THE IMPACT OF TRAINING SET SIZE AND FEATURE DIMENSIONALITY ON SUPERVISED OBJECT-BASED CLASSIFICATION: A COMPARISON OF THREE CLASSIFIERS

by
Gerhard Myburgh

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in the Faculty of Science at Stellenbosch University.

Supervisor : Dr Adriaan van Niekerk

December 2012

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 7 September 2012

SUMMARY

Supervised classifiers are commonly used in remote sensing to extract land cover information. They are, however, limited in their ability to cost-effectively produce sufficiently accurate land cover maps. Various factors affect the accuracy of supervised classifiers. Notably, the number of available training samples is known to significantly influence classifier performance and to obtain a sufficient number of samples is not always practical. The support vector machine (SVM) does perform well with a limited number of training samples. But little research has been done to evaluate SVM's performance for geographical object-based image analysis (GEOBIA). GEOBIA also allows the easy integration of additional features into the classification process, a factor which may significantly influence classification accuracies. As such, two experiments were developed and implemented in this research. The first compared the performances of object-based SVM, maximum likelihood (ML) and nearest neighbour (NN) classifiers using varying training set sizes. The effect of feature dimensionality on classifier accuracy was investigated in the second experiment.

A SPOT 5 subscene and a four-class classification scheme were used. For the first experiment, training set sizes ranging from 4-20 per land cover class were tested. The performance of all the classifiers improved significantly as the training set size was increased. The ML classifier performed poorly when few (<10 per class) training samples were used and the NN classifier performed poorly compared to SVM throughout the experiment. SVM was the superior classifier for all training set sizes although ML achieved competitive results for sets of 12 or more training samples per class. Training sets were kept constant (20 and 10 samples per class) for the second experiment while an increasing number of features (1 to 22) were included. SVM consistently produced superior classification results. SVM and NN were not significantly (negatively) affected by an increase in feature dimensionality, but ML's ability to perform under conditions of large feature dimensionalities and few training areas was limited.

Further investigations using a variety of imagery types, classification schemes and additional features; finding optimal combinations of training set size and number of features; and determining the effect of specific features should prove valuable in developing more cost-effective ways to process large volumes of satellite imagery.

OPSOMMING

Gerigte klassifiseerders word gereeld aangewend in afstandswaarneming om inligting oor landdekking te onttrek. Sulke klassifiseerders het egter beperkte vermoëns om akkurate landdekkingskaarte koste-effektief te produseer. Verskeie faktore het 'n uitwerking op die akkuraatheid van gerigte klassifiseerders. Dit is veral bekend dat die getal beskikbare opleidingseenhede 'n beduidende invloed op klassifiseerderakkuraatheid het en dit is nie altyd prakties om voldoende getalle te bekom nie. Die steunvektormasjien (SVM) werk goed met beperkte getalle opleidingseenhede. Min navorsing is egter gedoen om SVM se verrigting vir geografiese objek-gebaseerde beeldanalise (GEOBIA) te evalueer. GEOBIA vergemaklik die integrasie van addisionele kenmerke in die klassifikasie proses, 'n faktor wat klassifikasie akkuraathede aansienlik kan beïnvloed. Twee eksperimente is gevolglik ontwikkel en geïmplementeer in hierdie navorsing. Die eerste eksperiment het objekgebaseerde SVM, maksimum waarskynlikheids- (ML) en naaste naburige (NN) klassifiseerders se verrigtings met verskillende groottes van opleidingstelle vergelyk. Die effek van kenmerkdimensionaliteit is in die tweede eksperiment ondersoek.

'n SPOT 5 subbeeld en 'n vier-klas klassifikasieskema is aangewend. Opleidingstelgroottes van 4-20 per landdekkingsklas is in die eerste eksperiment getoets. Die verrigting van die klassifiseerders het beduidend met 'n toename in die grootte van die opleidingstelle verbeter. ML het swak presteer wanneer min (<10 per klas) opleidingseenhede gebruik is en NN het, in vergelyking met SVM, deurgaans swak presteer. SVM het die beste presteer vir alle groottes van opleidingstelle alhoewel ML kompeterend was vir stelle van 12 of meer opleidingseenhede per klas. Die grootte van die opleidingstelle is konstant gehou (20 en 10 eenhede per klas) in die tweede eksperiment waarin 'n toenemende getal kenmerke (1 tot 22) toegevoeg is. SVM het deurgaans beter klassifikasieresultate gelewer. SVM en NN was nie beduidend (negatief) beïnvloed deur 'n toename in kenmerkdimensionaliteit nie, maar ML se vermoë om te presteer onder toestande van groot kenmerkdimensionaliteite en min opleidingsareas was beperk.

Verdere ondersoeke met 'n verskeidenheid beelde, klassifikasie skemas en addisionele kenmerke; die vind van optimale kombinasies van opleidingstelgrootte en getal kenmerke; en die bepaling van die effek van spesifieke kenmerke sal waardevol wees in die ontwikkelling van meer koste effektiewe metodes om groot volumes satellietbeelde te prosesseer.

TREFWOORDE

Gerigte klassifikasie, landdekking, steunvektormasjien, naaste naburige klassifikasie, maksimum waarskynlikheidsklassifikasie, geografiese objekgebaseerde beeldanalise

ACKNOWLEDGEMENTS

I sincerely thank:

- Dr. Adriaan van Niekerk, my supervisor, for continued guidance, meaningful advice and helpful suggestions.

- The staff of the Department of Geography and Environmental Studies for helpful comments and constructive criticism during scheduled feedback sessions.

- Dr. Pieter de Necker for thorough editorial work on this document.

CONTENTS

TABLES

FIGURES

ACRONYMS AND ABBREVIATIONS

ANN	Artificial neural network
ASTER	Advanced spaceborne thermal emission reflection radiometer
AVIRIS	Airborne visible/infrared imaging spectrometer
CTA	Classification Tree Analysis
DA	Discriminant analysis
DAIS	Digital airborne imaging system
DE	Density estimation
DN	Digital number
DT	Decision tree
ESRI	Environmental Systems Research Institute
ETM	Enhanced thematic mapper
ETM+	Enhanced thematic mapper plus
GEOBIA	Geographic object-based image analysis
GDAL	Geospatial data abstraction library
GIS	Geographical information systems
GLCM	Grey level co-occurrence matrix
HSI	Hue saturation intensity
ICM	Iterated conditional modes
kNN	k-nearest neighbour
Libsvm	Library for support vector machines
ML	Maximum likelihood
NN	Nearest neighbour
OpenCV	Open source computer vision
PolSAR	Polarimetric synthetic aperture radar
OSH	Optimal separating hyperplane
RBF	Radial basis function
RS	Remote sensing
SPOT	Système Pour l'Observation de la Terre
SVM	Support vector machine
TM	Thematic mapper

CHAPTER 1: INTRODUCTION

Land cover refers to the physical characteristics of the earth's surface (Campbell 2006), and spatial knowledge about these characteristics is crucial for environmental and socio-economic research (Heinl *et al.* 2009; Lu & Weng 2007). Thematic maps are typically used to represent land cover information spatially and detailed, accurate and up-to-date land cover maps are required by many applications. Remotely sensed imagery of the earth's surface is a convenient source of information from which land cover maps may, through the application of image classification techniques, be derived (Foody 2002). This has long been a driving force for research on remote sensing (RS) image classification (Lu & Weng 2007). RS techniques are less costly than traditional ground survey methods and offer large area coverage and more frequent data availability (Foody 2009; Pal & Mather 2004). The success of image classification is, however, influenced by a wide variety of factors (Lu & Weng 2007) and resulting land cover maps are often inadequate for operational use (Foody 2002). Consequently, RS research is often focused on finding ways of improving classification accuracies (Foody & Mathur 2004b; Lu & Weng 2007). Automatic and semi-automatic processing of RS imagery is currently limited and research on the factors that influence classification accuracies, the comparison of different classifiers and the introduction of novel classification techniques is driven by the need of finding cost-effective ways to process the ever increasing volumes of available RS data (Baraldi *et al.* 2010).

Supervised classification is an approach commonly employed for digital image classification tasks within the field of RS. Supervised classifiers are theoretically well-founded algorithms requiring a set of known samples (training samples) to predict samples of unknown identity. Numerous, accurate, well-distributed and sufficiently representative training samples are typically required to perform a successful classification (Campbell 2006; Lu & Weng 2007). The collection and delineation of adequate training data is a considerable drawback of supervised classification (Stephenson & Van Niekerk 2009) as it is a time-consuming, expensive and tedious process, and often necessitates a number of field visits and the study of maps and aerial photographs (Campbell 2006).

Many supervised classifiers, each with their own advantages and disadvantages, have been applied in RS and the selection of an appropriate classifier is a key consideration for all image classification problems. Various factors, such as the nature of the study area, the spatial resolution of the remotely sensed data, the classification scheme, the number of training samples available and the number of features used may impact classification results differently depending on the choice of classifier (Lu & Weng 2007).

