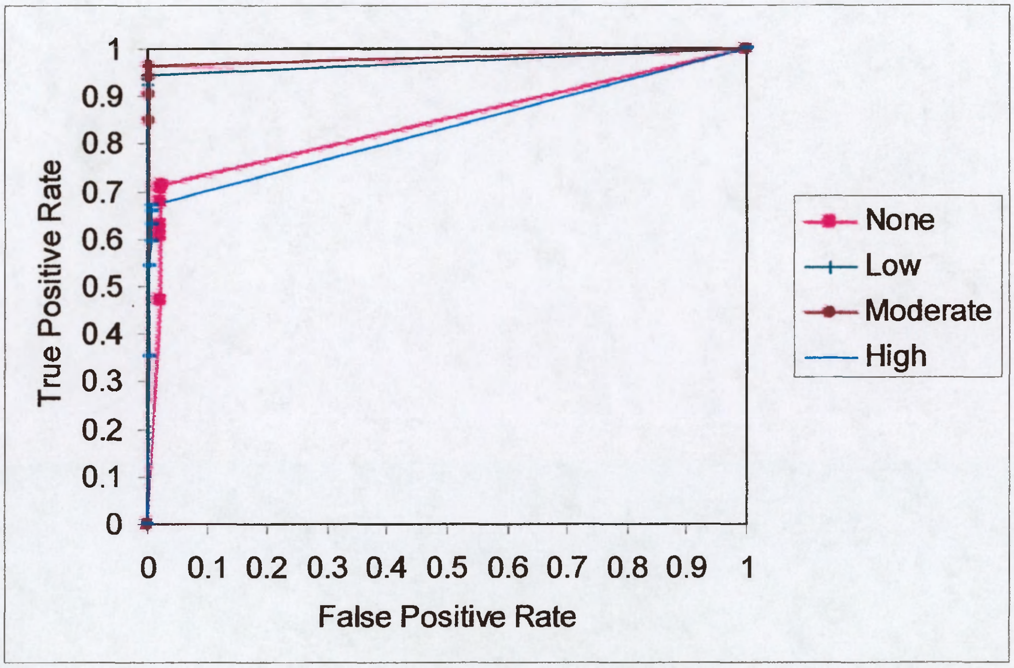


Figure 27: ROC curves for KNN Missoula land development experiment

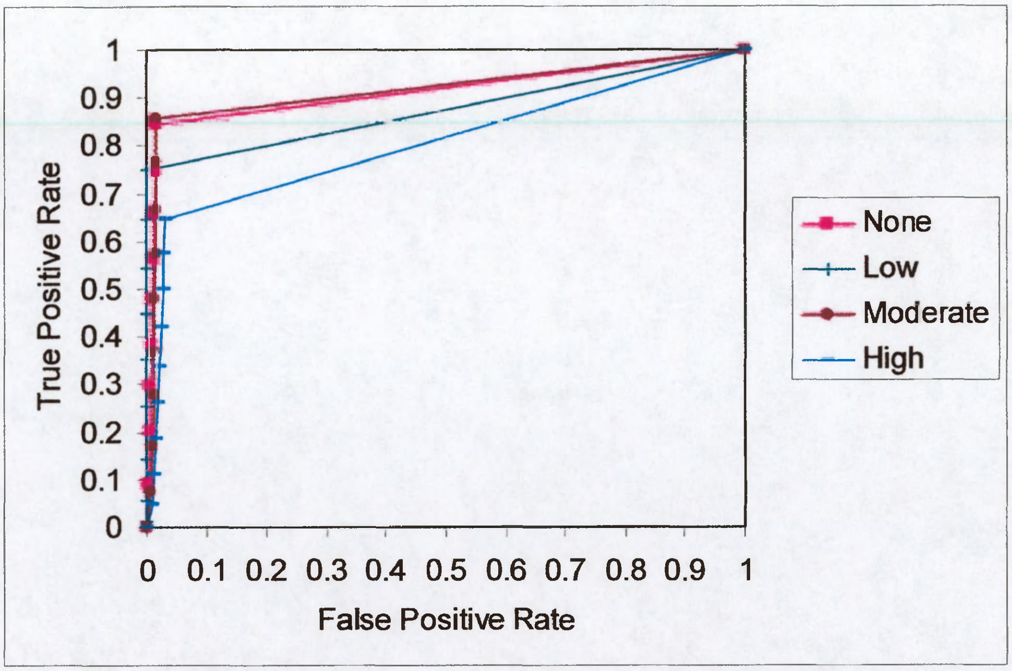


Figure 28: ROC curves for ANN Missoula land development experiment

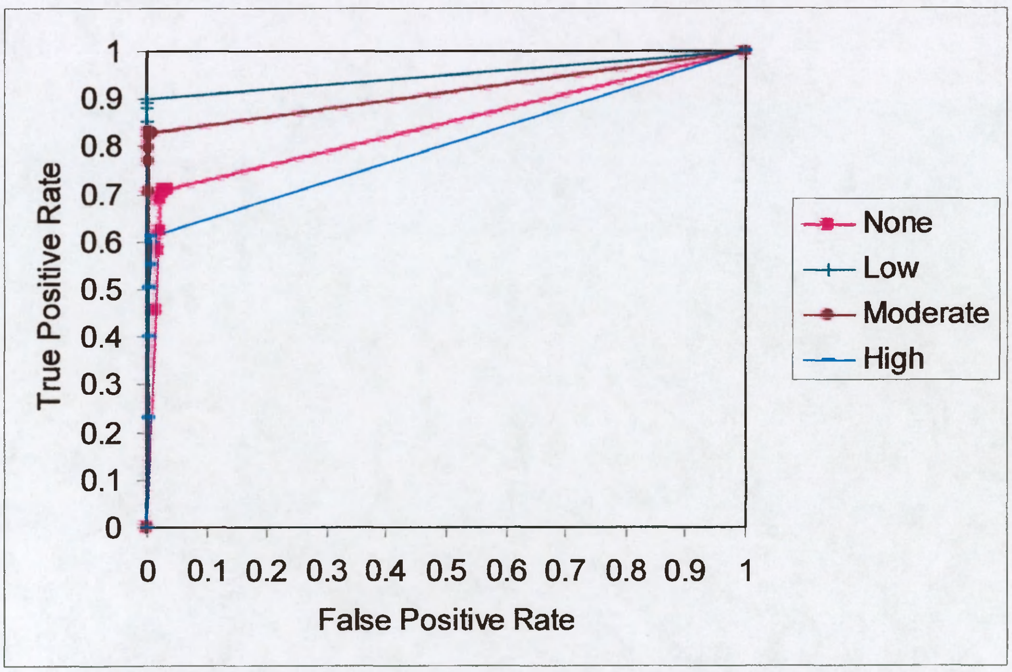


Figure 29: ROC curves for NB Missoula land development experiment

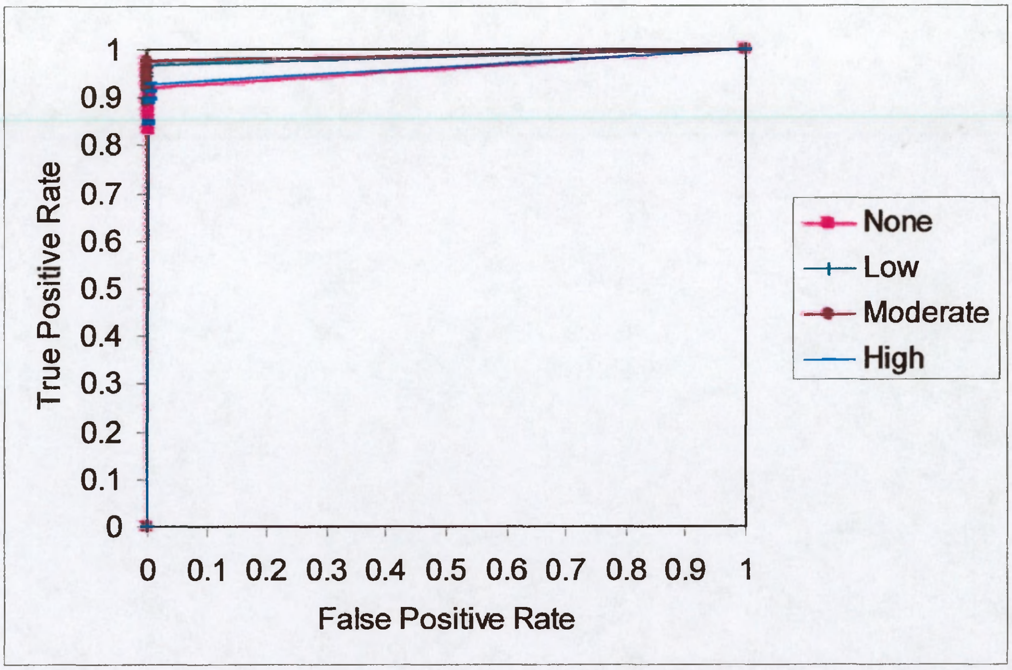


Figure 30: ROC curves for 9-learner ensemble Missoula land development experiment

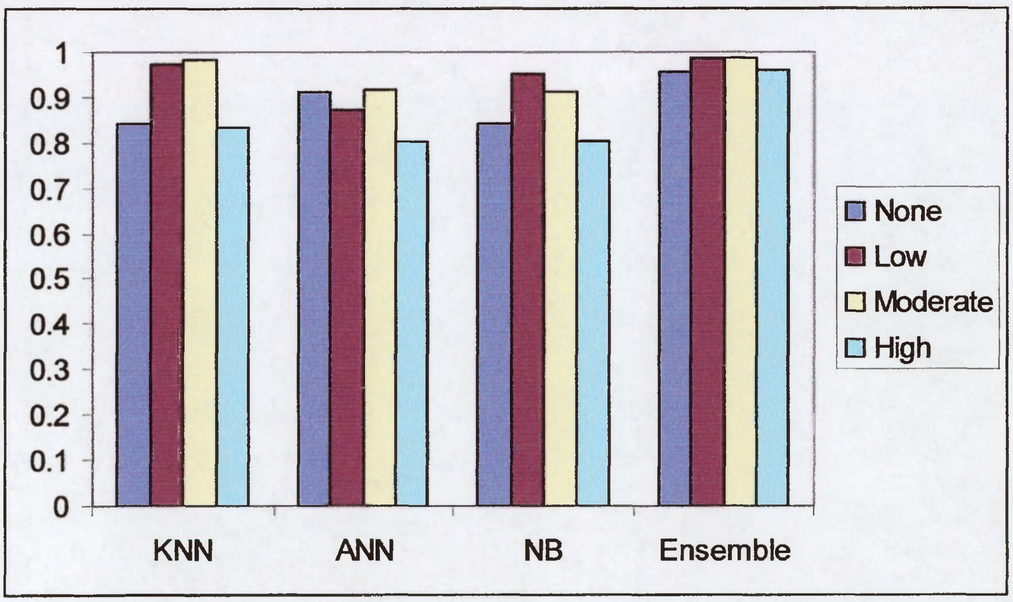


Figure 31: Approximate area under curve for Missoula land development experiment

5.4. Conclusions

Overall, the experiments illustrate the effectiveness of classifying features of change using classifiers and image processing techniques. The use of ensembles for combining multiple output classes and using multiple instances of the classifiers in combination with bagging shows additional utility in detecting the features of change. Additional bagging instances of each of the classifiers would have increased the percent correct and improved the area under the curve. The experiments also pointed out the inherent difficulty in defining a training set that represents the various degrees of change of the given feature especially in a mixed class. The learners were effective at building their own prediction of the separation and the bagging ensemble was able to merge them together to produce a better overall separation while generally increasing the prediction correctness of the individual classes.

6. FUTURE WORK

Through the study of this set of experiments further improvements could be made to the technique to improve classification. Well-known image processing techniques could be used to further enhance the images before the learning process. Further study concerning the relationship between feature complexities versus the learning parameters needed to successfully learn a concept is an important place to start. Class separation especially in terms of degrees of change needs to be investigated with respect to resolution and classifier bias. Ensembles also show great promise in increasing the effectiveness of feature extraction with detecting features of change. The techniques presented also beg further study of the application of discovering feature changes in dynamical systems using multiple temporal scenes over an extended period of time.

6.1. Integrating Image Processing

The application of image processing techniques has been studied for decades but the influence of these techniques on feature extraction with inductive learners has not. In particular with this set of experiments all images assume a two dimensional view of features which caused the landscape to hide features as well as shadows of those features to hide features based on the position of the light sources. Techniques have been development to remove shadows that could reduce some of the negative effects, but more elaborate models will need to be constructed to account for features dependent on landscape characteristics that hide parts of those features. Some improvements were found by biasing each band based on the most probable portions of the training set. This would be another potential area for exploration.

6.2. Feature Complexity vs. Learning Parameters

The Missoula land development experiment took a look at using the adaptive learning process to successively define changes in development. This process of adapting the learner so that the size of the input and learner parameters improves the classification accuracy of a feature by minimizing the complexity of the scene has many areas of study. The time of learning the feature classification can be greatly reduced for a large scene by compressing the image if the feature is sufficiently large. Further refinement could also be improved by using the abstract classification to focus a more complex learner to only the areas where the features were initially detected so that noise is eliminated and a learner can focus on the task to distinguish the desired feature from more similar disjunctive features.

6.3. Ensembles

The experiments presented in this study illustrated the utility of using ensembles to combine multiple classes together as well as using multiple sets of classifiers and bagging to improve feature extraction efforts of change detection. Other ensemble techniques such as Boosting (Opitz, 1999c) and Genetic Ensembles (Opitz, 1999b; Prabu, 2000) need to be investigated to study their utility to change detection. The construction of the ensembles in terms of classifiers, output classifications, resolutions, thresholds, data distributions can occur in many different ways and it would be useful to investigate various configurations of ensembles to find effective constructions for change detection.

6.4. Detection of Changing Features in Dynamical Systems

The experiments presented in this study focused on changes in features over a scene taken over an extended time period. Further investigation needs to look at combining learners that detect features with the learners that detect changes in features so that the learners can adapt to the newly evolved features and manage the evolution of the features into new classifications. Dynamical systems in the real world evolve over time and necessitate the need to keep track of the evolutionary characteristics of entities in the system. Machine vision has focused its efforts particularly in this pursuit and the adaptive feature extraction techniques with a system of keeping track of the features could provide a foundation for building a system to monitor features from a particular viewpoint in a dynamical system.

7. CONCLUSION

The goal of this thesis is to extend the existing feature extraction techniques to new areas and to improve the classification of features through the use of adaptive feature extraction techniques. The study presents the need for a technique to learn change characteristics in features of satellite images over time, the complexity of feature extraction in imagery, and a set of approaches to adapt the imagery and temporal features to the learner to improve the classification. The technique adapts the imagery to the learner, refines the classification, and successfully classifies temporal features in a variety of situations.

8. REFERENCES

Bain, W., (2000). MS Thesis, *A Comparison of Three Machine Learning Algorithms for Auotmated Feature Extraction from Digital Images*. Director: David W. Opitz, University of Montana.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.

Burl, M. et.al. (1998). Learning to Recognize Volcanoes on Venus, *Machine Learning Journal Special Issue on Applications and Knowledge Discovery.* 30:165-194.

Burl, M.C., U. Fayyad, P. Smyth, P. Perona and M.P. Burl (1994). Automating the hunt for volcanoes on Venus. In *Preceedings of the 1994 Conference on Computer Vision and Pattern Recognition.*

Ford, R., Ma, Z., Barsness, S. & Redmond, R. (1997). Rule-Based Aggregation for Region Identification. In *Preceedings of the 1997 American Society for Photogrammetry and Remote Sensing Annual Conference,* Seattle, WA.