The support vector machine (SVM) is a supervised classifier that has recently generated interest from the RS community (Mountrakis, Im & Ogole 2011). While SVMs are not yet well known, they have produced equivalent or superior results for remote sensing classification problems compared to traditionally used classifiers (Camps-Valls & Bruzzone 2005; Camps-Valls *et al.* 2004; 2006; Dixon & Candade 2008; Foody & Mathur 2004a; Huang, Davis & Townshend 2002; Kavzoglu & Colkesen 2009; Keuchel *et al.* 2003; Melgani & Bruzzone 2004; Mercier & Lennon 2003; Oommen *et al.* 2008; Pal & Mather 2004; 2005; Szuster, Chen & Borger 2011; Tzotsos & Argialas 2008). SVM is particularly suited for dealing with RS problems as it performs well with limited training samples (Foody & Mathur 2004b; Li *et al.* 2010; Lizarazo 2008; Mountrakis, Im & Ogole 2011; Pal & Mather 2005) and it is robust in issues of input dimensionality (Oommen *et al.* 2008). Comparative analyses of SVM have been restricted to traditional pixel-based classification approaches and the investigation for object-based image classification problems has been limited. Tzotsos & Argialas (2008) favourably compared SVM to the nearest neighbour (NN) classifier for object-based land cover classification while other studies have successfully applied object-based SVM in a remote sensing context (Li *et al.* 2010; Lizarazo 2008; Meng & Peng 2009; Tzotsos, Karantzalos & Argialas 2011; Wu *et al.* 2009).

1.1 PROBLEM FORMULATION

Given the recent shift from pixel-based to object-based research on the classification of remotely sensed data (Tzotsos, Karantzalos & Argialas 2011) and the significant differences existing between the two approaches (Blaschke 2010), the potential of SVM for object-based land cover classification calls for investigation. The performance of SVM using few training areas is particularly appealing as the number available samples is typically smaller in the case of geographic object-based image analysis (GEOBIA) than in traditional pixel-based approaches (Tzotsos & Argialas 2008). GEOBIA also allows the easy incorporation of additional spectral, textural and contextual features which could significantly affect classification accuracies. Little research has been done to evaluate SVM's performance for GEOBIA. The ability of a classifier to perform well under small training set-size and high feature dimensionality conditions is crucial for GEOBIA and a comparison between SVM and traditional classifiers, such as maximum likelihood (ML) and nearest neighbor (NN), is necessary to assess SVM's potential for object-based land cover classification under such conditions.

1.2   AIM AND OBJECTIVES

The aim of this research is to compare the performance of SVM, NN and ML classifiers for object-based land cover classification and to evaluate each classifier according to two key variables, namely the number of training samples and the number of additional object features.

To achieve this aim, the objectives of the study are to:

1. Review the literature on general and specific remote sensing concepts relevant to the study.

2. Obtain and prepare suitable satellite imagery.

3. Develop a software system capable of performing object-based SVM, NN and ML classification as well as automated accuracy assessment.

4. Use the software system to conduct a robust experiment to evaluate the SVM, NN and ML classifiers according to the number of training samples used to train each classifier.

5. Conduct a similar experiment to evaluate SVM, NN and ML when more object features are added as classification input.

6. Report and interpret the results of the experiments as they relate to land cover classification from remotely sensed data.

1.3   RESEARCH METHODOLOGY AND AGENDA

An experimental approach was followed in this research. Two experiments were carried out using empirically derived datasets (digital satellite imagery and selected class samples) and quantitative methods (SVM, NN and ML classification algorithms). The two experiments investigated the influence of two variables – number of training samples and number of object features respectively – on the outcomes of the three methods.

Figure 1.1 shows the research design and the order of the thesis chapters. The research problem and the aims and objectives have been set out in this chapter. Chapter 2 overviews the characteristics of remotely sensed imagery, common approaches to image classification (unsupervised, supervised and rule-based classification) and the differences between pixel-based and object-based classification. A discussion of literature on SVMs' potential for land cover classification is also included.

The design and the results of the first experiment (an investigation of the influence of the training set size on classifier accuracies for object-based classification) are provided in

Chapter 3, while Chapter 4 describes the design and results of the second experiment (an investigation of the effect of feature dimensionality). These chapters also provide brief theoretical discussions on the SVM, ML and NN classifiers and details on the study area and satellite imagery that was used. It should be noted that Chapter 3 and Chapter 4 were prepared for submission to respective scientific journals and that, due to the same methods and data being used for both experiments, some text, figures and tables are duplicated in these chapters. The findings of both exercises are summarized in the final chapter which concludes with suggestions for future research.
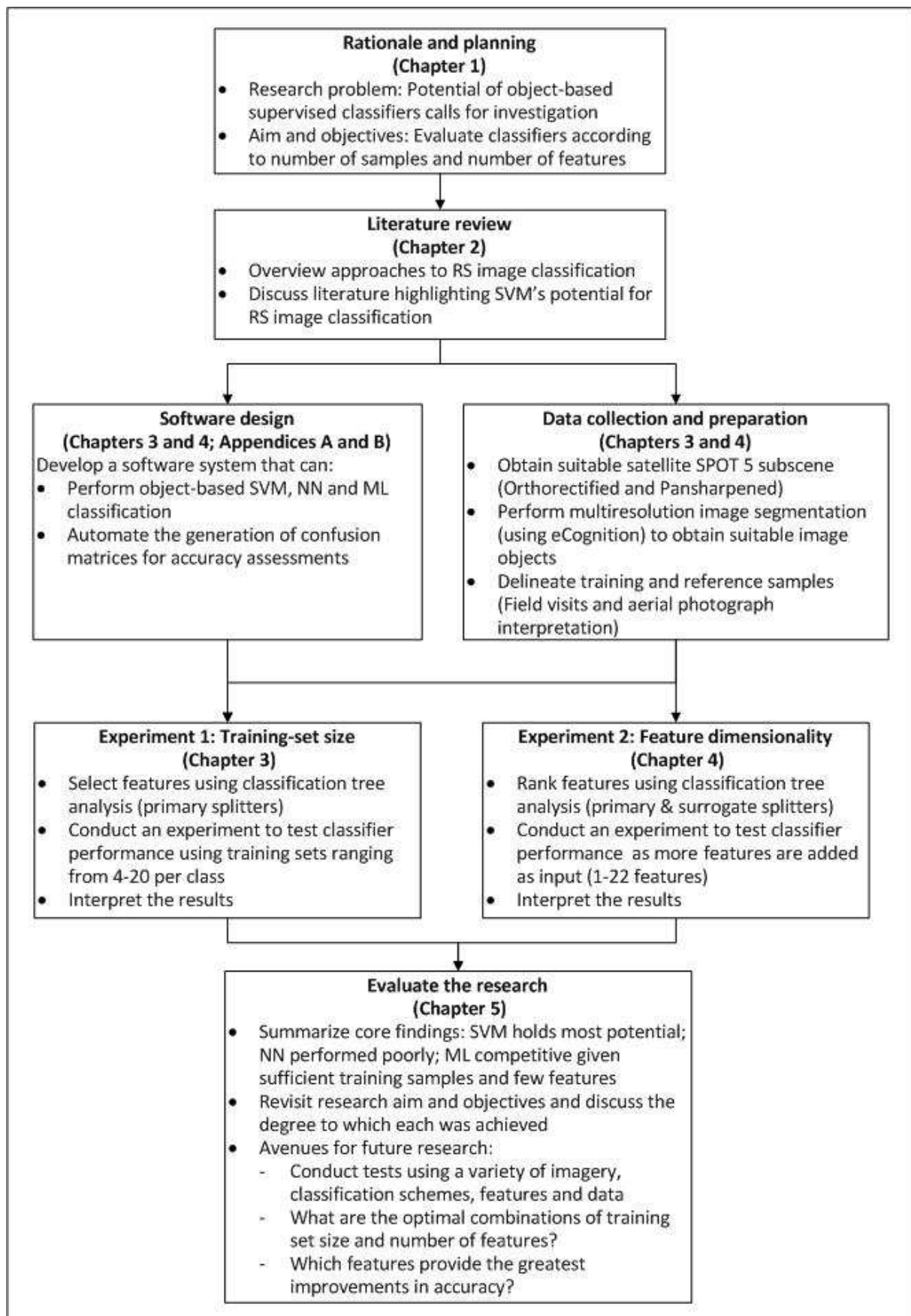
Figure 1.1: Research design for evaluating the performance of object-based SVM, NN and ML classifiers according to the number of training samples and feature dimensionality.

A review of relevant literature is presented in the next chapter.

CHAPTER 2: APPROACHES TO IMAGE CLASSIFICATION

The adoption of a suitable classification approach is crucial for successfully classifying RS data. This chapter first overviews the characteristics of remotely sensed imagery as this knowledge is essential for making informed decisions about specific problems concerning land cover classification. A discussion follows on classification approaches that have been successfully applied in RS. Finally, special attention is given to literature regarding the performance of SVMs for RS classification.

2.1 REMOTELY SENSED IMAGERY

The term *remote sensing* (RS) refers to the acquisition of information from a distance (i.e. the device collecting the data is not in physical contact with the object or phenomenon under investigation) (Campbell 2006; Lillesand, Kiefer & Chipman 2008). RS, with such a broad definition, may comprise many activities. However, modern usage of the term is commonly reserved for the science concerned with the observation of the earth's surface and atmosphere through the measurement of reflected or emitted electromagnetic energy (Campbell 2006; Mather 2004). RS will be regarded as such throughout this thesis.

All objects on the earth's surface reflect or emit certain amounts of the sun's electromagnetic energy at different wavelengths depending on their physical characteristics. Remotely sensed data is typically obtained from sensors, on board satellites or aircraft, designed to measure and record the amount of reflected or emitted energy for specific regions (bands) of the electromagnetic spectrum. While the sun's energy is the source of radiation recorded by most sensors (passive sensors), active sensors supply their own source of energy and record the portion of energy that is scattered back from the earth (Campbell 2006; Mather 2004). The recorded data is represented as a digital image consisting of a regularly spaced array of pixels for each band (Gao 2009). Such an image is known as a *raster* image (Mather 2004). Each pixel represents an area of the earth's surface as determined by its *cell size,* each has a location in two-dimensional space and each has a *digital number* (DN) as label (Gao 2009). The DN of a pixel is an integer value representation of the reflected or emitted energy measured by the sensor.

The characteristics of remotely sensed data vary among the range of currently operational systems. The spatial, spectral, radiometric and temporal resolutions of a system are its defining characteristics and determine the usefulness of the data for specific RS problems. The *spatial resolution* of a system is the dimensions of the smallest area that can be separately recorded and it is in most cases synonymous with the cell size of a raster image (Campbell

2006; Gao 2009). Higher levels of detail can be achieved at higher resolutions and the selection of an appropriate spatial resolution depends on the scale of the problem (Chuvieco & Huete 2010). *Spectral resolution* is the number of operational bands and their individual spectral bandwidths (Chuvieco & Huete 2010). Conventional *multispectral* sensors measure spectral responses in a handful of broadly defined channels while *hyperspectral* imagery consists of many narrowly defined spectral bands (Campbell 2006). The number of available bands and their spectral ranges affects the discernibility of certain features and requires consideration according to the problem at hand (Chuvieco & Huete 2010). The number quantization levels used to express the DNs of an image is known as its *radiometric resolution* (Mather 2004). This determines the range of DN values and affects the contrast of an image and the ability to detect subtle variations in target objects (Gao 2009). *Temporal resolution*, also known as *revisit time*, is the time elapsed between successive measurements of the same ground area (Mather 2004; Gao 2009). While not always a critical consideration, a fine temporal resolution (i.e. short intervals between consecutive scans of the same area) is desirable for monitoring dynamic phenomena (Campbell 2006).

As mentioned, the characteristics of remotely sensed imagery are critical to dealing with the problem at hand. For the challenge of land cover classification, the characteristics of different sensors must be carefully considered in conjunction with the classification scheme, the temporal requirements of the project and the availability of resources (time, money, computational power). The selection of an appropriate method of classification depends heavily on such variables. The following section briefly discusses various classification methods that have been applied in the field of RS.

2.2   IMAGE CLASSIFICATION

Digital image classification, the process of assigning image pixels or objects to informational classes (Campbell 2006), consists of two stages: The recognition of the categories of interest (the informational classes) and the labelling of the entities through the use of a specific classification algorithm, or *classifier* (Mather 2004). Classifiers are useful tools for extracting valuable information from remotely sensed images. Consequently, numerous classifiers, each with particular strengths and weaknesses, have been applied for a wide range of RS problems (Lu & Weng 2007). The two traditional approaches to image classification, *unsupervised* and *supervised* classification, as well as the *rule-based* approach, are discussed in this section. *Pixel-based* classification is also contrasted to the more recent geographic object-based image analysis (GEOBIA) approach.

2.2.1 Unsupervised approach

Unsupervised classification involves the process of *clustering;* the identification of natural groups within a feature set. Clustering algorithms identify and label the number of distinct classes according to the nature of the data in the feature set (Campbell 2006; Mather 2004). It is the user's task to assign these natural groupings, or *spectral classes,* to appropriate informational classes by making use of some form of reference data (Lillesand, Kiefer & Chipman 2008).

Unsupervised classifiers are useful when prior information on the study area is unavailable, and they perform best when the desired informational classes are spectrally distinct and can be easily clustered (Gao 2009). Because minimum user intervention is required, unsupervised classification is relatively easy and fast to implement (Gao 2009). However, it is not uncommon for the spectral classes resulting from clustering not to correspond to the informational classes of interest (Campbell 2006; Gao 2009; Stephenson 2010). As a result, unsupervised classifiers are often considered less useful, and used less, than supervised classifiers (Gao 2009; Stephenson & Van Niekerk 2009). Yet, popular unsupervised classifiers such as ISODATA, k-means and the modified k-means algorithms have been applied to a variety of RS classification problems (Calvo, Ciraolo & Loggia 2003; Duda & Canty 2002; Lang *et al.* 2008; Nolin & Payne 2007; Smith *et al.* 2002; Tapia, Stein & Bijker 2005)

2.2.2 Supervised approach

Supervised classification requires training samples of known identity to be supplied prior to classification. Supervised classifiers use the statistical information contained in a training set to predict the class membership of the remaining image samples. The approach offers greater control by forcing the user to determine the informational classes prior to classification. This allows the categorization of classes to be tailored to the needs of a project and also to the nature of the data (Campbell 2006). Compared to unsupervised classifiers, supervised classifiers are robuster, more suitable for complex classification problems (Gao 2009) and are more commonly applied in the field of remote sensing. However, they do have drawbacks, the most significant of which is their dependence on a training set. Successful classification requires enough accurate, well distributed and representative training samples (Campbell 2006; Hubert-Moy *et al.* 2001; Lillesand, Kiefer & Chipman 2008; Lu & Weng 2007; Stephenson & Van Niekerk 2009); conditions that cannot always be met due to limited resources (Campbell 2006).

When opting to use a supervised classifier, its categorization as either *parametric* or *non-parametric* is important. Parametric, or *statistical,* classifiers assume that the data follows a known distribution. The estimation of certain statistical parameters essential to the classification process relies on this assumption (Jain, Duin & Mao 2000). In contrast, non-parametric classifiers make no assumptions about the distribution of the data and they do not rely on the estimation of parameters. This is a noteworthy advantage as distribution assumptions often do not hold for remotely sensed data.

The *parallelepiped* and *minimum distance* classifiers offer the advantages of simplicity and speed, although the more complex ML classifier surpasses these methods regarding reliability and accuracy (Chuvieco & Huete 2010). The most commonly used supervised classifier in remote sensing is ML (Albert 2002; Stephenson 2010; Waske *et al.* 2009), and it assumes that the data is normally distributed. ML relies on estimates of the mean vector and the variance–covariance matrix which, in turn, are used to calculate class probabilities for unknown samples. A sample is assigned to the class for which the highest probability is calculated. ML produces high classification accuracies for RS applications (Albert 2002; Gao 2009; Pal & Mather 2003; Szuster, Chen & Borger 2011; Waske *et al.* 2009) and the classifier is often used as a benchmark when evaluating other classification techniques (Stephenson 2010). ML is, however, highly sensitive to the quality of training data (Campbell 2006) and the classifier's intrinsic assumption that data is normally distributed is often untenable. These limitations may lead to poor performance in RS applications.

A simple distance-based, non-parametric technique often employed in RS applications and for benchmarking is the *k*-nearest neighbour (*k*NN) classifier. The *k*NN rule assigns an unknown sample to the class that occurs most frequently among its *k*-nearest neighbours (Campbell 2006; Cover & Hart 1967). In its simplest form, referred to as nearest neighbour (NN) classification, the variable *k* is set to one and an unknown sample is assigned to the class of the closest training sample in feature space. *k*NN and NN classifiers offer simplicity and provide a practical advantage over statistical classifiers for use when data that is not normally distributed (Campbell 2006).