Friedman, J. H., Bentley, J. L., Finkel, R. A. (1977). An Algorithm for Finding Best Matches in Logarithmic Expected Time, *ACM Transactions on Mathematical Software, Vol 3, No 3, September 1977.*

Gruen, A., E.P. Baltsavias, O. Henricsson Birkhäuser Verlag Basel, Boston, Berlin, (1997). Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) *Bi-annual Ascona conference, held in Ascona, Switzerland.*

Hansen, L. and Salamon, P., (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 993-1001.

Haykin, S., (1999). *Neural Networks: A Comprehensive Foundation*, 2nd ed, Prentice Hall, Inc.

Hernández, E. and Weiss, G., (1996). *A First Course on Wavelets*, CRC Press.

Hubbard, B., (1996). *The World According to Wavelets*, Wellesley, MA: A K Peters.

Jensen, J. R. and D. C. Cowen, (1997). "Remote Sensing of Urban/Suburban Socioeconomic Attributes," *Land Satellite Information in the Next Decade II: Sources and Applications,* Bethesda: American Society for Photogrammetry & Remote Sensing, (December 5); CD, 19 p.

Kramber, W. J., and A. Morse, (1994). "Integrating Image Interpretation and Unsupervised Classification Procedures." *ACSM/ASPRS Annual Convention & Exposition*. Baltimore: ACSM/ASPRS, 1994. 1:327-336.

Langley, P., (1996). *Elements of Machine Learning*, Morgan Kaufman Publishers, Inc.

Lillesand, T., and Keifer,R., (1999). *Remote Sensing and Image Interpretation*, 4th ed: John Wiley & Sons, Inc..

Lunetta, R., and Elvidge, C., (1999). *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, Taylor & Francis.

Maloof, M.A. (2002). On machine learning, ROC analysis, and statistical tests of significance. *Proceedings of the Sixteenth International Conference on Pattern Recognition*, 204-207, Los Alamitos, CA: IEEE Press.

Maloof, M. et al. (1998). Learning to detect rooftops in aerial images. In *Proceedings of the Image Understanding Workshop*, (pp. 835-845).

Mangrich, M., (2001). MS Thesis, *Hierarchical Feature Extraction: A Stepwise Approach to Image Classification*. Director: David W. Opitz, University of Montana.

Mather, P. (1999). *Computer Processing of Remotely-Sensed Images, An Introduction*, 2nd ed.

Mitchell, T. (1997). *Machine Learning*, Boston: MIT Press.

Opitz, D. (1999a). Feature Selection for Ensembles. *Sixteenth National Conference on Artificial Intelligence (AAAI)*, (379-384). Orlando, FL.

Opitz, D. and Shavlik, J. (1999b). A Genetic Algorithm Approach for Creating Neural Network Ensembles. Combining Artificial Neural Nets. *Amanda Sharkey (ed.)*. (pp. 79-97). *Springer-Verlag, London.*

Opitz, D. and Maclin, R. (1999c). Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research*, (pp. 169-198).

Opitz, D., W. Bain, M. Mangrich, J. Zeiler, and M. Sweet, (2000). *Rapid Extraction of Objects from Digital Images.* Department of Computer Science, University of Montana.

Palhang, M., A. Sowmya, (1997). Two Stage Learning, Two Stage Recognition, *Proceedings of Australian Joint Conference on Artificial Intelligence*, Perth Australia, 2-4 Dec. 1997, 1997.

Prabu, Ganesh, (2000). MS Thesis, Genetic Ensemble Feature Selection. Director: David W. Opitz, University of Montana.

Seul, M., L. O'Gorman, M. Sammon, (2000). *Practical Algorithms for Image Analysis: Description, Examples, and Code.*, Cambridge Univ Press.

Stollnitz, E., Derose, T., and Salesin, D., (1996). *Wavelets for Computer Graphics Theory and Applications*, Morgan Kaufmann Publishers, Inc.

Swets, J.A. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–1293.

Theodoridis, S., Koutroumbas, K., (1999). *Pattern Recognition*, Academic Press.

Watt, A. and Policarpo, F., (1997). *The Computer Image*, Addison-Wesley Publishing.

Williams, G., (1999). *Chaos Theory Tamed*, Joseph Henry Press/National Academy Pres.

Woodcock, C.E., and V.J. Harward, (1992). Nested-Hierarchical Scene Models and Image Segmentation, *International Journal of Remote Sensing*, 13(16):3167-3187.

Adaptive Feature Extraction:

Exploring the Search Space with Change Detection using Inductive Learners and Image Processing

By

Joseph D. Zeiler

BS, Computer Science, The University of Montana

Presented in partial fulfillment of the requirements for the
Degree of Master of Science
The University of Montana
2003

Approved by:

Chairman, Board of Examiners

Dean of the Graduate School

10-10-03

Date

Zeiler, Joseph D., M.S., October 2003 Computer Science

Adaptive Feature Extraction, Exploring the Search Space with Change Detection using Inductive Learners and Image Processing

Director: David W. Opitz DWO

Remotely sensed digital imagery provides snapshots of the earth at given time frames giving scientists information necessary to analyze changes to the earth. Due to the steadily increasing number of images available and the dynamic nature of the earth scientist need tools to help them identify the features that are changing as well as the nature of the changes. Current techniques for feature extraction suffer from being unable to classify complex features and similarly do not generalize well to other datasets with similar features. Machine learning techniques using inductive learners have shown great promise in identifying the desired features in information with minimized user interaction and have greatly reduced the time needed to correctly classify objects in imagery over traditional techniques. We present a technique for adapting existing feature extraction efforts using inductive learners and image processing techniques to improve the quality of identifying features in remotely sensed digital imagery that change over time. The technique compares snapshots of an area taken at different times creating a difference image which is incorporated into the learning process as additional information. Additionally, multiple features are incorporated into the learning task along with the addition of multiple sets of classifiers forming ensembles. The basic idea is that individual classifiers are used to learn each feature. The classifiers are successively combined to produce the predicted output. Multiple "bootstrapped" ensembles of these classifiers then predict the final output for the image classification improving the accuracy and generalization of the classification task. The adaptive feature extraction technique successfully adapts the imagery to the learner, refines the classification, and classifies temporal features in a variety of situations.

Table of Contents

List of Figures

1. INTRODUCTION

The universe is a complex dynamical system. Scientists attempt to understand the behavior of the universe by gathering information at varied intervals of time and analyze the information to determine the state of a given system, its history, and even to predict the future events of a system. Satellite technology provides a way by which to collect information. The process of extracting the information from space over time via satellite and converting the information from the various sensors into digital form is often referred to in literature as remotely sensed digital imagery. This information is analyzed to look at the state of the system at a given time as well as to look at the changes in the state of the system over time. Analysis frequently looks for specific features that are of interest in information. Feature Extraction is an area of research that studies the process of identifying features of interest in information. *This paper presents a technique for adapting existing feature extraction efforts using inductive learners and image processing techniques to improve the quality of identifying features in remotely sensed digital imagery that change over time.*

Current machine learning techniques using inductive learners have shown great promise in identifying the desired features in information with minimized user interaction and have greatly reduced the time needed to correctly classify objects in imagery over traditional techniques (Bain, 2000; Burl et al., 1998; Maloof et al., 1998; Mangrich, 2001). Currently, there is need to improve the current feature extraction techniques that use inductive learners since there are many complex features that are difficult to identify with a single model. Current techniques also suffer from not being able to generalize well to other data sets with similar features.

There is potential promise in the use of incorporating multiple models by using multiple images of the same area of space taken over time. Changes are of primary interest to scientists since they provide information necessary to understand complex dynamical systems. Current research has worked with change detection and has found the problem to be very difficult due to the different sampling parameters of the various sensors in the satellites as well as environmental conditions altering the sampling values. Even though a direct comparison is difficult, change detection has shown potential promise in incorporating the additional data as another model for the inductive learners to use by finding change amongst images and incorporating it into the learning task.

Particularly, this process will help to find features that change dramatically from image-to-image and focus the learner on these features. For example, scientists are often interested in changes that occur after extreme natural phenomena such as earthquakes or fires. Scientists need to be able to ascertain damage after such an event and could use before and after satellite images to detect change and then learn the various levels of change amongst the damage. This could greatly benefit planners working to assess clean-up efforts and prioritize damaged areas.

Increasing the generalization ability of an inductive learner is also a difficult task. Currently various additive models have been used to increase the complexity of the learner until a certain degree of generalization is obtained. Other techniques start with a complex learner and subtract elements of the model until an acceptable level of generalization is reached. The other common approach is to modify the search space of the dataset so that it is more compatible with the learner. Ultimately all of the techniques work to increase the level of generalization while maintaining as much accuracy as possible for the given task at hand.

Together, change detection and image processing techniques incorporated into feature extraction contain the potential to increase the ability of the learner to find more complex features. Incorporating change detection is an important addition to the feature extraction process and will provide analysts the techniques needed to detect change in features and allow them to adapt past feature extraction efforts to new images. This is an important step that will begin to provide the foundation to build predictors for future events. Adaptation of the learning process to other images through image compression techniques will provide a means to increase the generalization ability of the learners to classify other images.

2. BACKGROUND

Earth is a complex dynamical system within the universe that is of primary interest to scientists. Satellites extract information about the earth giving a picture of the state of a given area of space based on the sensors of the device creating an image of that space at a given time at a defined resolution. Scientists interpret images to recognize important features that will help them analyze the region of space the images cover. Satellites sample information over varied intervals of time and hence create many images. Image interpretation can quickly become difficult for scientists since thousands of images can be taken of a given area of space in minutes. This makes traditional feature extraction efforts overly time consuming for scientists. Thus computers are often used as a tool to facilitate the interpretation of images.

2.1. Computer Images

A computer image is a numerical representation of an image (or picture). The computer image is represented in two-dimensions consisting of a set of l bands each of which consists of a $m \times n$ dimension array where m represents the rows and n represents the columns of the image, respectively. Each element of the band, or pixel, contains z bits used to represent the intensity of a specified area of a band in the image. (Watt, 1999)

Images created using satellites contains sensors that sample the given space at a particular time for each sensor producing a finite representation of the space. The information is limited by the sensors' capabilities to extract information from the given area of space and further limited by the precisions of the computer used to store the information (bit depth of each pixel and the number of pixels used to represent that area of space). The process of extracting information with the sensors is also further limited by the state of the system at the time of sampling due to interference from environmental conditions that prevent optimal sampling.

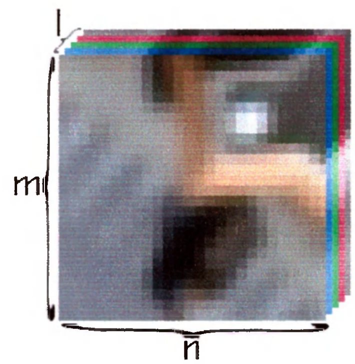


Figure 1: Sample Digital Image

2.2. Image Processing

The goal of image processing is to enhance an image in some way so that it is more easily interpretable. This often means producing an image that is more easily viewable or to reduce noise in an image. Image processing techniques are operations that transform an image either spectrally or spatially in the spatial domain or transform an image based on the spatial frequency in the Fourier domain. Wavelet transforms simultaneously transform the spatial and frequency domains. Image transformations fall into two broad categories global image transformations and local image transformations.

Global image transformations look at the global characteristics of the image. These transformations include intensity, image combining, geometric, and color transformations. Image intensity transformations deal with the histogram of the color intensities of an image. Common operations on the histogram alter the visible contrast of an intensity range about an object of interest. Image combining transformations select bands from images and compare them. Often logical operations such as AND, OR, and XOR are used to perform these operations. Geometric image transformations such as scaling, rotation, and shearing are common operations performed on images. Color image transformations transform the color bands of an image into alternate representations such as hue, intensity, and saturation components versus tradition red, green and blue components.

Local image transformations look at local characteristics of an image and alter them in a local context. Most local image transformations evolve around spatial domain convolution filters. The two most common filters are low-pass filters (smoothing) and high-pass filters (edge enhancement). The operations work in the spatial context by computing values using neighboring pixels to compute the new pixel values based on spatial context.

Convolution filters work on each pixel or a group of pixels of an image. A convolution kernel **H** represented as a matrix containing the coefficients for each pixel in the neighborhood of a defined pixel is used to transform the given point, P, of an image at that pixel. The basic process starts by choosing the point, P, to apply the convolution filter. Multiplication with each convolution kernel coefficient and the respective pixel in the image then occurs. Each product is then summarized, normalized, and placed in the output image at P. This can be represented by $P = \sum_{i,j} I[i,j] * H[i,j]$ as illustrated in *Figure 2* below (Seul, 2000).

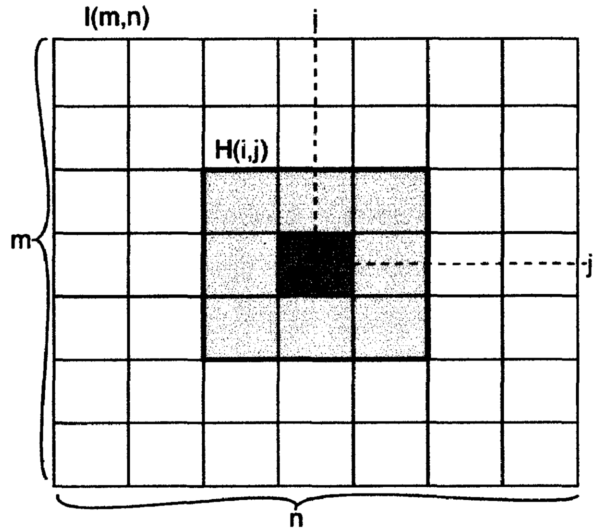


Figure 2: Sample Convolution Filter

2.3. Feature Extraction

Feature extraction is the process of identifying features of interest in a dataset. In remotely sensed digital imagery, satellites take snapshots of a given area of space at a given time producing a representation of that area in digital form at a specific resolution. Feature extraction for digital images looks at the data and identifies areas of an image that represent the specified feature of interest. The term image classification is used to refer to the process of classifying image features into classes or themes. In image classification, algorithms known as image classifiers analyze images and produce classes of image features, which are often represented pictorially as thematic maps.

Extracting information from remotely sensed digital imagery can be a difficult task. Feature extraction is not only limited by image resolution and physical sampling but also the features themselves can be limiting. Features often relate spectrally and spatially to other objects in an image. The term used to refer to the degrees of similarity or separation of objects is disjunctive. For instance, given an aerial view of a city, one can easily identify features like buildings, trees, roads, etc. with similar color and shapes.

It is often difficult to distinguish between different types of similar features since their spectral signatures are similar. If the task at hand were to find all fir trees in an image, this task would be quite difficult if many different types of evergreens also existed in the image. Other information would be needed to help identify the fir trees in the image. This is also true with shapes of similar features.

Most feature extraction efforts gather as many bands as possible incorporating different sensor ranges with as much resolution as possible. This information is then analyzed and processed to select the information that will help to classify the desired features. The image classifiers attempt to exploit spectral differences in the various bands as well as spatial similarities. Even with all of this information, features can be complex and their representation is often nonlinear in the feature space making detection difficult. Thus there has always been a need for improved techniques.

2.4. Current Techniques

Originally, feature extraction for images evolved from data processing techniques already used to analyze data. These techniques evolved with the advent of the computer helping scientists create techniques to handle larger amounts of data producing more complex techniques. Digital images began to emerge and refined the techniques for computer images creating the field of image processing. The evolution of feature extraction started in the late 1950s and continues to the present.

Feature extraction techniques are heavily rooted in artificial intelligence. A branch of artificial intelligence known as machine vision began by exploring techniques to reproduce human vision for robotics. These algorithms started analyzing just spectral information and quickly found need for spatial context and shape.