Artificial neural networks (ANNs) have been in use as alternative non-parametric methods for RS image classification since the early 1990s (Chen & Ho 2008; Mas & Flores 2008). They are complex classification algorithms designed to simulate the human learning process. An ANN consists of an *input layer* which consists of the source data (i.e. spectral information), an *output layer* which consists of the desired output classes and one or more *hidden layers.* ANN establishes an association between the input and output layers by determining weights

in the hidden layers. Repeated associations between classes and the digital values contained in the training data, strengthen the weights in the hidden layers. A fully trained ANN is able to assign correct labels to input data based on the weights in the hidden layers (Campbell 2006). ANNs typically produce higher classification accuracies compared to traditional statistical classifiers, they can handle noisy data well, and their non-parametric nature allows the effective incorporation of multisource and ancillary data (Mas & Flores 2008; Kavzoglu & Mather 2003). Consequently, ANNs have become a widely researched topic in RS (particularly for land cover classification). The extent of this research is reviewed by Mas & Flores (2008). ANNs have, however, been criticized for their complex nature, long training times, the trial-and-error-based design of the network architecture and their variable results (Kavzoglu & Mather 2003; Mas & Flores 2008; Mather 2004; Stephenson 2010).

More recently, a number of RS studies have concentrated on the application of support vector machines (SVMs) (Mountrakis, Im & Ogole 2011). SVMs, introduced by Vapnik (1995), are theoretically well-founded supervised classifiers based on statistical learning theory and structural risk minimization (Roli & Fumera 2000). Developed as a binary classifier, SVM relies on identifying the optimal separating hyperplane (OSH) as a decision boundary to separate two classes. The OSH ensures a maximum margin between the hyperplane and the closest training samples of each class (termed *support vectors*) and it is calculated by standard quadratic programming optimization techniques (Pal & Mather 2005). The support vectors are the only training samples used in this calculation. To accommodate data that is not linearly separable, SVM is extended by introducing slack variables and applying a kernel function to solve the optimization problem in higher-dimensional space (Mountrakis, Im & Ogole 2011) (see Section 3.3.3). Kernel functions need to fulfil Mercer's theorem and linear, polynomial, radial basis function (RBF) and sigmoid kernels are often used (Tzotsos & Argialas 2008). Methods such as one-against-one, one-against-all and direct acyclic graph are used to extend SVM for multiclass classification problems (Mountrakis, Im & Ogole 2011).

Cited advantages of SVMs include superior classification accuracies, good performance with limited training samples and robustness to large input dimensionalities (Foody & Mathur 2004b; Li *et al.* 2010; Lizarazo 2008; Mountrakis, Im & Ogole 2011; Pal & Mather 2005). However, the selection of an appropriate kernel function and the assignment of kernel parameters is problem specific and may significantly affect classification results (Mountrakis, Im & Ogole 2011). Because the promise shown by SVMs for land cover classification was a key motivator for conducting this research, a separate section (2.3) is devoted to elaborating on the potential of these supervised classifiers.

Another often used supervised approach, decision tree (DT) classifiers, is discussed in the next subsection on rule-based classification because the generation of a rule set, and the nature in which data is classified from such a rule set, distinguishes DT classifiers from the supervised classifiers discussed above.

2.2.3   Rule-based approach

Whereas traditional classifiers consider all available features simultaneously to make a single membership decision for each unknown sample (Pal & Mather 2003), *rule-based* classifiers apply a chain of informed rules (a *rule set*) in a structured or layered approach (Mather 2004). An advantage of this approach is that these decision rules can be applied to a wide variety of input data so allowing the efficient incorporation of ancillary data (Chuvieco & Huete 2010). A distinction is made between classifiers requiring manual creation of rule sets by an experienced analyst (expert systems) and supervised algorithms that extract decision rules automatically from training samples. These approaches are briefly discussed here.

DT classifiers are versatile tools for supervised rule-based classification. These algorithms recursively split a training set into homogeneous subdivisions based on some statistical test (Chuvieco & Huete 2010; Friedl & Brodley 1997). From each such split, logical rules, capable of emulating the statistical divisions, are inferred resulting in a hierarchical rule set capable of image classification. The generated rule set offers increased interpretability (the most discriminating features can be easily identified through inspection of the rules) and flexibility (rules may be manually refined) compared to traditional classifiers (Brown de Colstoun *et al.* 2003; Friedl & Brodley 1997; Hansen, Dubaya & Defries 1996). However, the algorithm is still strictly supervised and successful classification requires sufficient training data. Several comparative studies have shown that DTs produce classification accuracies that are superior to those of ML and comparable to those of ANNs (Brown de Colstoun *et* al. 2003; Friedl & Brodley 1997; Pal & Mather 2003). Pal & Mather (2003) have noted that DTs are not recommendable for the classification of high-dimensional data sets as both ML and ANNs achieve superior results when the size of the feature set is increased.

An expert system employs expert knowledge to emulate the decision-making of a human expert for solving a specific problem (Skidmore *et al.* 1996). When considering the problem of RS image classification, one or more human experts develop a rule set capable of extracting predetermined target classes from the available data layers. Expert systems have the advantage of not requiring the prior definition of training samples and their flexibility makes them useful for land cover mapping (Aitkenhead & Aalders 2011). However, the

development of an effective rule set is time-consuming (Liu, Skidmore & Van Oosten 2002; Tseng *et al.* 2008). Expert rule sets are often employed in object-based image classification (discussed in the following section) by using the eCognition software package which has resulted in high classification accuracies being achieved (Bauer & Steinnocher 2001; Chen *et al.* 2009; Laliberte *et al.* 2006; Mallinis *et al.* 2008; Tansey *et al.* 2009).

2.2.4   Object-based vs pixel-based classification

Traditionally, a per-pixel approach has been adopted for RS image classification despite the use of pixels as units of analysis often receiving criticism (Blaschke & Strobl 2001; Cracknell 1998; Fisher 1997). For example, a pixel is not likely to represent a real world geographical object (Blaschke & Lang 2006) and per-pixel classifiers are limited in their use of spatial concepts (Blaschke & Strobl 2001). Pixel-based classification can be effective if the spatial resolution is similar to the land cover features of interest (Blaschke 2010; Fourie 2011) but problems arise when this is not true. *Mixed pixels* occur when boundaries between mapping units occupy a single pixel or the features of interest exist at a sub-pixel level (Fisher 1997). The mixed pixel effect lowers classification accuracy (Campbell 2006; Fourie 2011; Shaban & Dikshit 2001). More sub-class elements may become detectable at finer resolutions, implying high within-class spectral variance which results in lower classification accuracies (Shaban & Dikshit 2001). Misclassifications caused by these spectral variances may lead to the well-known salt-and-pepper effect with homogeneous regions containing some scattered, incorrectly classified pixels (Blaschke *et al.* 2000)

The concept of GEOBIA gained widespread interest in the fields of remote sensing and GIS around 2000, although it builds on concepts used in image analysis since the 1970s (Blaschke 2010; Blaschke, Lang & Hay 2008). GEOBIA methods do not consider individual pixels for analytical purposes, rather *objects* that comprise several pixels. A segmentation algorithm subdivides an image into homogeneous interlocking regions (the objects) based on the spectral properties of the underlying image and some user-defined constraints (Campbell 2006). The partitioning of an image into meaningful geographical objects is akin to human interpretation of landscapes (Addink, De Jong & Pebesma 2007; Hay & Castilla 2006, 2008). GEOBIA classification has several advantages over pixel-based approaches, for example the use of objects reduces within-class spectral variance and typically solves the salt-and-pepper problem (Liu & Xia 2010). Consequently, GEOBIA is well suited for the classification of high- and very-high-resolution imagery (Bauer & Steinocher 2001; Laliberte *et al.* 2006; Mallinis *et al.* 2008; Tansey *et al.* 2009). Also, additional spectral, spatial, textural and

contextual features are contained in, or easily derived from, image objects and ancillary data sources (Hay & Castilla 2006, 2008; Liu & Xia 2010). Such additional variables can significantly improve classification accuracies (Campbell 2006; Heinl *et al.* 2009).

The outstanding drawback of GEOBIA is its reliance on segmentation which, as Hay & Castilla (2008: 84) put it, is an "*ill-posed* problem" having "no unique solution". Segmentation quality does affect classification accuracies (Addink, De Jong & Pebesma 2007; Kim Madden & Warner 2009). Whether a segmentation is "good" is difficult to determine (Hay & Castilla 2006, 2008) and the quality depends on the scale of the classification problem (Benz *et al.* 2004; Liu & Xia 2010). Obtaining an appropriate segmentation relies heavily on the analyst's knowledge and often involves a time consuming process of trial-and-error tweaking of segmentation parameters (Fourie, Van Niekerk & Mucina 2011, 2012).

As mentioned in Section 2.2.3, expert rule-based classifiers are often applied for GEOBIA. The additional inherent features of image objects are convenient for developing rule sets (Stephenson 2010). Supervised methods have also been successfully applied to object-based classification (Berberoglu *et al.* 2000; Li *et al.* 2008; Lizarazo 2008; Mansor, Hong & Shariff 2002; Tzotsos & Argialas 2008). Results of comparisons of supervised methods for per-pixel and object-based classification is inconclusive. Duro, Franklin & Dubé (2012) compared the accuracies of pixel-based and object-based classifications of three classifiers (DT, ML and random forest) using Landsat enhanced thematic mapper plus (ETM+) imagery. While the object-based classifiers produced visually appealing results when compared to their pixel-based counterparts, improvements in overall accuracies were not statistically significant. Conversely, Li *et al.* (2008) showed object-based SVM classification of a polarimetric synthetic aperture radar (PolSAR) image to be about 40% more accurate than a pixel-based SVM classification using the same data. Clearly, the nature of the data and the chosen classification scheme influence the suitability of adopting either a pixel-based or an object-based approach for supervised image classification.

The inherent differences between GEOBIA and traditional pixel-based analysis can significantly influence supervised classification. The nature of the training data and the use of additional features – two key aspects that are affected by adopting an object-based approach – are discussed in the following subsections.

2.2.4.1 Training data

The characteristics of a training set influence the accuracy of supervised classification (Campbell 2006; Lu & Weng 2007). Recall (Section 2.2.2) that supervised classifiers require an adequate number of accurate, well-distributed and representative training samples. For traditional pixel-based classification, Campbell (2006) suggests using at least 100 training pixels per class while Mather (2004) recommends a minimum of $30p$ pixels, where p is the number of features. It is important that homogeneous groups of pixels consisting of about 10 to 40 pixels each be selected to obtain reliable estimates of the spectral characteristics of each class (Campbell 2006). Deviations from such recommendations are often necessary because of limited resources (Mather 2004). New ways to achieve high accuracies by using fewer training samples can improve the cost-effectiveness of mapping land cover from large volumes of imagery. While advanced non-parametric classifiers, such as ANNs and SVMs, are less sensitive to the size of training sets compared to traditional statistical classifiers (Mas & Flores 2008; Mountrakis, Im & Ogole 2011), the nature of the training set may have a greater effect on classification accuracies than that of the selected classifier (Foody & Mathur 2004a).

A GEOBIA approach significantly changes the nature of the data being analysed, and consequently also the nature of the training set. When selecting a homogeneous group of pixels for pixel-based classification, each pixel within such a group is regarded as an indivdual training sample by the classifier. In GEOBIA pixels are grouped into homogeneous objects prior to an analysis and only the mean values of such objects are used. This effectively reduces the number of samples available to the classifier (Tzotsos & Argialas 2008). It is generally unfeasible in GEOBIA to select a sufficient number of samples according to the above recommendations by Campbell (2006) and Mather (2004). Classification methods that perform well under conditions of limited training set sizes are therefore crucial in object-based supervised classification.

2.2.4.2 Additional features

In addition to the original spectral bands, variables such as vegetation indices, transformed images, textural information, contextual information and ancillary data are often incorporated into, and may significantly influence the accuracy of, RS image classification (Heinl *et al.* 2009; Lu & Weng 2007). Heinl *et al.* (2009) found that the addition of topographic measures, the normalized difference vegetation index (NDVI) and texture measures resulted in greater classification accuracies for ML, ANN and discriminant analysis (DA) classifiers. Berberoglu

et al. (2000, 2007) have reported that the incorporation of textural features leads to increased classifier performance.

Compared to traditional pixel-based analysis, GEOBIA incorporates such additional features more effectively (Hay & Castilla 2006, 2008). Recall that additional spectral and spatial features are inherent to image objects. Consequently GEOBIA has greater potential for using additional discriminating features for image classification and it follows that it is important to consider the dimensionality of feature space when applying supervised classification. The Hughes effect (Hughes 1968) limits the performance of some classifiers when a large number of features is used. The Hughes effect is the phenomenon that classification accuracy decreases after the number of features is increased beyond a certain point, *unless* the number of samples is increased proportionally (Chen & Ho 2008). This problem is more likely to be encountered when working with a limited training set size, as is typically the case with GEOBIA. Consequently, feature selection methods are often employed to determine optimal features for GEOBIA classification. The Bhattacharyya distance, the Jeffreys-Matusita (JM) distance, genetic algorithms, feature space optimization (FSO) and classification tree analysis (CTA) are all methods that have been used to select optimal features for object-based classification (Addink *et al.* 2010; Carleer & Wolff 2006; Chubey, Franklin & Wulder 2006; Herold, Liu & Clarke 2003; Laliberte, Browning & Rango 2010, 2012; Laliberte, Fredrickson & Rango 2007; Marpu *et al.* 2008; Van Coillie, Verbeke & De Wulf 2007; Yu *et al.* 2006; Zhang, Feng & Jiang 2010). In a comparison between the JM distance, FSO and CTA feature selection methods for object-based classification, Laliberte, Browning & Rango (2010, 2012) concluded that CTA was the best suited due to its ability to efficiently rank and reduce features.

SVMs perform well with limited training sets and are less susceptible to the Hughes effect (Mountrakis, Im & Ogole 2011). SVMs' non-parametric nature also promotes the integration of various data sources. As such, it assumed that SVMs are well suited for object-based supervised classification using a large number features. The next section reviews a number of case studies in which the performance of SVMs were evaluated for land cover classification.

2.3   THE POTENTIAL OF SVM FOR LAND COVER CLASSIFICATION

SVM, a relatively new supervised machine learning technique (Kotsiantis 2007), is receiving keen attention from the RS community for its ability to generalize well with small training sets and its robustness for large input dimensionalities (Foody & Mathur 2004b; Li *et al.* 2010; Lizarazo 2008; Mountrakis, Im & Ogole 2011; Pal & Mather 2005). SVM-related

research on remote sensing problems has proliferated in recent years (Mountrakis, Im & Ogole 2011) and SVM's potential for RS image classification has been the subject of a number of comparative studies.

2.3.1   Pixel-based comparative studies

Gualtieri & Cromp (1998) applied SVM for hyperspectral image classification using an airborne visible/infrared imaging spectrometer (AVIRIS) scene. Their seminal study on SVM for RS image classification found that the overall accuracy produced by the SVM classifier was superior to those of various classifiers tested by Tadjudin & Landgrebe (1998) using the same data. They noted that, despite the high feature dimensionality of the data, SVM did not suffer from the Hughes effect. Hermes *et al.* (1999) tested three SVM variants (regular SVM, probabilistic SVM and a probabilistic SVM with iterated conditional modes (ICM)) for classifying Landsat thematic mapper (TM) imagery. The SVM approaches outperformed three other classifiers (NN, ML and Gaussian mixture model), the SVM with ICM achieving the best results. Subsequently a number of studies that compare SVM with more commonly used RS methods have emerged for both multispectral and hyperspectral image classification.

Huang, Davis & Townshend (2002) compared SVM, ML, DT and ANN classifiers using Landsat TM data. They included a test in which the training set size was investigated and found that SVMs outperformed the other classifiers in most cases. Using TM imagery, Keuchel *et al*. (2003) did a comparative study which adopted a 10-class classification scheme. SVM yielded the highest accuracy (93.3%) compared to ML (90.2%) and ICM (88.5%). Foody & Mathur (2004a) similarly reported higher accuracies by SVM in a classification of a Deadalus 1268 airborne thematic mapper scene compared to DT and ANN and they noted that the size of the training set significantly influenced the performance of each classifier. Candade & Dixon (2004) compared radial basis function (RBF), linear and polynomial kernel SVMs with ANN, the polynomial kernel SVM producing the most accurate classification. Dixon & Candade (2008) reported that polynomial SVM achieved a significantly higher overall accuracy (79.2%) than ML (50.6%) and slightly better than ANN (78.4%) in the classification of eight land use classes using Landsat 5 TM data and a fixed training set. SVM also has considerable potential for effective multisource classification compared to ANN, ML and DT as shown by Watanachaturaporn, Arora & Varshney (2008). Kavzoglu & Colkesen (2009) applied polynomial and RBF kernel SVMs and compared the results with a ML classification. They adopted a seven-class classification scheme and the classifiers were applied to Landsat ETM+ and to Advanced Spaceborne Thermal Emission Reflection

Radiometer (ASTER) imagery respectively. The SVM approaches outperformed the ML classifier (by approximately 4%) for all data sets, the RBF kernel producing the best results. The classifications from the Terra ASTER image were more accurate than those obtained from the Landsat ETM+ scene for each classifier. ASTER data was also used by Szuster, Chen & Borger (2011) to compare SVM, ML and ANN for coastal land cover and land use change. The classifiers achieved similar results regarding overall accuracy, but SVM separated spectrally similar classes better.