Many existing image classification techniques currently exist. The two primary ways of classification are supervised and unsupervised. Supervised techniques require user interaction where a user iteratively refines the features in the image until an acceptable level of classification for the image is reached. Unsupervised techniques attempt to automate the entire process. This thesis only looks at supervised techniques that adapt images to the existing feature space to augment feature extraction.

Classic feature extraction techniques include template matching, adaptable pattern recognizers, statistical classifiers, and trainable pattern recognizers. The earliest form of pattern matching became know as template matching. Template matching consists of storing a prototype or template of features for each class and then checking each image for the features based on their similarities to the prototypes. This early method suffered from being able to only correctly classify well-defined features with little change and relied on just image intensity information.

Adaptable pattern recognizers attempt to reduce some of the limitations imposed by the early template matching schemes by making the prototypes deformable. The techniques allowed variations in the prototypes to exist, which allowed for more robust classification. For instance, if an aerial image was taken of a city, and another image was taken of this same area at a different angle, the classic template matching technique would not produce a good match since the templates would differ greatly. These techniques often take an existing template and rotate it at various angles to see if a given image would match the template more closely and perform other transformations. Altering the image intensities to account for different daylight conditions has also extended this technique. In addition to intensify information, spatial context became an important component of this technique. Machine learning researchers have built genetic algorithms that perform these transformations on the templates to perform feature extraction

Statistical classifier feature extraction techniques look at features of an image and attempt to build a representative set of image elements for each classification represented as a multivariate probability distribution function. Each new pattern could then be run through the functions to determine the class of the given feature. Decision tree algorithms, K-Nearest Neighbor and Naïve Bayesian algorithms are current machine learning techniques that have evolved from statistical classifier techniques for image classification.

Trainable pattern classifier feature extraction techniques are inductive learners that accept patterns of different classes. These patterns are given to the algorithm and the algorithm attempts to build a predictor that when given new patterns it will classify them according to its experience gained in training.

2.5. Machine Learning

A branch of artificial intelligence known as machine learning has emerged as an effective discipline for feature extraction. This field evolved from the realm of mathematics particularly in statistics, computation complexity, and information theory as well as from many fields of science such as biology, chemistry, and physics (Mitchell, 1997). Many types of algorithms have evolved from this heavily researched field providing powerful tools to analyze information (Mitchell, 1997). This thesis focuses on using inductive learners for feature extraction.

The general form of an inductive learner accepts a set of examples depicting the features used for training from the hypothesis space. The hypothesis space for a learner is the set of all possible hypotheses (predictions). Each example consists of input and/or output values. The learner analyzes the training examples and builds a target function that models the training examples. The target function accepts an example, or set of examples, as input and produces an output based on its construction inferred from the existing knowledge gained by the training examples. In other words, the learner builds a model from the given set of inputs mapped to the given set of outputs and when given an unseen instance can infer the output. The inductive learning hypothesis states that given a sufficiently large set of training examples that map well to the target function of the hypothesis space the learner will approximate the target function well over unseen instances (Mitchell, 1997).

2.5.1. Problem Representation

For feature extraction of images using inductive learners, inputs are constructed from the images spectral intensity values from selected bands of the image. Given an image, examples can be constructed by selecting a point in an image and retrieving the intensity value of each pixel at that location from the image and assigning the set of input values an output value according to the assigned classification for that pixel. In a simple two-class image classification problem, the sample could be assigned as true or false with respect to the given class. This would form a training example. More elaborate training examples can be constructed by taking a subimage (window) from the existing image. This technique also accounts for spatial information by including neighboring pixels as part of a training example. Spatial context provides important information about the features shape and context. Together, spectral pattern recognition and spatial pattern recognition create the structure of the examples for the given learner. Image preprocessing is often performed on the image before (while) gathering test examples. These processing techniques exploit characteristics of the image that will facilitate the learner in realizing an ideal target function.

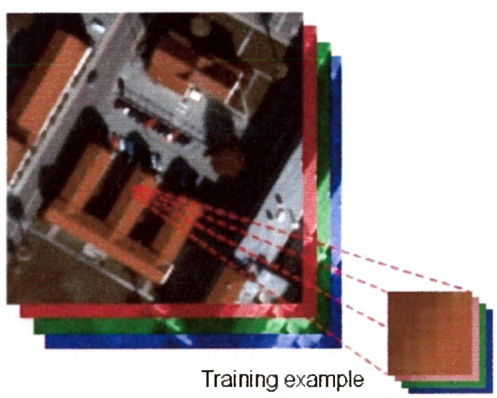


Figure 3: Sample 5x5 roof top training example

2.5.2. Artificial Neural Network

The artificial neural network (ANN) works well as a trainable pattern recognizer. The ANN is capable of learning pattern classes that are not linearly separable which is true of most pattern recognition classes in the real world. They are also robust to noisy training data and converge quickly to the target function. The ANN is a network of interconnected nodes. Each node contains a weight. Training examples are input into the network and produce an output classification. Back-propagation is the type of ANN used in this thesis. Back-propagation alters the weights by comparing the output of training examples with the output of the network at a given time and produces a vector of error terms that are then propagated back into the network to alter the weight in such a way that when the examples are sent into the network again the error term will be smaller. As the examples continue to evolve the network, the error term of the network approaches zero. The error term going to zero is limited by the size and structure of the network as well as the quality of the data. Examples are continuously fed into the network until a fixed number of iterations (epochs) or until convergence criteria is met. (Haykin, 1999)

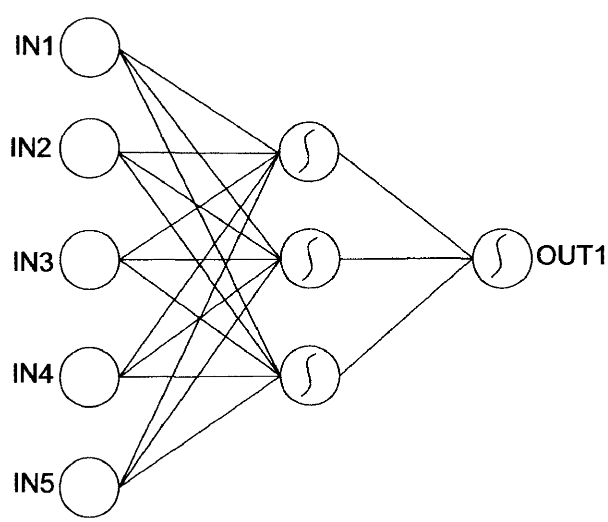


Figure 4: A neural network with 3 layers, 1 hidden layer

2.5.3. K-Nearest Neighbor

The k-nearest neighbor (KNN) also works well as a pattern recognizer. The KNN is an instanced-based learning algorithm. The algorithm accepts an unseen instance and compares it to the existing set of training examples using a distance metric such as the Euclidean distance. The k-nearest training examples is then used to classify the unseen instance as the most common classification. Since KNN does not build an explicit model each new unseen instance must be compared to the set of training examples available. For large sets of training examples, this can be very slow. KNN works well with small sets of training examples and has been found to work well on pruned sets of training examples using a hierarchical learning process where a neural network refines the initial features and then the KNN algorithm can be used to refine the results (Mangrich, 2001). KNN is robust to noisy training data and also is able to handle classes that are not linearly separable (Friedman et al, 1977; Mitchell, 1997).

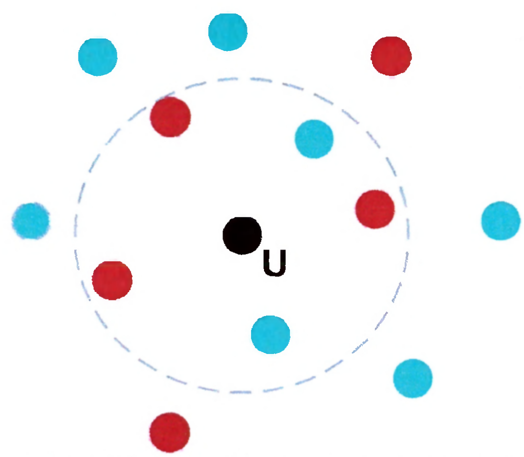


Figure 5: A 2-dimensional instance of k-nearest neighbor, k=5, where U is assigned the classification of red

2.5.4. Naïve Bayes

The Naïve Bayes (NB) classifier is the third type of pattern classifier. The NB classifier is a probabilistic learning algorithm. The NB classifier calculates probabilities of a set of classifications using a simplified version of Bayes theorem that naively assumes that attributes are probabilistically independent. The classifier computes the probabilities of each attribute value/class conjunction and multiplies them with the prior probabilities of the attribute values to produce an estimation of the entire set of attributes. The learner infers the most probable outcome using this simplified set of rules to classify unseen instances. This set of assumptions leads to a computationally efficient algorithm that can quite effectively classify a variety of problems especially where the attributes are independent. Even though this is an unrealistic assumption, the algorithm still produces acceptable results on a wide variety of problems including digital imagery. (Theodoridis, 1999)

2.5.5. Ensemble

An ensemble is an extension to traditional learning techniques. An ensemble combines multiple learners known as predictors. Each predictor is trained using a set of training examples. The predictors in the ensemble are each fed an unseen instance to produce predicted output. The output for each predictor is then compared and analyzed to determine the final output of the ensemble. To produce an effective ensemble, each predictor needs to produce as much disagreement as possible. This can be accomplished by varying the types of predictors or learners, varying the distribution of training examples, and/or modifying the training examples. This in effect produces bias so as to favor different properties in the learner. If the desired feature to learn was buildings, learners could be constructed that favor each roof type. These learners could then effectively find each roof type and then use a weighted average favoring the specialized learner to specify buildings. (Opitz, 1999a)

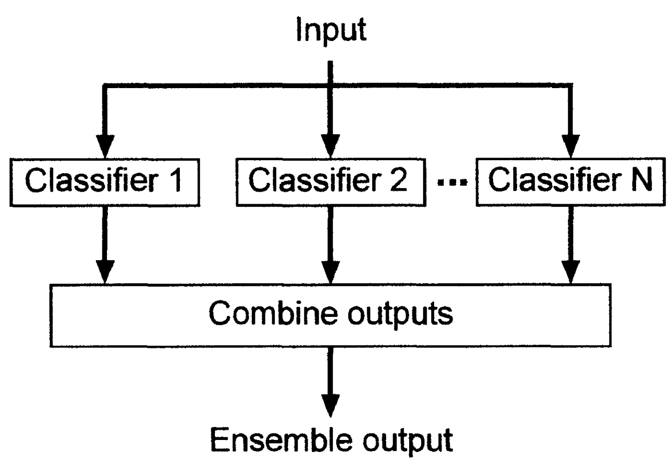


Figure 6: Sample ensemble

Bagging (Breiman, 1996) is the ensemble technique used in this thesis where each of the *N* classifiers is trained on a random resample (with replacement) of the original training set and then combined to produce a prediction of the output. Bagging has often been shown to reduce errors in the overall classification task better than a single classifier since the distribution of the resampled training sets produces sufficient diversity among the classifiers to reduce the impact of increased error rates among the individual classifiers. (Opitz, 1999c)

2.6. Change Detection

Change detection is an important concept for future feature extraction efforts. Dynamical systems and chaos theory are integrally connected with change detection. Computer imaging especially video compression algorithms exploit changes in a scene and perform difference operations to find the changes in images. This operation exploits the fact that there is usually little change from one image to the next. The differences can be stored instead of each image and hence reduce the size of the video greatly. Similarly, difference operations can illustrate the changes of images taken over time and represent additional knowledge for a learner to learn a disjunctive concept. These difference operations compare bands taken over periods of time. Difference operations can be taken from each set of images to produce a difference image that represents a phase change between images. Successive phase change images can then be used to illustrate changes between each phase. This knowledge can greatly help a learner find disjunctive concepts that change over time. (Lunetta, 1999)

2.6.1. History

Change detection for satellite images first analysis efforts attempted to detect changes in the luminosity and position of stars from Landsat MSS images taken at different time periods. The technique was further refined to visually analyze the various locations of space by filming the changes in position of the stars using color addition/subtraction on the changes. This process evolved from a purely visual interpretation process to an analytical process. Two primary change detection analysis methods emerged: post-classification and pre-classification methods.

2.6.2. Post-Classification Change Detection

Post-classification techniques perform categorization on each dataset independently and then analyze the differences between the classifications. Classification categorization techniques do not require any data normalization between multitemporal datasets, but do often require additional time to separately classify each set. Classification can also be different between the multitemporal data due to changes in the satellites making it difficult to compare the separate classifications. The separate classifications of each image when combined also propagate errors due to the errors inherent in the separate classifications.

2.6.3. Pre-Classification Change Detection

Pre-classification techniques work with multitemporal data directly performing the necessary analysis and transformations to the data and then produce classifications based on the multitemporal data. Five basic pre-classification methods currently exist: composite analysis, image differencing, principle component analysis, change vector analysis, and spectral mixing analysis.

Composite analysis performs a single analysis of a multitemporal dataset where the dataset is collected under similar conditions from different years. Composite analysis uses standard pattern recognition and spectral classification.

Image differencing performs a difference operation between two multitemporal datasets producing a map of the degrees of change between two time periods. Image differencing is often accompanied by a pre-processing step that normalizes two images to correct different sampling parameters.

Principle component analysis is a data redundancy technique that explores multitemporal data and locates the principle components in the data that accounts for the greatest amount of variance in the data. Subsequently, additional components in the data are accounted for to provide additional information to help with classification.

Change vector analysis looks at the changes in multitemporal data as it changes over time storing the changes from one time step to another in change vectors. The lengths of the change vectors are analyzed to determine the magnitude of change and the direction of the given vector that indicate the nature of change.

Spectral mixing analysis looks at high spectral resolution multitemporal data combining multiple spectral signatures of the data that contribute to the overall reflectance of the image that can be analyzed to classify subtle land cover changes.

This thesis focuses on image differencing using satellite imagery taken over extended periods of time with the primary intent to extract changes that have occurred between the time periods.

3. ADAPTIVE FEATURE EXTRACTION

Satellites have been creating images of the earth for many years. Millions of images have been taken of different locations at different times with varying environmental conditions using widely varying equipment and techniques. This imagery has been used to analyze the earth to help scientists to understand the evolving environmental conditions of earth. Due to the dynamics of the earth, previous feature extraction efforts quickly become outdated as new imagery becomes available. New technologies also provide higher resolution imagery with additional sensors providing the capabilities to analyze the image much more thoroughly than was possible in the past. Thus there is great need to refine techniques that will allow existing feature extraction efforts to facilitate enhanced feature extraction efforts.

This thesis presents an adaptive feature extraction approach that uses change detection techniques to incorporate changes into the learning process to iteratively refine the feature extraction process. Additionally, image compression is used to simplify the learning process and to adapt existing feature extraction efforts at lower resolutions to higher resolution images. This in effect improves the classification of the images at the higher resolution by refining the coarser approximations made at lower resolutions.

Image processing is the first component of the adaptive feature extraction approach. The images are first projected into the same coordinate system and referenced so that each point in space of the image coincides to the same point of space in the other image. The images are then normalized to minimize differences in sampling parameters. Any additional image preprocessing techniques are then used to enhance the bias of the features being extracted for the learner from the images.

Change detection is the second component of the adaptive feature extraction approach. Change detection first analyzes two images taken at different periods of time and produces a phase change image, or difference image, that represents the changes over time. These changes direct the learner to features that change at various levels. Typically, not all of the features in an image change at the same rate. For instance, in a small city that has had recent development. The buildings will contain large differences where new development occurs and help a learner learn the new development. The resulting phase change image can then be incorporated into the learning process as additional information for the learner.