Pal & Mather (2005) conducted two classification experiments using multispectral (Landsat-7 ETM+) and hyperspectral digital airborne imaging system (DAIS) data respectively. The first experiment tested the performance of one-against-all and one-against-one SVM implementations using two software packages. The one-against-one implementation using Libsvm (Chang & Lin 2011) achieved the highest accuracy (87.9%) and outperformed ML (82.9%) and ANN (85.1%) classifiers. The second experiment compared the performance of SVM, ML and ANN classifiers using an increasing number of features (DAIS spectral bands). Classification accuracies generally increased as more features were added although slight reductions in accuracy occurred with all three classifiers when the number of features exceeded 50. Oommen *et al.* (2008) used multispectral (Landsat-7 ETM) and hyperspectral (Hyperion) data to compare SVM and ML classifiers. Their experiments, based on the size of training sets and number of band combinations, concluded that SVM gave higher accuracies, was robust and did not suffer from dimensionality issues.

Several comparative studies have concentrated on SVMs in hyperspectral image classification (Camps-Valls & Bruzzone 2005; Camps-Valls *et al.* 2004, 2006; Melgani & Bruzzone 2004; Pal & Mather 2004). The findings show that SVMs generally produce higher accuracies than classifiers such as ANN, ML, *k*NN and DT and SVMs are only slightly affected, if not unaffected, by the input space dimensionality. SVMs are therefore well suited for problems where the number of input features is high and feature selection is not a viable option, although feature selection is still recommendable (Camps-Valls *et al.* 2004). Other advantages of SVMs identified by these studies include robustness to noisy data, lower computational cost and their simplicity when compared to that of ANN methods.

2.3.2   Object-based SVM in remote sensing

From the previous section it is clear that there is much interest in SVMs for RS image classification in the pixel-based paradigm. However, similar studies concerning GEOBIA are sparse. Tzotsos & Argialas (2008) found in an object-based environment that SVM

outperformed (by 5%) NN classifiers for mapping land cover from Landsat TM imagery. Li *et al.* (2010) report similar accuracy gains for SVM over NN in the object-based classification of high-resolution QuickBird data.

Object-based SVMs have also been applied in other remote sensing studies: Lizarazo (2008) favourably compared object-based SVM to pixel based SVM; Li *et al*. (2008) found object-based SVM to be far more accurate (91.6 %) than a pixel-based SVM (51.4 %) for crop classification using PolSAR data; Meng & Peng (2009) used a fuzzy SVM approach for object-based building extraction from QuickBird imagery; Wu *et al.* (2009) applied an object-based SVM classification in their evaluation of the maximum mutual information feature selection method; Liu & Xia (2010) used SVM to investigate the impact of over- and under-segmentation on classification accuracies; Tzotsos, Karantzalos & Argialas (2011) applied object-based SVM classification as the final step in their advanced GEOBIA approach; and Duro, Franklin & Dubé (2012) directly compared pixel-based and object-based implementations of SVM, DT and DA classifiers and found that the results were not significantly affected by the choice of approach. No studies exist where the performance of SVM has been evaluated comparatively for object-based classification with varying sizes of training sets and feature sets.

SVM's ability to perform well under conditions of small training set sizes and high feature dimensionalities has generally been lauded in many pixel-based comparative studies (see Section 2.3.1). Although these advantages should qualify SVM as well suited for GEOBIA, clearly more research is necessary to evaluate SVM's potential as an object-based classifier. Consequently, this research undertook two experiments to compare object-based SVM, NN and ML classifications. The first that investigates the impact of the training set size on the classifiers' performance is discussed in the next chapter and the second experiment, focusing on the effect of feature dimensionality, is documented in Chapter 4.

CHAPTER 3: IMPACT OF TRAINING SET SIZE ON OBJECT-BASED LAND COVER CLASSIFICATION: A COMPARISON OF THREE CLASSIFIERS[*]

3.1 ABSTRACT

Supervised classifiers are commonly employed in remote sensing to extract land cover information, but various factors affect their accuracy. The number of available training samples, in particular, is known to have a significant impact on classification accuracies. Obtaining a sufficient number of samples is, however, not always practical. The support vector machine (SVM) is a supervised classifier known to perform well with limited training samples and has been compared favourably to other classifiers for various problems in pixel-based land cover classification. Very little research on training-sample size and classifier performance has been done in a geographical object-based image analysis (GEOBIA) environment. This paper compares the performance of SVM, nearest neighbour (NN) and maximum likelihood (ML) classifiers in a GEOBIA environment, with a focus on the influence of training set size. Training set sizes ranging from 4–20 per land cover class were tested. Classification tree analysis (CTA) was used for feature selection. The results indicate that the performance of all the classifiers improved significantly as the size of the training set increased. The ML classifier performed poorly when few (<10 per class) training samples were used and the NN classifier performed poorly compared to SVM throughout the experiment. SVM was the superior classifier for all training set sizes although ML achieved competitive results for sets of 12 or more training areas per class.

3.2 INTRODUCTION

Detailed, accurate and up-to-date land cover information is critical for environmental and socio-economic research (Heinl *et al*. 2009; Lu & Weng 2007). A large number of satellite platforms are operational that have the capability to provide remotely sensed imagery at various spatial and temporal scales (Foody 2002). This abundance of available data offers great potential for generating frequently updated thematic maps as remotely sensed images cover large areas, are acquired at regular intervals and are less costly than traditional ground-survey methods (Foody 2009; Gao 2009; Pal & Mather 2004; Szuster, Chen & Borger 2011).

---

[*] This chapter was submitted for publication to the International Journal of Remote Sensing and consequently conforms to the prescribed structure of that journal.

Current image-processing techniques are, however, limited in their ability to extract accurate land cover features automatically (Baraldi *et al.* 2010). Many factors also affect the accuracy of image classification (Lu & Weng 2007) and the quality of many land cover maps is often perceived as being insufficient for operational use (Foody 2002).

Supervised classification, an approach commonly used for the classification of remote sensing images, requires samples of known identity (training samples) to construct a model capable of classifying unknown samples. Apart from selecting a suitable classifier, the number and quality of training samples are key to a successful classification (Hubert-Moy *et al.* 2001; Lillesand, Kiefer & Chipman 2008; Lu & Weng 2007). A sufficient number of training samples is generally required to perform a successful classification and the samples need to be well distributed and sufficiently representative of the land cover classes being evaluated (Campbell 2006; Gao 2009; Mather 2004; Lu & Weng 2007). In remote sensing applications, the availability of labelled training samples is often limited (Gehler & Schölkopf 2009; Mountrakis, Im & Ogole 2011) as their collection is time-consuming, expensive and tedious, often requiring the study of maps and aerial photographs and carrying out field visits (Campbell 2006).

Support vector machines (SVM) have been shown to improve the reliability and accuracy of supervised classifications (Oommen *et al.* 2008). SVM are known for their good generalizing ability even when few training samples are available and it has been suggested that SVM produce superior results compared to other statistical classifiers when fewer training samples are available (Foody & Mathur 2004b; Li *et al.* 2010; Lizarazo 2008; Mountrakis, Im & Ogole 2011; Pal & Mather 2005).

The introduction of SVM to remote sensing has led to a number of comparative studies involving SVM and other classifiers of land cover (Camps-Valls & Bruzzone 2005; Camps-Valls *et al.* 2004; Dixon & Candade 2008; Foody & Mathur 2004a; Gualtieri & Cromp 1998; Huang, Davis & Townshend 2002; Kavzoglu & Colkesen 2009; Keuchel *et al.* 2003; Melgani & Bruzzone 2002, 2004; Mercier & Lennon 2003; Oommen *et al.* 2008; Pal & Mather 2004, 2005; Szuster, Chen & Borger 2011; Tzotsos & Argialas 2008). Although the results of such studies depend on the data and classification scheme used in each case, it was generally found that SVM produced either superior or equivalent classification accuracies when compared with methods such as maximum likelihood (ML), nearest neighbour (NN), artificial neural networks (ANN) and decision trees.

Most of the comparative studies published to date were carried out using a traditional pixel-based classification approach. Geographical object-based image analysis (GEOBIA) has emerged as an alternative approach to pixel-based image processing (Blaschke 2010; Blaschke & Lang 2006; Hay & Castilla 2006, 2008). GEOBIA involves a segmentation step during which image pixels are grouped into homogeneous interlocking regions as determined by a specific segmentation algorithm (Campbell 2006). All subsequent analyses, such as classification, are based on the various attributes of these image objects. The grouping of multiple pixels into single objects often means that fewer training samples are available to the classifier when supervised classification is performed. A classifier's ability to perform well with a limited number of training samples is consequently of great importance for supervised GEOBIA. When applied in an object-based environment, Tzotsos & Argialas (2008) found that SVM outperformed NN classifiers for mapping land cover from Landsat TM imagery. Although object-based SVM has been implemented in other remote sensing studies (Duro, Franklin & Dubé 2012; Li *et al.* 2008, 2010; Liu & Xia 2010; Lizarazo 2008; Meng & Peng 2009; Tzotsos, Karantzalos & Argialas 2011; Wu *et al.* 2009), none have investigated the comparative performance of SVM under conditions of limited training set sizes. Given the significant differences between pixel-based and object-based classification and the suitability of GEOBIA for classifying high-resolution imagery (Blaschke 2010), a comparative analysis of SVM and other supervised classifiers will provide insights into their suitability for object-based supervised classification. In addition, an investigation of the influence of training set size on classification accuracy may shed light on the potential of supervised object-based image analysis for the cost-effective processing of large volumes of imagery.

The aim of this paper is to investigate the performance of object-based SVM for land cover classification compared to NN and ML classifiers, with a focus on the number of training samples used. The NN and ML classifiers were chosen for benchmarking since the latter is regarded as the most commonly used supervised classification method in remote sensing (Albert 2002, Stephenson 2010; Waske *et al.* 2009) and NN is the supervised method most commonly employed for object-based classification (Campbell 2006).

The rest of the paper is structured into four sections, the first of which overviews the NN, ML and SVM classifiers. This is followed by descriptions of the data used, the experimental design and the development of the software that automated the assessments. The results are discussed in Section 4, and the final section summarizes the findings and makes suggestions for further research.

3.3 OVERVIEW OF THE TESTED SUPERVISED CLASSIFIERS

Supervised classifiers are typically categorized as either parametric or non-parametric. Parametric classifiers assume that the data follows a known distribution and they involve the estimation of certain statistical parameters from training data (Jain, Duin & Mao 2000). In contrast, non-parametric classifiers make no assumptions about the distribution of the data and do not require the calculation of statistical parameters. Supervised classifiers can also be categorized as density estimation (DE) or boundary estimation (BE) classifiers. DE classifiers obtain decision boundaries indirectly from density functions, while BE obtain decision boundaries directly from training data by solving some optimization problem (Fourie 2011; Jain, Duin & Mao 2000). The three classifiers applied in this paper have different properties regarding these definitions.

3.3.1 Maximum likelihood

ML is the most commonly used supervised classification method in remote sensing and is derived from the Bayesian decision rule for the case where all classes have equal prior probabilities (Albert 2002, Waske *et al.* 2009). It is a parametric, DE method that calculates the mean vector and the variance-covariance matrix from the training data and uses it to construct an estimated distribution for each class. The probability (*P(x)*) of an unknown sample, represented by a vector *x* of *p* features, belonging to a certain class (*i*) is calculated from this distribution by

$$P(\boldsymbol{x}) = 2\pi^{-0.5p}|\boldsymbol{S}_i|^{-0.5}\mathrm{e}^{-0.5y} \qquad\qquad (3.1)$$

where $\boldsymbol{S}_i$ is the variance−covariance matrix for class *i* and *y* is the Mahalanobis distance. The sample is assigned to the class for which the highest probability is calculated. ML has the advantage of accounting for variability within each class through use of the covariance matrix. Although ML has been shown to produce high classification accuracies (Albert 2002; Gao 2009; Pal & Mather 2003; Stephenson 2010; Szuster, Chen & Borger 2011; Waske *et al.* 2009), its assumption that all input data is normally distributed (Albert 2002; Oommen *et al.* 2008) and its sensitivity to training-data quality (Campbell 2006) have been cited as reasons for the poor performance in certain remote sensing applications (Fourie 2011).

3.3.2 Nearest neighbour

NN is a simple distance-based, non-parametric decision procedure and is also a DE method. The NN decision rule is a special case of the *k*-nearest neighbour (*k*NN) rule with the free

parameter k set to one. The kNN rule assigns an unknown sample to the class that occurs most frequently among its k-nearest neighbours (Campbell 2006; Cover & Hart 1967). NN simply assigns a sample to the class of the training sample it is closest to in multidimensional feature space. In other words, if the training instances are represented by vectors x_1, $x_{2, \ldots}$, x_n and the Euclidean distance between two vectors is $d(x_1, x_2)$, the nearest neighbour x_{nn} of an unknown sample x can be defined as:

$$\mathbf{x}_{nn\,=\,}\mathbf{x}_j \quad \text{where} \quad \mathbf{x}_j \in \{\ \mathbf{x}_1, \mathbf{x}_{2, \ldots}, \mathbf{x}_n\} \tag{3.2}$$

if

$$d(x_j,\, x) = \min d(x_i, x) \quad \text{for} \quad i = 1, \ldots, n \tag{3.3}$$

The unknown sample is classified to the same class as x_{nn}. The NN and kNN rules are often employed in remote sensing as their simplicity and effectiveness in classifying heterogeneous classes often provide a practical advantage (Campbell 2006). In this paper, the NN rule was employed as small training sets were used during the experiment. Using larger values of k is more appropriate when the number of samples is large (Cover & Hart 1967).

3.3.3 Support vector machines

SVM is a non-parametric, BE supervised classification technique that is based on statistical learning theory (Vapnik 1995). In its simplest form, SVM is a binary classifier capable of separating two linearly separable classes. SVM relies on identifying the optimal separating hyperplane between two sets of classes in a multidimensional feature space. Once identified, the hyperplane is used as a decision surface to classify unknown samples. As shown in Figure 3.1(a), many hyperplanes may exist that are capable of separating two classes, but there can only be one optimal hyperplane that maximizes the margin between the hyperplane and the closest training samples (called the support vectors) (see Figure 3.1(b)). The support vectors lie on two hyperplanes parallel to the optimal hyperplane (Oommen *et al.* 2008) and can be defined as:

$$w\, \mathbf{x}_i + b = \pm\, 1 \tag{3.4}$$

where $\mathbf{x}$ is a point on the hyperplane, w is a vector normal to the hyperplane and b is the bias. Rescaling the parameters w and b allows the margin between these support planes to be

expressed as 2/||*w*// (Oommen *et al.* 2008, Fourie 2011). The optimization problem for maximizing this margin is given by:

Minimize $\quad \frac{1}{2} \|w\|^2$ (3.5)

subject to the constraints

$y_i(w\ \mathbf{x}_i + b) - 1 \geq 0 \quad$ for $\ i = 1, 2, ... N$ (3.6)

where $y_i \in \{1,-1\}$ is the class label and *N* is the number of training samples. This optimization problem can be regarded as the basic definition of SVM for the case of two linearly separable classes. Target classes are, however, often spectrally similar and the training set may include outliers that reduce the likelihood that classes will be linearly separable for the majority of remote sensing classification problems.
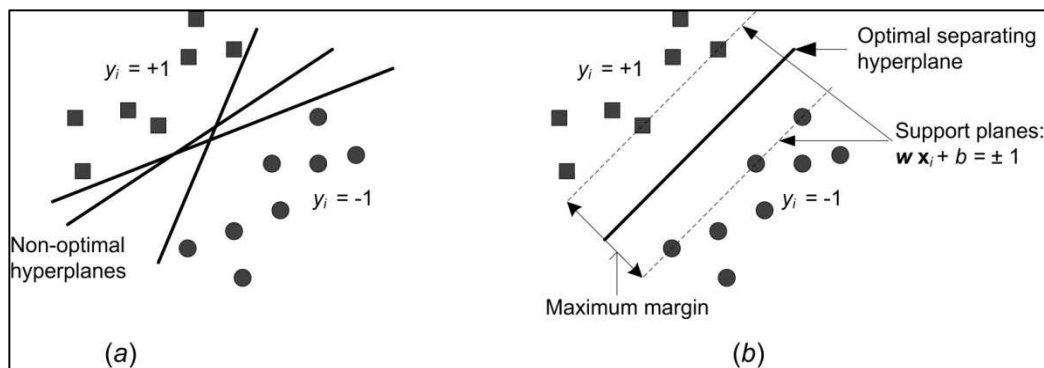
Figure 3.1: Conceptual view of SVM showing how (a) multiple hyperplanes may separate two classes and how (b) SVM relies on identifying the optimal separating hyperplane.

SVM can be extended for the linearly inseparable case through the introduction of slack variables and by applying a kernel function (Mountrakis, Im & Ogole 2011). Equations (3.5) and (3.6) are then reformulated as:

Minimize $\quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$ (3.7)

subject to the constraints

$y_i(w\ \Phi(x_i) + b) - 1 \geq 1 - \ \xi_i \ $ and $\xi_i \geq 0, i = 1, 2, ..., N$ (3.8)

where the ξ_i's are the slack variables, C is a user-defined penalty parameter and Φ is a mapping function used to map the data (x_i) to a higher-dimensional feature space and where a linear hyperplane can be fitted that is non-linear in the original input space (Waske *et al.* 2009). To solve this optimization problem, it is typically reformulated as a Lagrange function from which the following decision rule can be obtained:

$$f(\boldsymbol{x}) = sign(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i,\mathbf{x}) + b) \quad \text{for } i = 1, 2, \ldots N \tag{3.9}$$

where the α_i's are non-zero Lagrange multipliers and $K(\mathbf{x}_i,\mathbf{x})$ is a kernel function which satisfies Mercer's theorem and is equal to the dot product $\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x})$. Using a kernel function allows the training of a classifier without explicit knowledge of the mapping function Φ (Huang, Davis & Townshend 2002). Some commonly used kernel functions that satisfy Mercer's condition include the linear, polynomial, radial basis functions as well as sigmoid kernels (Tzotsos & Argialas 2008).