Image resolutions continue to increase as satellite technology evolves providing much more information. Adapting existing feature extraction efforts requires adjustments to existing images and feature extraction efforts. Image compression allows images to be compressed to fit the existing images. Learners can then transfer existing knowledge gained from previous efforts to the new images. The new images can then be uncompressed progressively along with the new features extracted from previous efforts.

Multiple predictors will be needed in each phase to classify each feature. Separate learners are used to classify each feature. The learners will then combine their output forming an ensemble to determine the final image classifications for the various classes. Ensembles will be used at each phase to classify the individual features. Bagging will also be used on combinations of the classifiers to produce a unique classification for each pixel in the image. Similarly, multiple ensembles can also be combined to produce a unique classification for each pixel in the image.

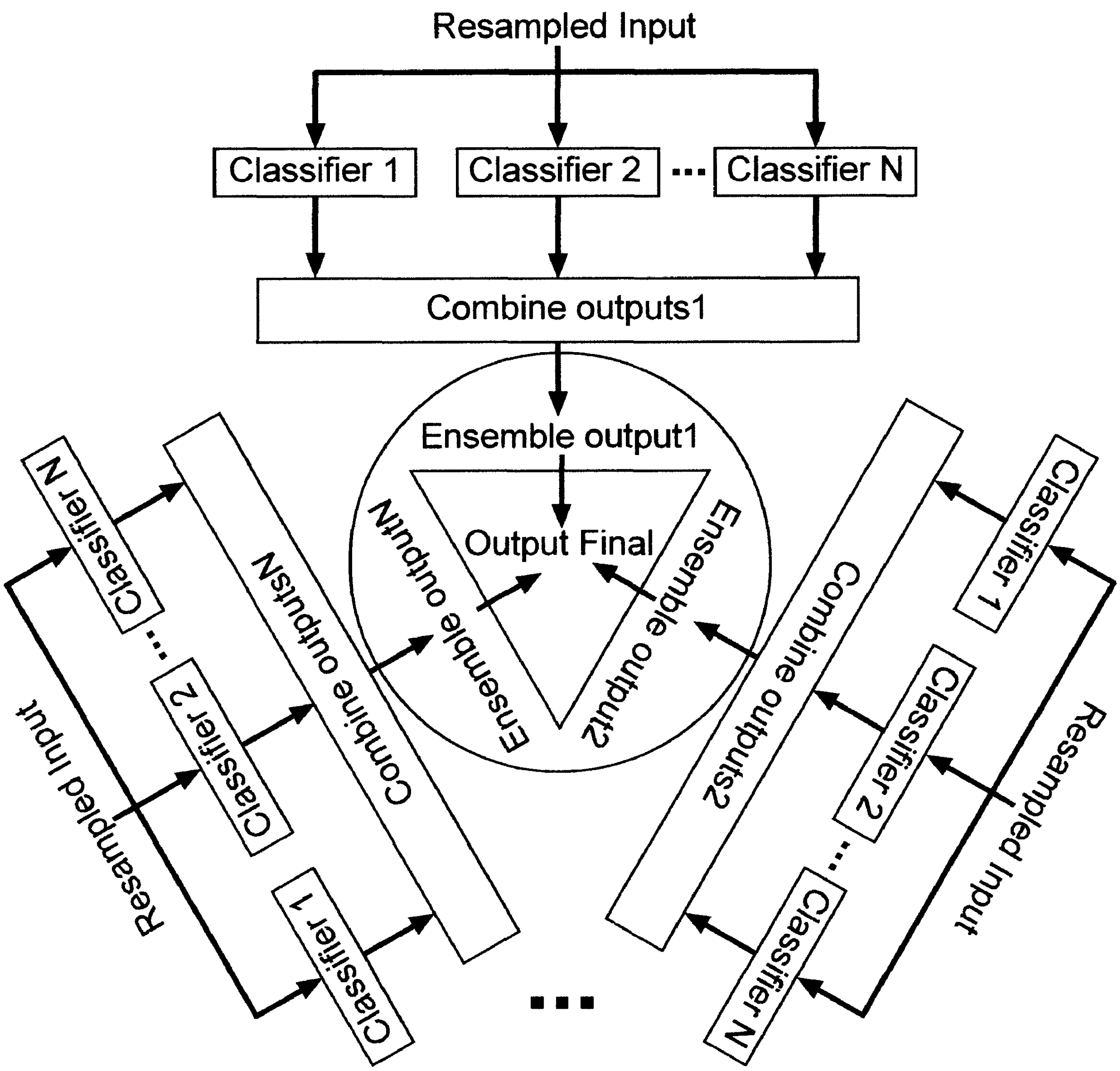


Figure 7: Collection of Ensembles Combining Predictions

4. METHODOLOGY

Three study areas are presented for the adaptive feature extraction techniques: (1) and (2) classifying the severity of wildfire to vegetation, and (3) classifying built-up area in two urban scenes at low resolution. The experiments illustrate learning to classify changes in features using temporal satellite imagery and the utility of adapting the images to the learners to improve the classification.

4.1. General Methodology

This section describes the general parameters for all three experiments performed in this thesis. The 30m LandSat images for each experiment were geo-rectified to previously terrain-corrected images using Erdas Imagine software and then clipped to the scene of interest. The before-and-after images were corrected for atmospheric scattering using histogram equalization (Seul, 2000). Difference images were constructed from these corrected images by taking the absolute difference of each band from the before-and-after images.

Training examples were selected by visually interpreting the image. Existing lifeform (tree, grass, shrub, etc.) layers and expert analysts assisted in the interpretation of the images to come up with a set of training examples and a truth image. Examples were constructed by defining point and polygonal representations (shapes) of the features using ESRI ArcMap GIS software and were then converted to Boolean mask bands for each feature. Inputs of the selected examples consisted of each band from the after satellite image and optionally each band of the difference image. The inputs along with the feature classification represented the training example for each feature.

The ANN classifier settings consisted of a learning rate of 0.1, momentum of 0.9, and a single hidden layer topology with a real-valued output between 0 and 1 (Mitchell, 1997). The KNN classifier used 5 neighbors with inverse squared real-valued distance weighting. The NB classifier computed outputs based on the real valued inputs merged into 100 values. The outputs from the inductive classifiers are real valued numbers normalized between 0 and 1. Percent error was computed by comparing the predicted output image to the truth image for each feature of interest.

For each individual classifier type (ANN, KNN, NB) experiment the final classification of a given pixel was determined by individually training an instance of the classifier type for each feature of interest. Examples were constructed such that the given feature examples were positive and all other features of interest were negative for the classifier designated to learn the given feature. Each learner was trained on the assigned feature and all outputs of the classifier were compared. The dominant learner (learner with the highest predicted output) then classified the pixel with its output type. This was done for each pixel in the scene and represents a simple ensemble classifier.

The ensembles experiments each used 9 classifiers composed of 3 ANN, 3 KNN, and 3 NB ensemble classifiers arranged as depicted in Figure 7. Each ensemble classifier's input consisted of random samples with replacement equal to the size of the original training set. Each classifier type formed another ensemble where the majority output classification for each of the ensemble classifiers represented the vote for the 3 learner ensemble. The outputs of each of the classifier types were then compared and a dominant type was selected. For the final output from the ensemble each classifier type ensemble (3 ANN, 3 KNN, 3 NB) voted on the predicted output type and the majority output type was selected.

4.1.1. Receiver Operating Characteristic Curves

A receiver operating characteristic (ROC) curve is an effective representation of the quality of the discriminatory power of a classification algorithm when looking at the positive and negative results of the learning algorithm on a given set of data under various evaluations of the output of the classifier. To construct the ROC curve the real valued output is tested at different thresholds and then plotted with the true positives on the y-axis and the false positives on the x-axis. The area under the ROC curve is approximated using the trapezoid rule on the true/false positive points and helps determine the effectiveness of the leaner at learning the given task and will be used to compare the effectiveness of the different learners to one another. ROC analysis removes the inductive bias of each of the learners and costs of unknown and unequal classification error. (Maloof, 2002)

		Truth: Yes	Truth: No
Prediction	Yes	True Positive	False Positive
	No	False Negative	True Negative

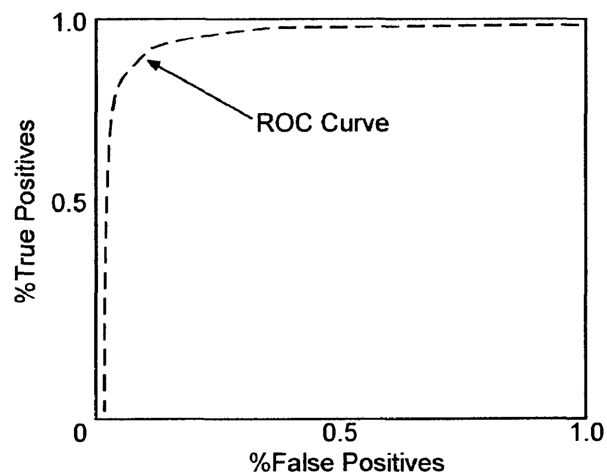


Figure 8: A sample confusion matrix and ROC curve

4.2. Ashland Fire Experiment

This experiment presents a scenario that explores the utility of including a phase change image as additional input for the learner. Given a set of Landsat-7 TM images of an area near Ashland Montana now known as the Fort Howes fire complex due to a 15,000-acre fire that burned in late July 2000 classify the fire burn characteristics. The images are snapshots of the area before the fire and after the fire. The area consists of two different Landsat-7 TM scenes, Path 35/Row 28 and Path 35/Row 29.

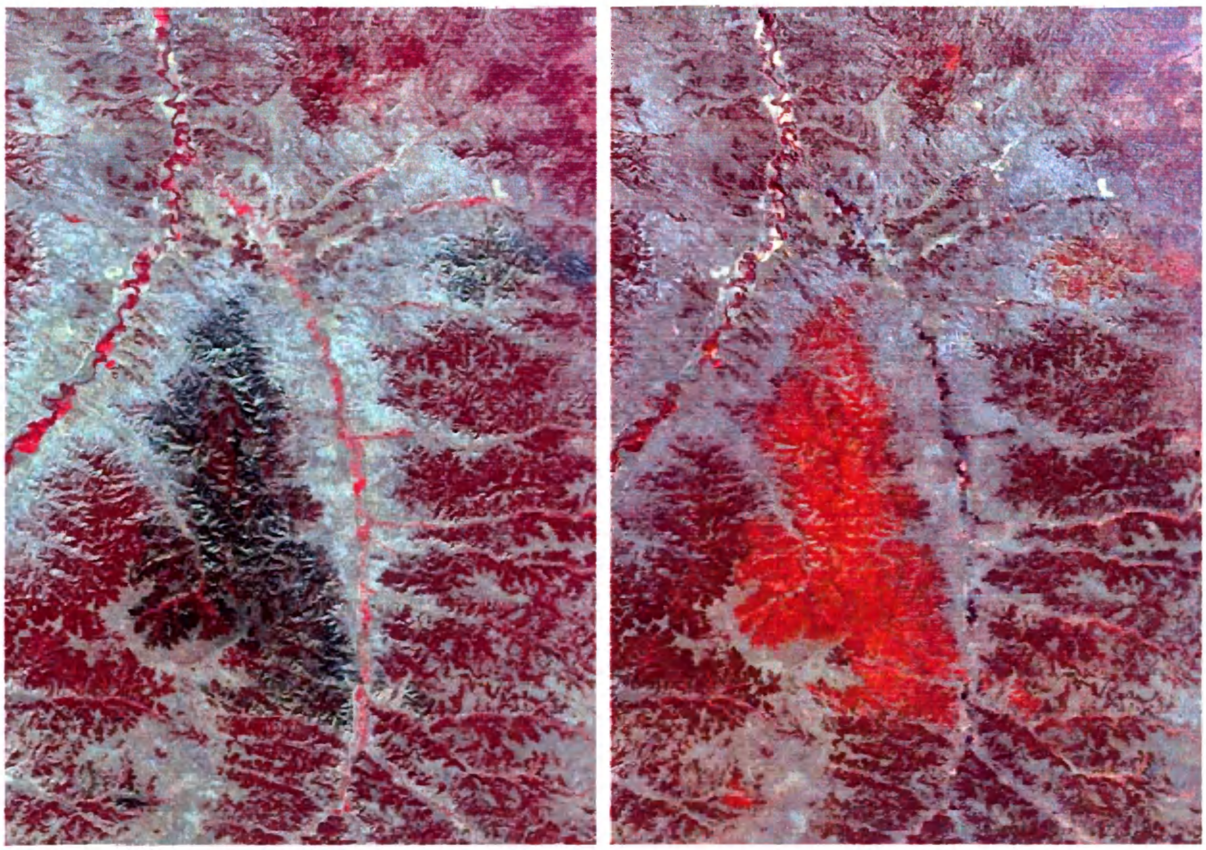


Figure 9: Ashland after fire image and difference image using bands 4, 3 and 2 as R, G and B

4.2.1. Task

The task is to classify the after fire image into eight classes:

1. Unburned Tree
2. Mixed Burn Tree (mosaic burn)
3. Burned Tree
4. Burned Grassland
5. Unburned Grassland
6. Burned Shrub Land
7. Unburned Shrub Land
8. Barren

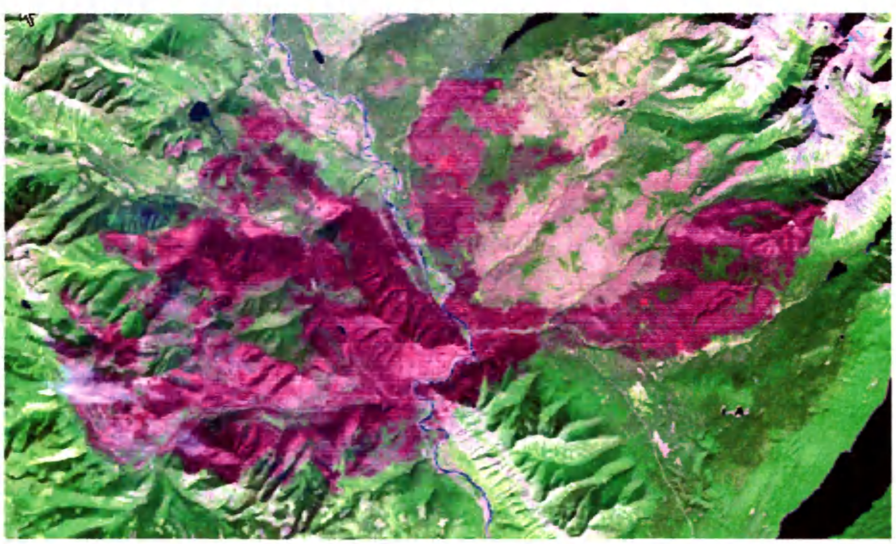
This training set consists of a set of examples (approximately 1,000 pixels from each layer for each class) selected by an analyst who derived them using air photo interpretation and an existing life form layer created before the fire.

4.2.2. Specific Methodology

Two experiments were conducted for this image set. The first experiment looked at only the after fire image. The second experiment compared the two images and produced a phase change image. This phase change image was then added as additional bands to the after fire image. The features were relatively small so inputs consisted of subimages of size 3 pixels by 3 pixels from each layer. Each classifier was trained using the standard bands and the standard bands with the phase change image. The results of the learner's classifications were combined as an ensemble and then classified each of the pixels into one of the eight classes giving the classification to the learner with the highest prediction.

4.3. Moose Fire Experiment

This experiment presents a scenario that also explores the utility of including a phase change image as additional input for the learner over the existing technique of principle component analysis. Given a set of Landsat-7 TM images of an area near Columbia Falls Montana now known as the Moose fire due to a 71,000-acre fire that burned in summer of 2001 classify the fire burn characteristics. The images are snapshots of the area before the fire and after the fire. The area consists of two different Landsat-7 TM scenes.



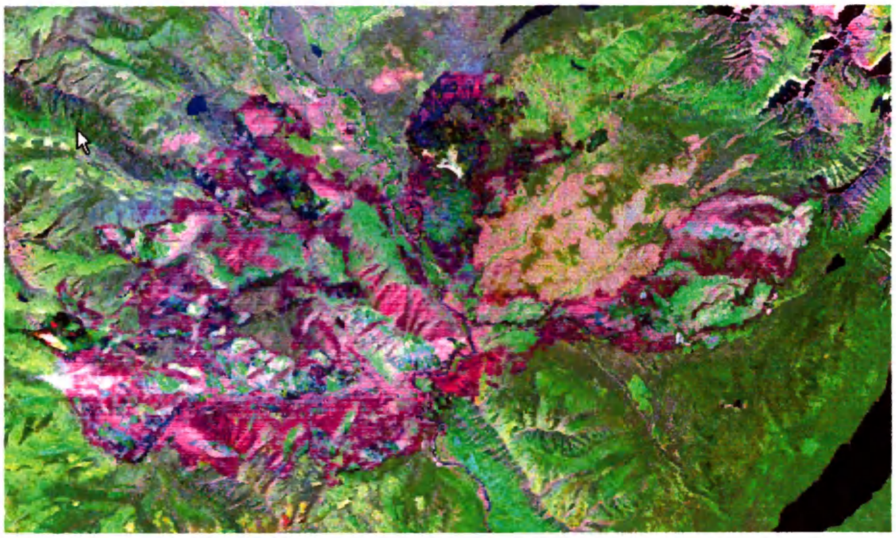


Figure 10: Moose before fire, post fire and difference images

4.3.1. Task

The task is to classify the after fire image into nine classes:

1. Unburned Tree
2. Mixed Burn Tree (mosaic burn)
3. Burned Tree
4. Burned Grassland
5. Unburned Grassland
6. Burned Shrub Land
7. Unburned Shrub Land
8. Barren
9. Water

This training set consists of a set of examples (approximately 10,000 pixels from each layer for each class) selected by an analyst who derived them using air photo interpretation and a life form layer created before the fire.

4.3.2. Specific Methodology

Two experiments were conducted for this image set. The first experiment compared the two images and produced a phase change image that was combined with the after fire image. The second experiment used principle component analysis. Each classification was trained using the standard bands with the phase change image. The principle component analysis used bands 4 and 5 to derive a fire perimeter which was then combined with a lifeform layer to separate the burned and unburned vegetation types out for each lifeform (tree, grass, and shrub). The tree class was further divided into mixed and burned. The results of the learner's classifications were compared to the results of the principle component analysis to show the utility of using the learners to learn change detection versus the more traditional principle component analysis.

4.4. Land Development Experiment

This experiment presents a scenario that explores the utility of including phase change images as additional input into the learner for land development. Given two LANDSAT-5 TM images taken over the city of Missoula spanning a 12-year period classify the changes in land development. The first image was taken in 1984 and the second image was taken in 1996.

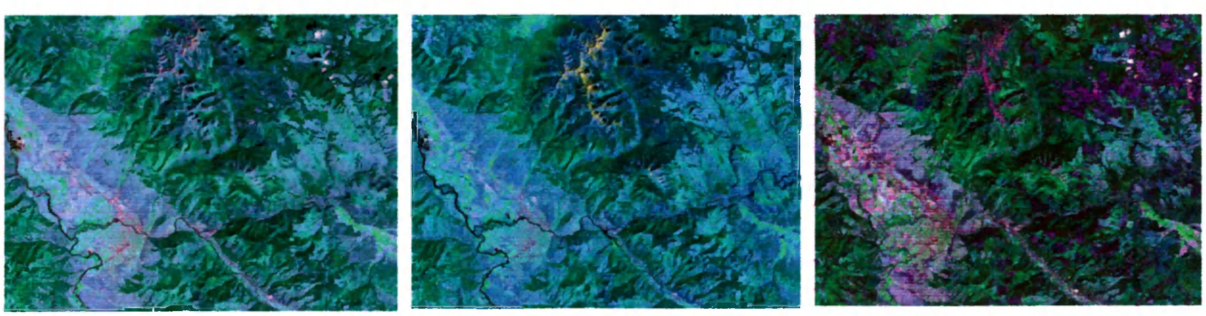


Figure 11: Missoula 1984, 1996 and difference images

4.4.1. Task

The task is to classify the after phase change image and the 1996 image into four classes:

1. No Change
2. Low Change
3. Moderate Change
4. High Change

This training set consists of a set of examples (approximately 15,000 pixels from each layer for each class) selected by an analyst who derived them using air photo interpretation.

4.4.2. Specific Methodology

This experiment compared the two images and produced a phase change image. This phase change image was then added as an additional band to the 1996 image. The features were relatively small so inputs consisted of subimages of size 3 pixels by 3 pixels for each layer. The training examples and truth image were constructed using the lifeform layer as well as city maps of Missoula and the surrounding area. Each classification was trained using the standard bands and the phase change image. These learners were combined as an ensemble and then classified each of the pixels into one of the four classes giving the classification to the learner with the highest prediction.

5. RESULTS AND DISCUSSION

Results for the three study areas are presented to illustrate the utility of the adaptive feature extraction approach in using the difference images and image compression to improve the feature extraction results. The first two experiments illustrate the use of difference images in classifying temporal features of interest for fires while the third experiment presents the utility of classifying temporal features of interest for land development in a scene.

5.1. Ashland Fire Experiment

Figure 12 illustrates the after fire scene as a false color image to highlight the areas that burned where bands 4, 5, and 3 represent the red, green, and blue components of the image. The maroon hues represent the living vegetation and the dark green hues represent the burn areas where the darker the green the more intense the burn. Figure 13 highlights the predicted burn fire perimeter area and was constructed using the hierarchical learning process with the available burn point data.

Figure 12: Ashland after fire false positive image, bands 4, 5, 3 (RGB)

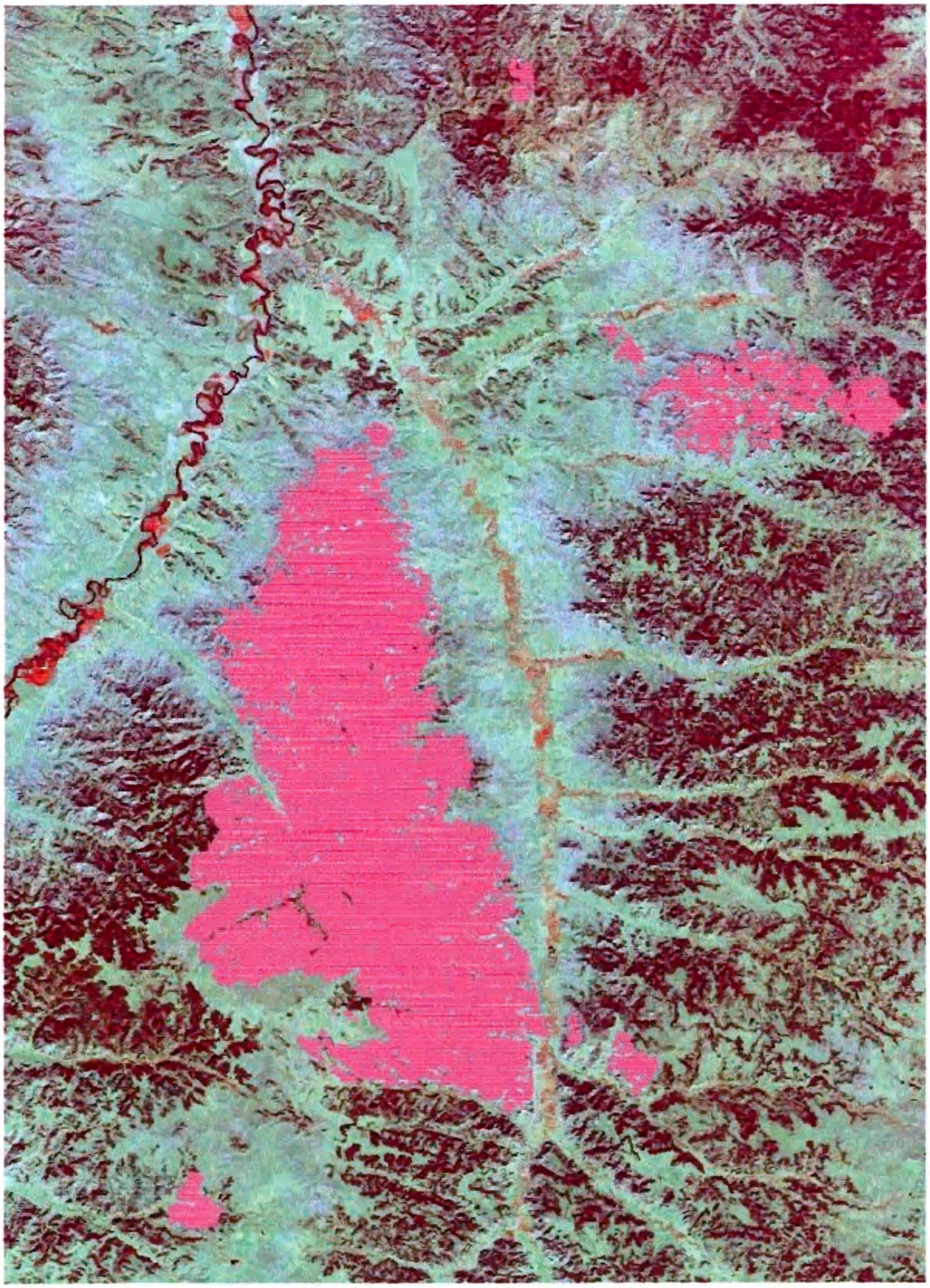
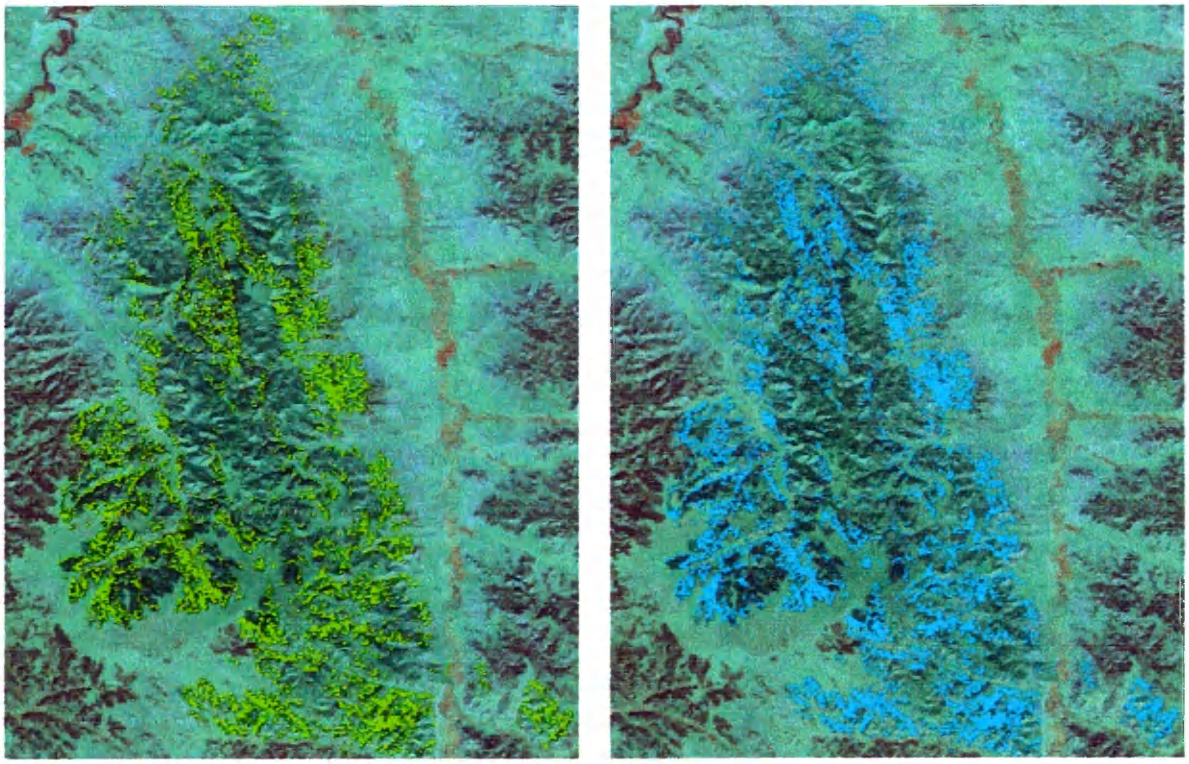


Figure 13: Ashland fire burn perimeter inferred using hierarchical learning process.



(A) (B)

Figure 14: Results for Ashland mixed burn classification. 12(A) Represents the after fire experiment and 12(B) represents the difference image experiment.

Table 1 illustrates the mean percent correct for all of the learners for the after fire experiment and the after and difference image experiment. In the Figure 14 image A represents the mixed burn classification for the after fire image and image B represents the difference fire image results. The differenced image improved the classification results by 9 percent.

Classification	After fire image	After and difference image
Unburned Tree	63	70
Mixed Burn Tree	54	57
Burned Tree	83	80
Burned Grassland	50	75
Unburned Grassland	75	85
Burned Shrub Land	80	100
Unburned Shrub Land	85	90
Barren	72	78
Mean	70	79

Table 1: Mean percent correct by learners for each class

The figures below illustrate the effectiveness of predicting the classification task for each type for each of the classifiers. Figure 15, Figure 16, Figure 17, and Figure 18 depicts the ROC curves for the KNN, ANN, NB and ensemble respectively and Figure 19 depicts the approximate area under the ROC curve for the respective classifiers.

The barren classification was particularly difficult for KNN and can likely be attributed to the limited number of example compared with the actual number of true barren points. The ANN and NB classifier were able, however, to learn the concept. The ensemble seemed to average the effect out and improve overall.

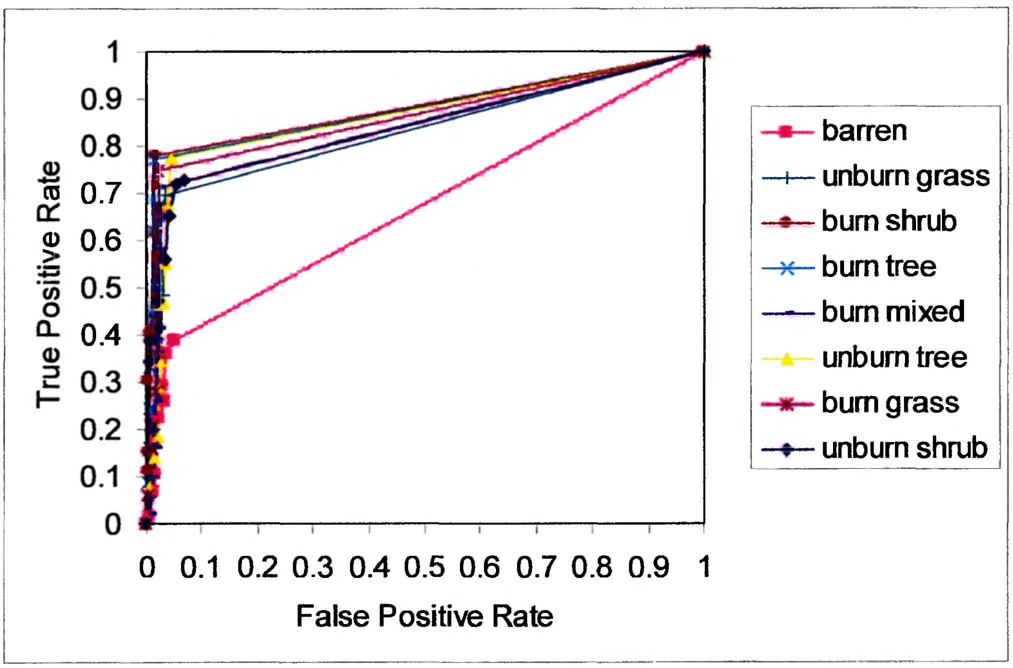


Figure 15: ROC curves for Ashland KNN after fire experiment

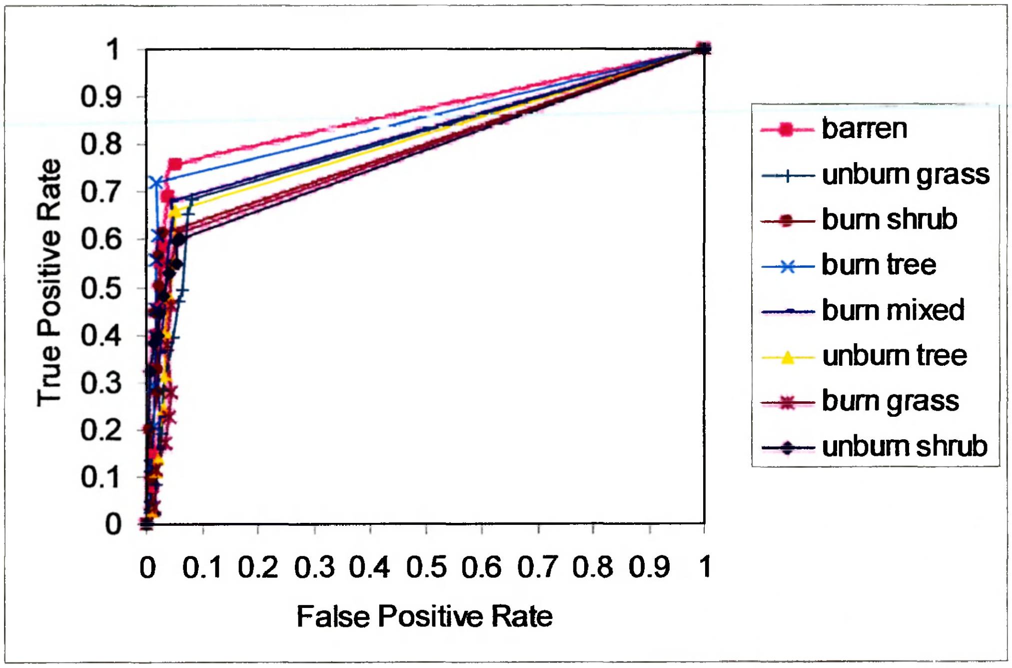


Figure 16: ROC curves for ANN Ashland after fire experiment

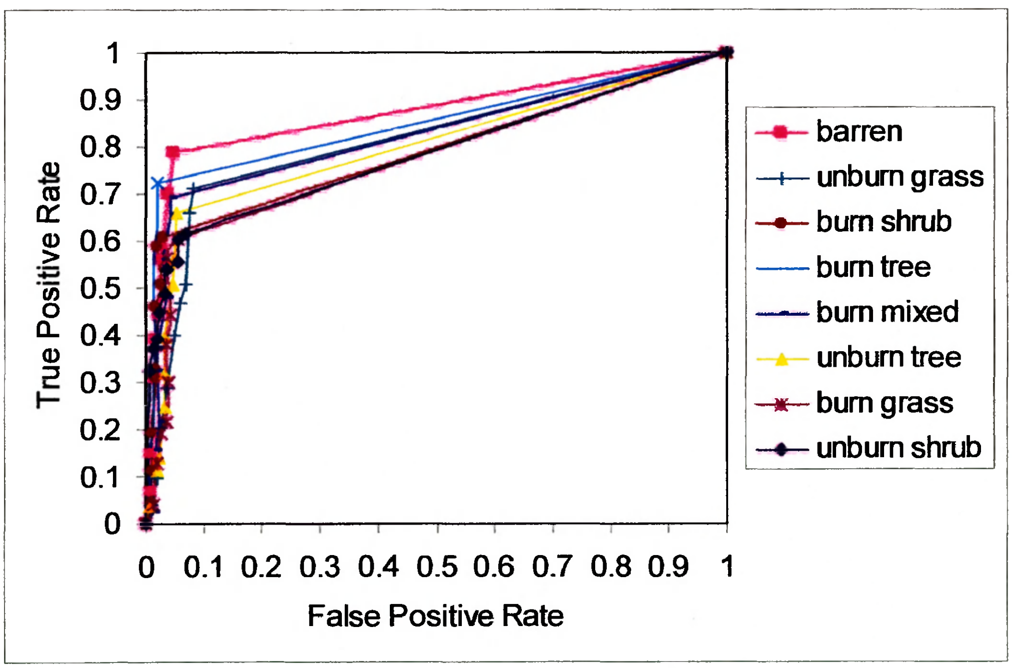


Figure 17: ROC curves for NB Ashland after fire experiment

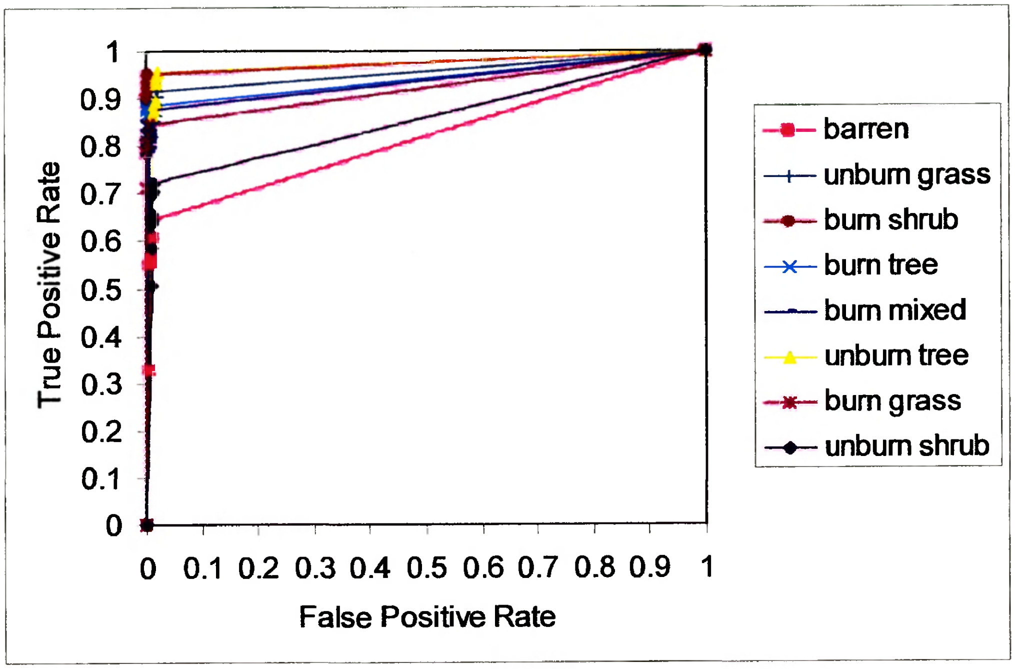


Figure 18: ROC curves for 9-learner ensemble after fire experiment

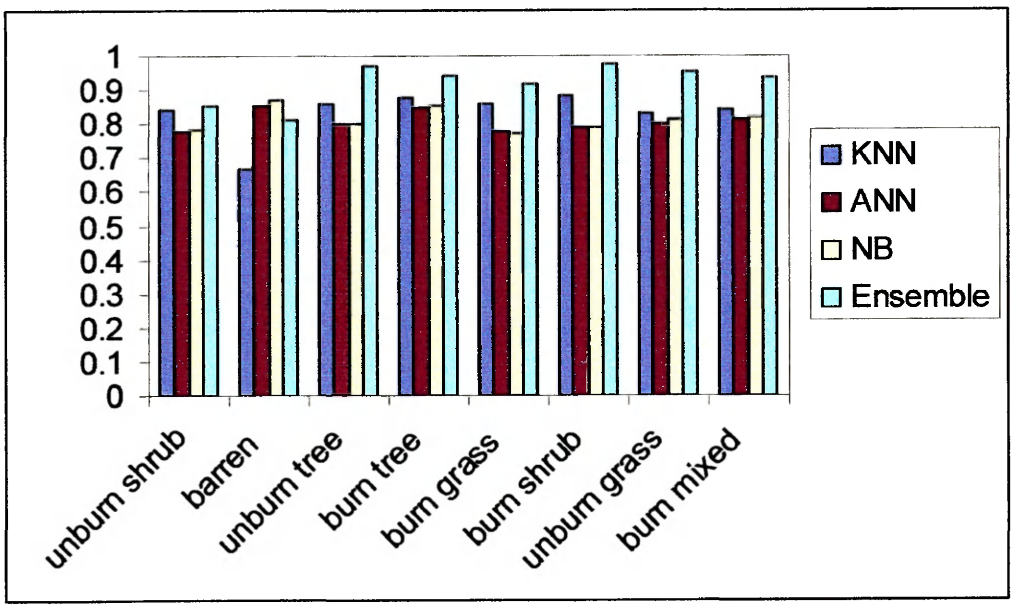


Figure 19: Approximate area under ROC curves for after fire experiment

5.2. Moose Fire Experiment

Figure 20 illustrates the changes that occurred over the moose fire scene from before and after the fire as a false color image using bands 4, 3, and 2 for the red, green and blue bands. The scene shows the changes in the red bands where the darker the purple the more the fire burned a given vegetation class. After one repetition of the hierarchical learning technique the changes in vegetation and the changes in development were classified with a high degree of accuracy.

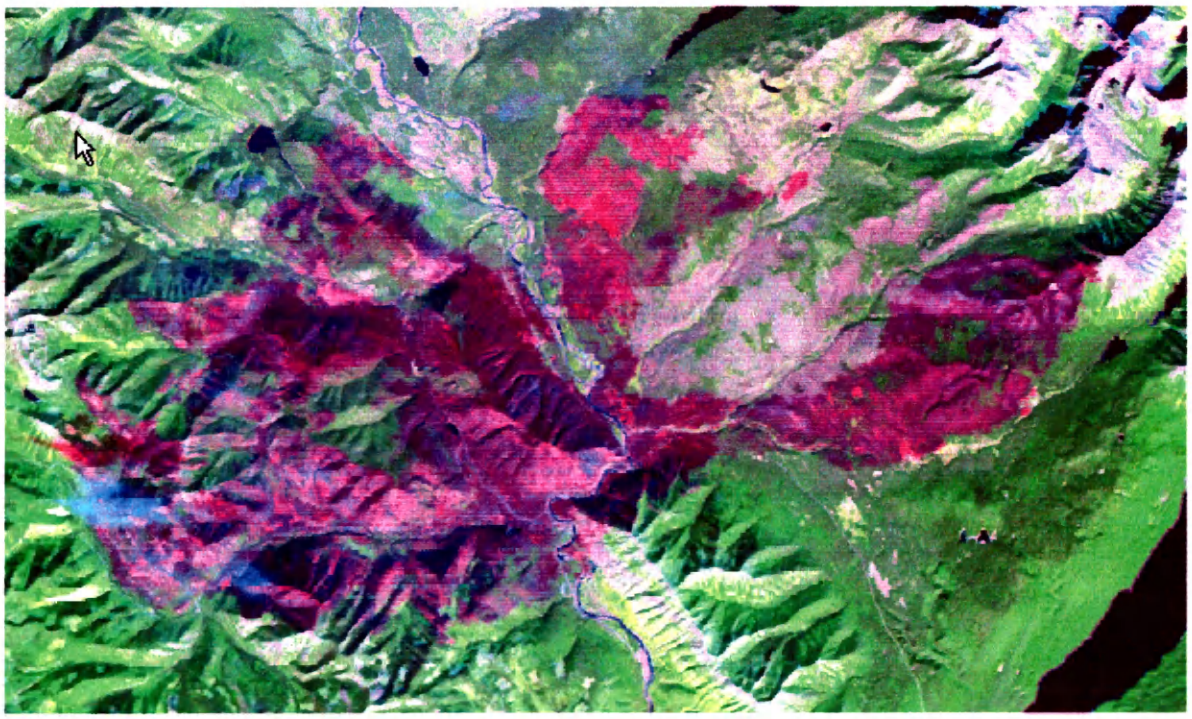


Figure 20: Missoula difference false color image

Table 2 represents the percent correct of each output class of each classifier as well as the percent correct for the principle component analysis of each output class. The figures below represent the classification of the dominant changes in the difference scene. Figure 21, Figure 22, Figure 23, and Figure 24 depicts the ROC curves for the KNN, ANN, NB and ensemble respectively and Figure 25 depicts the approximate area under the ROC curve for the respective classifiers.

The percent correct for each classifier of each output class shows that the ANN and KNN classifiers were effective at classifying the output classes. The PCA classifier was 20 percent less reliable on average from the lowest overall learner, the NB classifier. The ensemble improved the overall accuracy of the learners and illustrated the effectiveness at applying the bagging technique to a variety of classification tasks given all of the classifiers and a random resample of the input data for each classifier.

Classification	**ANN**	**KNN**	**NB**	**Ensemble**	**PCA**
Unburned Tree	93	93	85	96	77
Mixed Burn Tree	82	91	74	91	68
Burned Tree	77	78	70	88	82
Burned Grassland	94	94	74	99	54
Unburned Grassland	93	92	75	98	47
Burned Shrub Land	83	85	84	92	74
Unburned Shrub Land	93	92	85	95	66
Barren	96	97	82	98	59
Water	98	98	99	100	57
Mean	90	91	81	95	66

Table 2: Percent Correct for each classifier

All three classifiers did well overall though the mixed burn class was particularly challenging. The learners often misclassified mixed burn as tree burn or shrub burn. Grass output classes were better represented in the truth points and hence did much better this time than the Ashland experiment. Similarly, mixed burn was better represented in the Ashland experiment and thus did much better than the Moose experiment. In both experiments it was difficult to find a good separation in the mixed burn class since the mixed state varied so much.

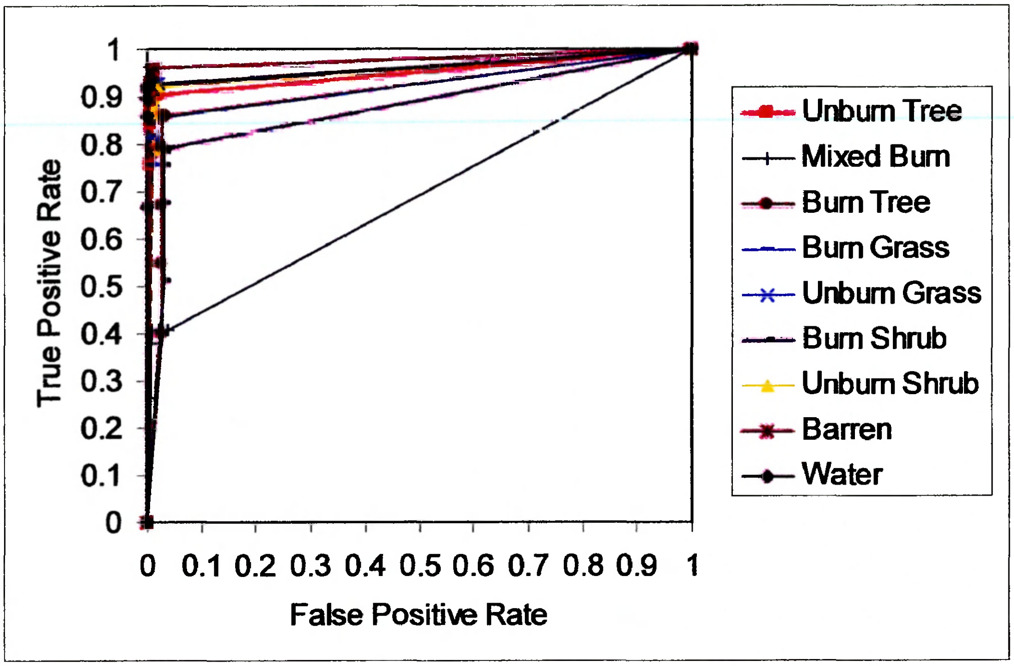


Figure 21: ROC curves for KNN Moose fire experiment

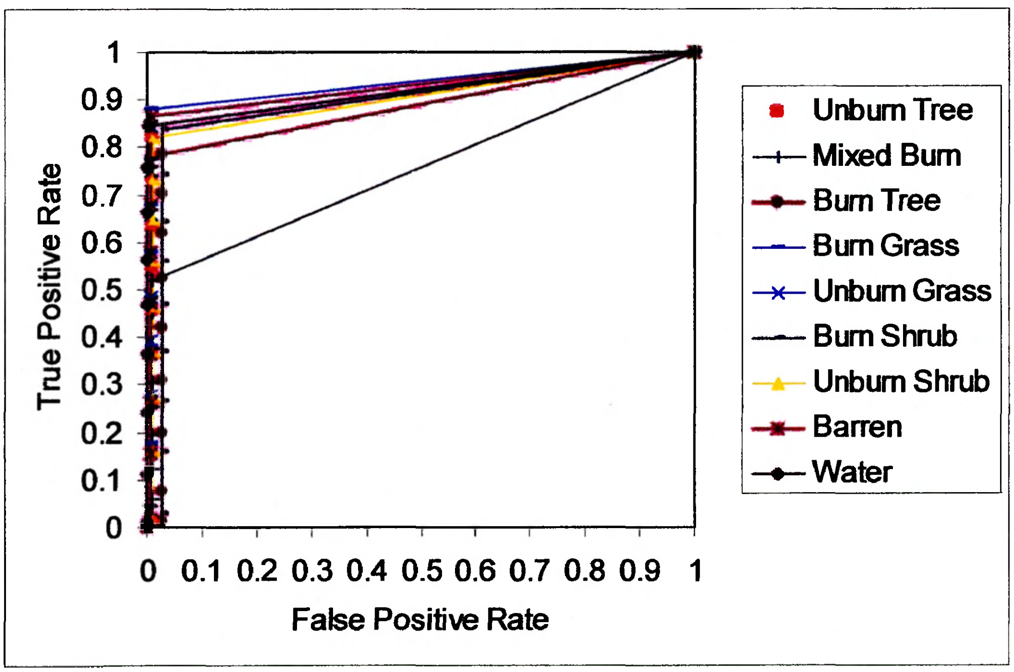


Figure 22: ROC curves for ANN Moose fire experiment

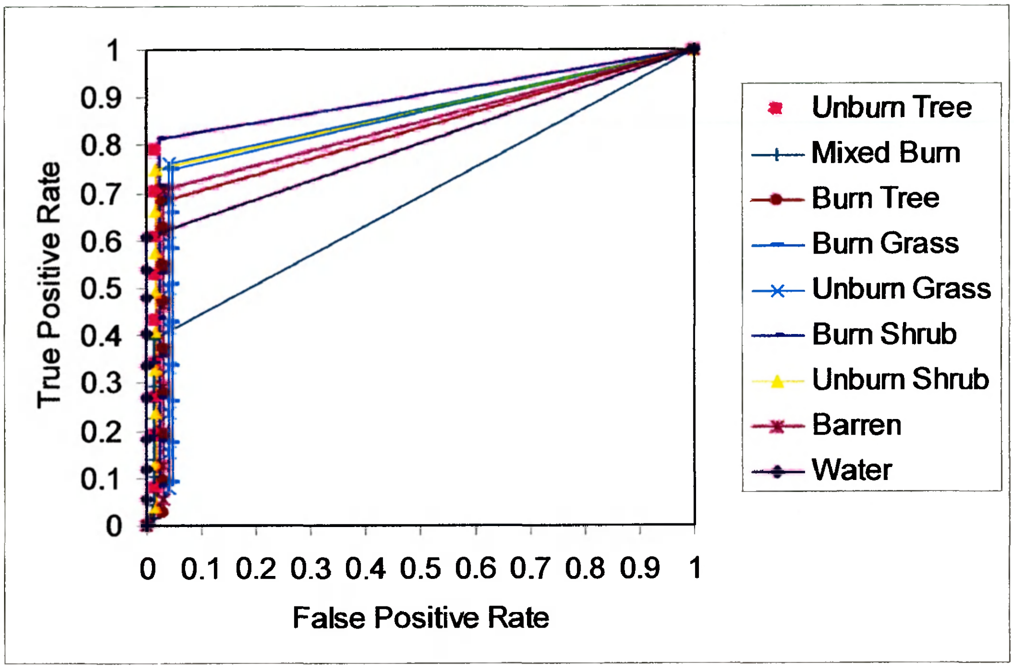


Figure 23: ROC curves for NB Moose fire experiment

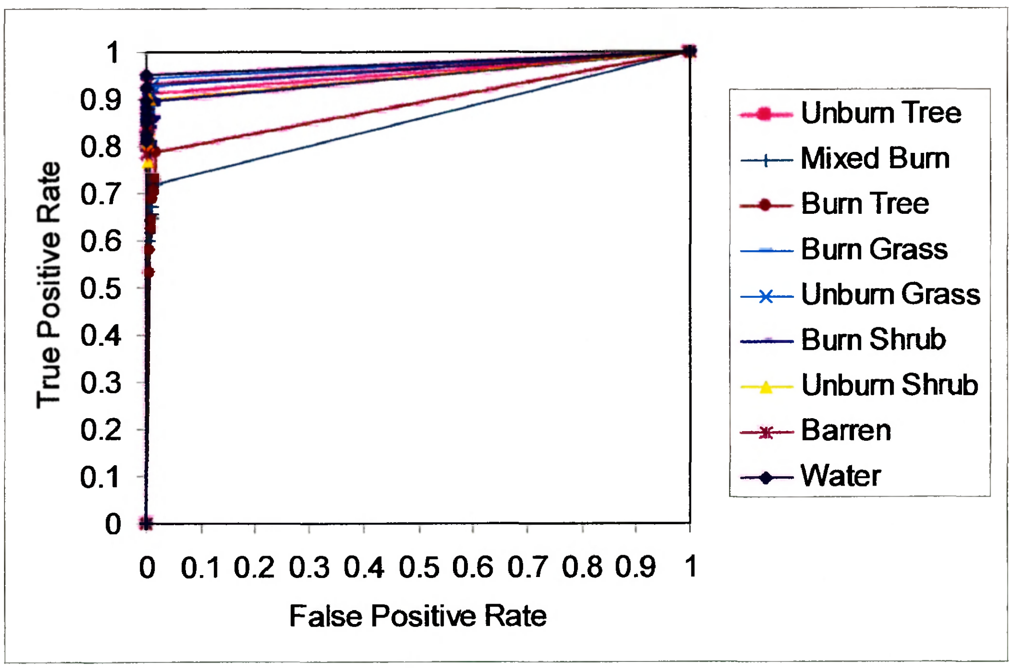


Figure 24: ROC curves for 9-learner ensemble Moose fire experiment

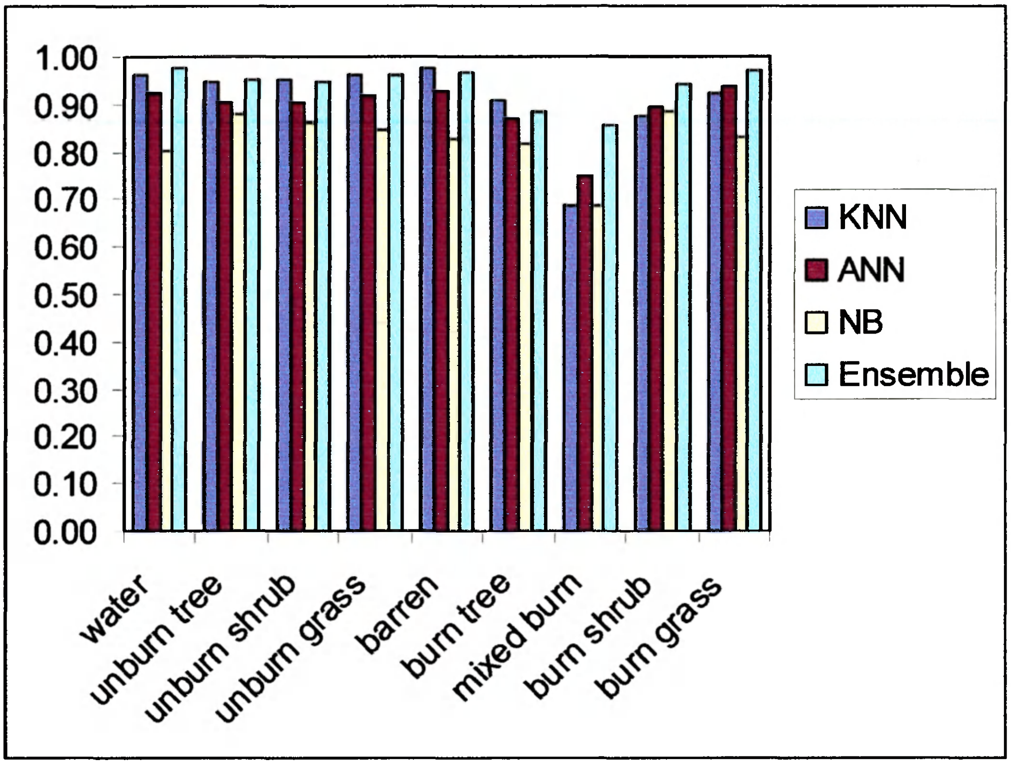


Figure 25: Approximate area under ROC curves for Moose fire experiment

5.3. Land Development Experiment

Figure 26 illustrates the changes that occurred over the Missoula scene from 1984 to 1996 as a false color image using bands 4, 3, and 2 for the red, green and blue bands. The scene shows the changes in the red bands where the darker the red the more the change that occurred particularly with vegetation changes and land development. The green band also depicted changes in development. After one repetition of the hierarchical learning technique the changes in vegetation and the changes in development were classified with a high degree of accuracy.

Figure 26: Missoula difference false color image

The figures below represent the classification of the dominant changes in the difference scene for each learner. Figure 27, Figure 28, Figure 29, and Figure 30 depicts the ROC curves for the KNN, ANN, NB and ensemble respectively and Figure 31 depicts the approximate area under the ROC curve for the respective classifiers.

The KNN performed the best of the three classifiers overall and each learner had the greatest difficulty discriminating between the no development and high development due to the inherent complexity of selecting a good breaking point from no development to low development and moderate development to high development. In the case of no development there was a great deal of mixture of features from roads to streams to vegetation to development to merge together. Breaking the vegetation, roads, streams, and existing development classes out would likely help these results. The ensemble was able to generalize much better while the KNN outputs illustrated a very large separation between no and low development and moderate and high development. Polygons were used to outline the boundaries of change for the truth set and included small portions of another class increasing the error of the training set.

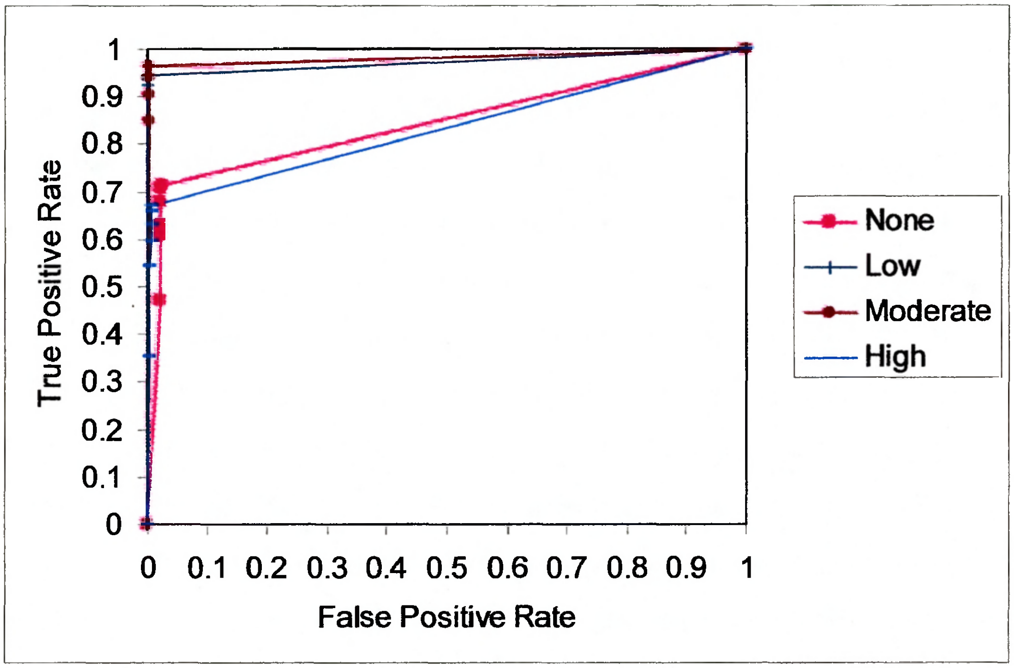


Figure 27: ROC curves for KNN Missoula land development experiment

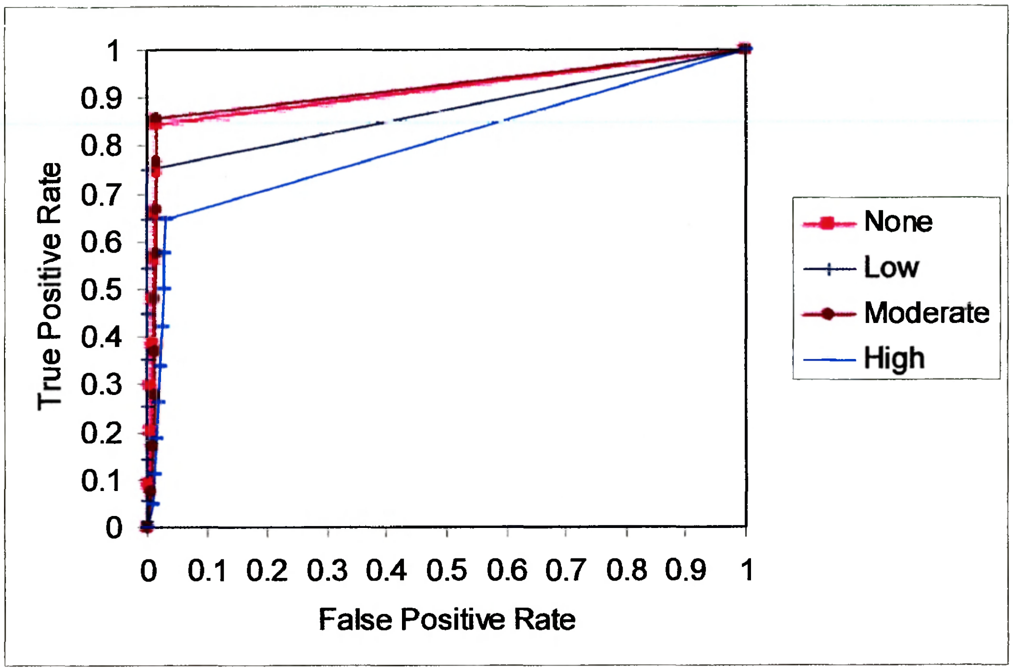


Figure 28: ROC curves for ANN Missoula land development experiment

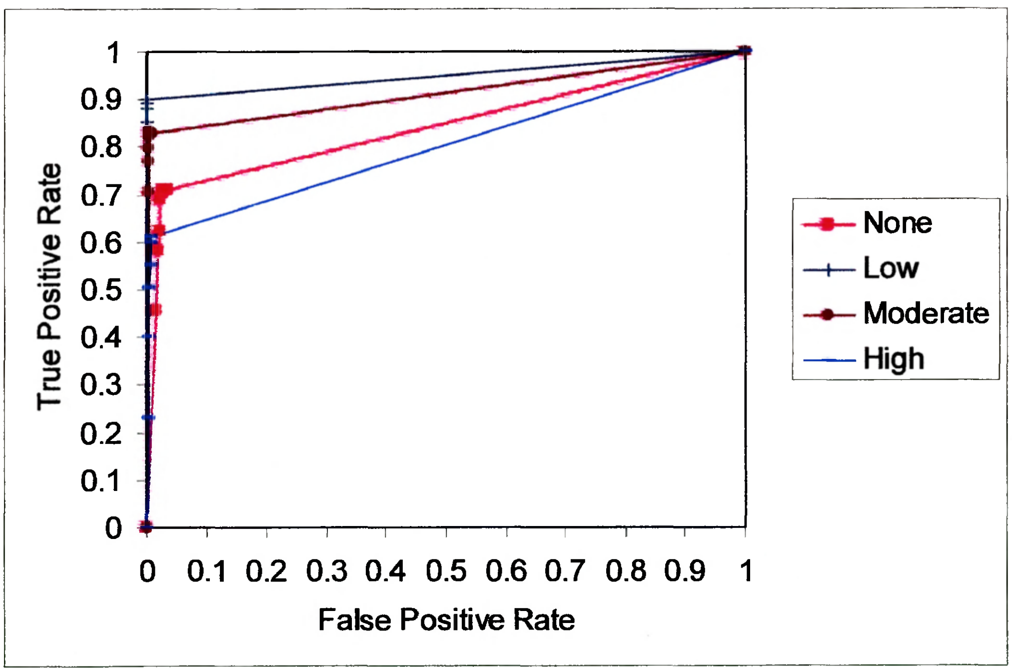


Figure 29: ROC curves for NB Missoula land development experiment

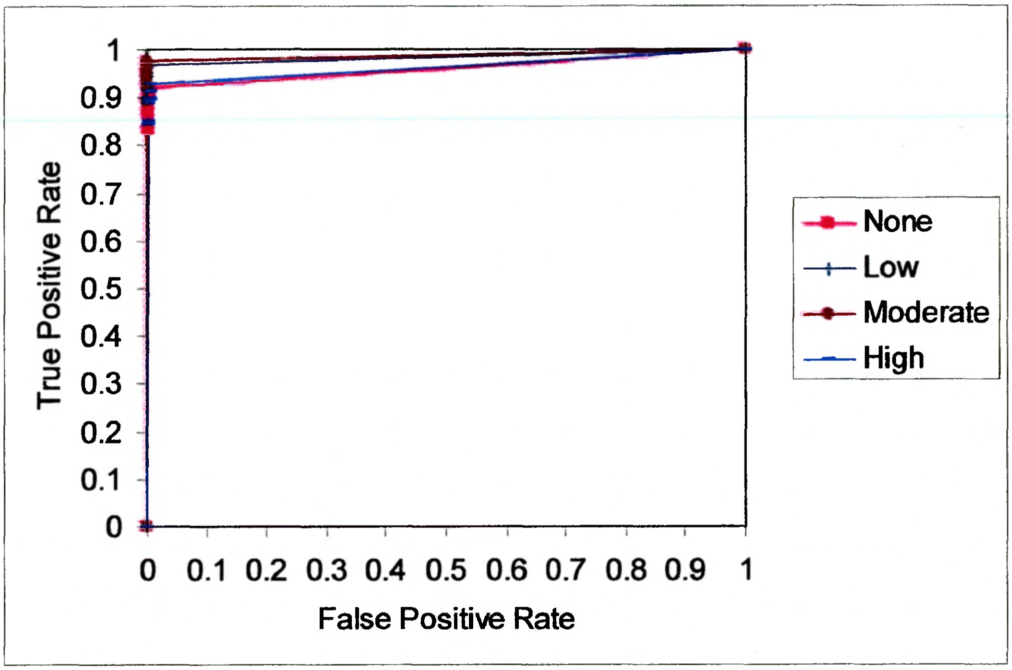


Figure 30: ROC curves for 9-learner ensemble Missoula land development experiment

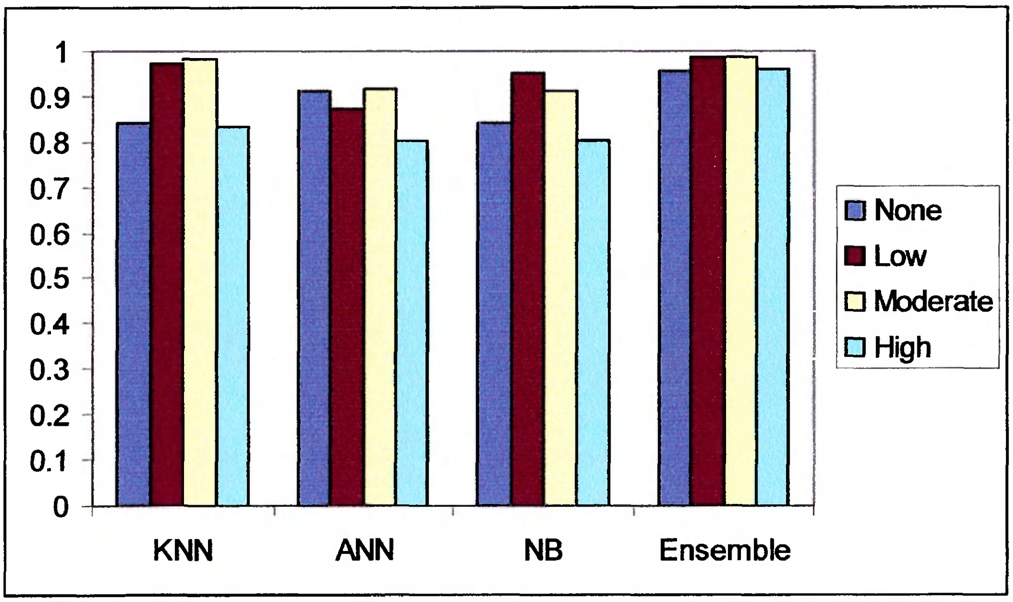


Figure 31: Approximate area under curve for Missoula land development experiment

5.4. Conclusions

Overall, the experiments illustrate the effectiveness of classifying features of change using classifiers and image processing techniques. The use of ensembles for combining multiple output classes and using multiple instances of the classifiers in combination with bagging shows additional utility in detecting the features of change. Additional bagging instances of each of the classifiers would have increased the percent correct and improved the area under the curve. The experiments also pointed out the inherent difficulty in defining a training set that represents the various degrees of change of the given feature especially in a mixed class. The learners were effective at building their own prediction of the separation and the bagging ensemble was able to merge them together to produce a better overall separation while generally increasing the prediction correctness of the individual classes.

6. FUTURE WORK

Through the study of this set of experiments further improvements could be made to the technique to improve classification. Well-known image processing techniques could be used to further enhance the images before the learning process. Further study concerning the relationship between feature complexities versus the learning parameters needed to successfully learn a concept is an important place to start. Class separation especially in terms of degrees of change needs to be investigated with respect to resolution and classifier bias. Ensembles also show great promise in increasing the effectiveness of feature extraction with detecting features of change. The techniques presented also beg further study of the application of discovering feature changes in dynamical systems using multiple temporal scenes over an extended period of time.

6.1. Integrating Image Processing

The application of image processing techniques has been studied for decades but the influence of these techniques on feature extraction with inductive learners has not. In particular with this set of experiments all images assume a two dimensional view of features which caused the landscape to hide features as well as shadows of those features to hide features based on the position of the light sources. Techniques have been development to remove shadows that could reduce some of the negative effects, but more elaborate models will need to be constructed to account for features dependent on landscape characteristics that hide parts of those features. Some improvements were found by biasing each band based on the most probable portions of the training set. This would be another potential area for exploration.

6.2. Feature Complexity vs. Learning Parameters

The Missoula land development experiment took a look at using the adaptive learning process to successively define changes in development. This process of adapting the learner so that the size of the input and learner parameters improves the classification accuracy of a feature by minimizing the complexity of the scene has many areas of study. The time of learning the feature classification can be greatly reduced for a large scene by compressing the image if the feature is sufficiently large. Further refinement could also be improved by using the abstract classification to focus a more complex learner to only the areas where the features were initially detected so that noise is eliminated and a learner can focus on the task to distinguish the desired feature from more similar disjunctive features.

6.3. Ensembles

The experiments presented in this study illustrated the utility of using ensembles to combine multiple classes together as well as using multiple sets of classifiers and bagging to improve feature extraction efforts of change detection. Other ensemble techniques such as Boosting (Opitz, 1999c) and Genetic Ensembles (Opitz, 1999b; Prabu, 2000) need to be investigated to study their utility to change detection. The construction of the ensembles in terms of classifiers, output classifications, resolutions, thresholds, data distributions can occur in many different ways and it would be useful to investigate various configurations of ensembles to find effective constructions for change detection.

6.4. Detection of Changing Features in Dynamical Systems

The experiments presented in this study focused on changes in features over a scene taken over an extended time period. Further investigation needs to look at combining learners that detect features with the learners that detect changes in features so that the learners can adapt to the newly evolved features and manage the evolution of the features into new classifications. Dynamical systems in the real world evolve over time and necessitate the need to keep track of the evolutionary characteristics of entities in the system. Machine vision has focused its efforts particularly in this pursuit and the adaptive feature extraction techniques with a system of keeping track of the features could provide a foundation for building a system to monitor features from a particular viewpoint in a dynamical system.

7. CONCLUSION

The goal of this thesis is to extend the existing feature extraction techniques to new areas and to improve the classification of features through the use of adaptive feature extraction techniques. The study presents the need for a technique to learn change characteristics in features of satellite images over time, the complexity of feature extraction in imagery, and a set of approaches to adapt the imagery and temporal features to the learner to improve the classification. The technique adapts the imagery to the learner, refines the classification, and successfully classifies temporal features in a variety of situations.

8. REFERENCES

Bain, W., (2000). MS Thesis, *A Comparison of Three Machine Learning Algorithms for Auotmated Feature Extraction from Digital Images*. Director: David W. Opitz, University of Montana.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.

Burl, M. et.al. (1998). Learning to Recognize Volcanoes on Venus, *Machine Learning Journal Special Issue on Applications and Knowledge Discovery.* 30:165-194.

Burl, M.C., U. Fayyad, P. Smyth, P. Perona and M.P. Burl (1994). Automating the hunt for volcanoes on Venus. In *Preceedings of the 1994 Conference on Computer Vision and Pattern Recognition.*

Ford, R., Ma, Z., Barsness, S. & Redmond, R. (1997). Rule-Based Aggregation for Region Identification. In *Preceedings of the 1997 American Society for Photogrammetry and Remote Sensing Annual Conference,* Seattle, WA.

Friedman, J. H., Bentley, J. L., Finkel, R. A. (1977). An Algorithm for Finding Best Matches in Logarithmic Expected Time, *ACM Transactions on Mathematical Software, Vol 3, No 3, September 1977.*

Gruen, A., E.P. Baltsavias, O. Henricsson Birkhäuser Verlag Basel, Boston, Berlin, (1997). Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) *Bi-annual Ascona conference, held in Ascona, Switzerland.*

Hansen, L. and Salamon, P., (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 993-1001.

Haykin, S., (1999). *Neural Networks: A Comprehensive Foundation*, 2nd ed, Prentice Hall, Inc.

Hernández, E. and Weiss, G., (1996). *A First Course on Wavelets*, CRC Press.

Hubbard, B., (1996). *The World According to Wavelets*, Wellesley, MA: A K Peters.

Jensen, J. R. and D. C. Cowen, (1997). "Remote Sensing of Urban/Suburban Socioeconomic Attributes," *Land Satellite Information in the Next Decade II: Sources and Applications,* Bethesda: American Society for Photogrammetry & Remote Sensing, (December 5); CD, 19 p.

Kramber, W. J., and A. Morse, (1994). "Integrating Image Interpretation and Unsupervised Classification Procedures." *ACSM/ASPRS Annual Convention & Exposition*. Baltimore: ACSM/ASPRS, 1994. 1:327-336.

Langley, P., (1996). *Elements of Machine Learning*, Morgan Kaufman Publishers, Inc.

Lillesand, T., and Keifer,R., (1999). *Remote Sensing and Image Interpretation*, 4th ed: John Wiley & Sons, Inc..

Lunetta, R., and Elvidge, C., (1999). *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, Taylor & Francis.

Maloof, M.A. (2002). On machine learning, ROC analysis, and statistical tests of significance. *Proceedings of the Sixteenth International Conference on Pattern Recognition*, 204-207, Los Alamitos, CA: IEEE Press.

Maloof, M. et al. (1998). Learning to detect rooftops in aerial images. In *Proceedings of the Image Understanding Workshop*, (pp. 835-845).

Mangrich, M., (2001). MS Thesis, *Hierarchical Feature Extraction: A Stepwise Approach to Image Classification*. Director: David W. Opitz, University of Montana.

Mather, P. (1999). *Computer Processing of Remotely-Sensed Images, An Introduction*, 2nd ed.

Mitchell, T. (1997). *Machine Learning*, Boston: MIT Press.

Opitz, D. (1999a). Feature Selection for Ensembles. *Sixteenth National Conference on Artificial Intelligence (AAAI)*, (379-384). Orlando, FL.

Opitz, D. and Shavlik, J. (1999b). A Genetic Algorithm Approach for Creating Neural Network Ensembles. Combining Artificial Neural Nets. *Amanda Sharkey (ed.)*. (pp. 79-97). *Springer-Verlag, London.*

Opitz, D. and Maclin, R. (1999c). Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research*, (pp. 169-198).

Opitz, D., W. Bain, M. Mangrich, J. Zeiler, and M. Sweet, (2000). *Rapid Extraction of Objects from Digital Images*. Department of Computer Science, University of Montana.

Palhang, M., A. Sowmya, (1997). Two Stage Learning, Two Stage Recognition, *Proceedings of Australian Joint Conference on Artificial Intelligence*, Perth Australia, 2-4 Dec. 1997, 1997.

Prabu, Ganesh, (2000). MS Thesis, Genetic Ensemble Feature Selection. Director: David W. Opitz, University of Montana.

Seul, M., L. O'Gorman, M. Sammon, (2000). *Practical Algorithms for Image Analysis: Description, Examples, and Code.*, Cambridge Univ Press.

Stollnitz, E., Derose, T., and Salesin, D., (1996). *Wavelets for Computer Graphics Theory and Applications*, Morgan Kaufmann Publishers, Inc.

Swets, J.A. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–1293.

Theodoridis, S., Koutroumbas, K., (1999). *Pattern Recognition*, Academic Press.

Watt, A. and Policarpo, F., (1997). *The Computer Image*, Addison-Wesley Publishing.

Williams, G., (1999). *Chaos Theory Tamed*, Joseph Henry Press/National Academy Pres.

Woodcock, C.E., and V.J. Harward, (1992). Nested-Hierarchical Scene Models and Image Segmentation, *International Journal of Remote Sensing*, 13(16):3167-3187.