The above formulation of SVM is applicable to binary classification problems which are not common in remote sensing (Waske *et al.* 2009). SVM is typically extended to the multiclass problem through either the one-against-all or the one-against-one approach. In the one-against-all approach, a set of binary classifiers is trained to separate each class from all the others. This results in *N* binary classifiers being trained for an *N*-class problem. A decision value is calculated for each class from these classifiers and data objects are classified to the class for which the largest decision value was determined (Tzotsos & Argialas 2008). The one-against-one approach applies a series of binary classifiers to each pair of classes resulting in $N(N-1)/2$ classifiers. An unknown sample receives a vote to the winning class from each binary classifier and is eventually assigned to the class having the most votes (Huang, Davis & Townshend 2002).

The full mathematical details of the theory and concepts of SVM are given by Vapnik (1995). Other useful sources include a SVM tutorial by Burges (1998) and the comprehensive introductory texts by Christianini & Shawe-Taylor (2000) & Hamel (2009). Steinwart & Christman (2008) provide an in-depth look at the success and recent developments of SVM and a good introduction to the concepts of kernel learning algorithms can be found in Gehler & Schölkopf (2009).

3.4 DATA AND EXPERIMENTAL DESIGN

3.4.1 Study area and data

The study area is located near Paarl in the Western Cape province of South Africa (see Figure 3.2). The boundaries of the study area match those of a Chief Directorate National GeoSpatial Information (CDNGI), 1:10 000 orthophoto map (3318DD5) and extend from 33°44´55˝ to 33°48´05˝S and from 18°56´54˝ to 19°00´06˝E (see Figure 3.3(*a*)). The area, measuring 4.9 km × 5.9 km, includes a diversity of land covers. It is also easily accessible by road and consequently suitable for carrying out field visits. SPOT 5 multispectral and panchromatic scenes (dated 29 March 2010) were acquired for the area. The scenes were orthorectified and subjected to atmospheric and topographic correction (ATCOR 3). The multispectral and panchromatic scenes were fused using a statistical fusion algorithm (PANSHARP) to create a single 2.5-metre-resolution multispectral image consisting of four spectral bands (green, red, near infrared and shortwave infrared). The image was subsetted to match the extents of the study area.
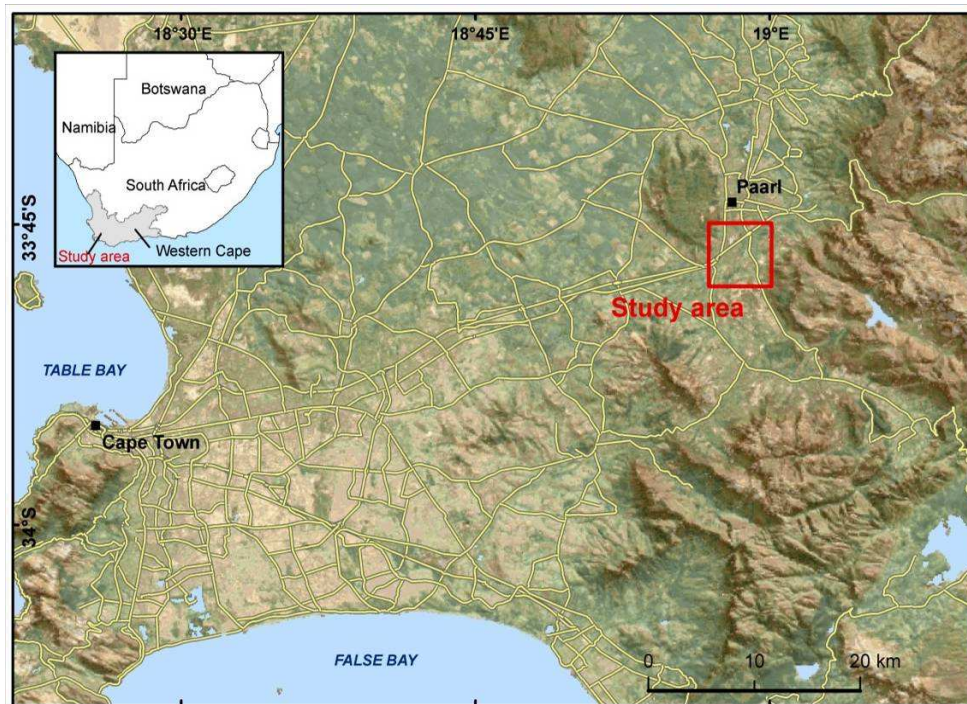
Figure 3.2: Location of the study area near Paarl in the Western Cape province of South Africa.
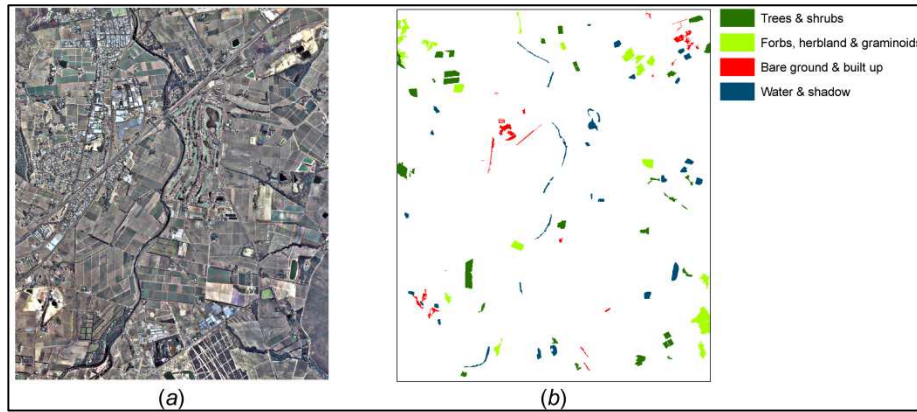
Figure 3.3: Aerial photograph of the study area (a) and the location of the selected land cover class samples (b).

3.4.2 Image segmentation, training data selection and feature selection

It is well known that a poor image segmentation can negatively affect the results of an object-based classification (Baatz, Hoffmann & Willhauk 2008; Fourie 2011; Hofmann, Strobl & Blaschke 2008; Tzotsos, Karantzalos & Argialas 2011). An optimal segmentation was, however, considered less critical for this study as it was assumed that all classifiers would be equally affected by a poor segmentation. Nevertheless, a good segmentation was attempted using the multiresolution segmentation (MRS) algorithm in eCognition 8.0. Various segmentation parameters were sequentially tested until a segmentation was obtained that, based on visual inspection, adequately represented all land cover features. A scale parameter of 30, a shape parameter of 0.2 and a compactness value of 0.3 produced the best results and provided a total of 6439 image objects with a high level of homogeneity.

A broad four-class classification scheme was adopted to limit subjectivity during generation of training sets (see Table 3.1). A number of field visits were carried out and class samples were selected by visual interpretation of a high-resolution (0.5m), colour aerial photograph. A total of 40 object samples per class were selected for use as training and reference data. Figure 3.3(*b*) shows the location of the object samples for each class.

Table 3.1: Land cover class descriptions

No	Class name	Description
1	Trees & shrubs	Mainly woody vegetation consisting of various trees, shrubs and bushes. Includes natural vegetation as well as plantations and agriculture.
2	Forbs, herbland & graminoids	Less woody vegetation consisting of forbs, herbs and grass areas. Includes mainly natural vegetation and recreational areas such as parks and fields.
3	Bare ground & built-up	Man-made, built-up features (buildings and roads) and artificial and natural bare areas.
4	Water and shadow	All dams, rivers and artificial water bodies and shadow areas that cannot be classified otherwise.

GEOBIA is well-suited to the calculation and incorporation of additional object features (Hay & Castilla 2006, 2008). A total of 47 object features, roughly based on the respective features used by Yu *et al.* (2006) and Laliberte, Browning & Rango (2010, 2012), were considered in this study. This included spectral, textural and geometric features, as well as image transforms. A classification tree analysis (CTA) was used to reduce the number of features and limit the impact of feature dimensionality on classification accuracies. CTA has been shown to be an effective feature selection method for remote sensing image classification (Gómez-Chova *et al.* 2003; Laliberte, Browning & Rango 2010, 2012; Laliberte, Fredrickson & Rango 2007; Otukei & Blaschke 2010; Yu *et al*. 2006). CART® software (by Salford Systems) was employed to perform the CTA and to select optimal features based on the primary splitters of the decision tree. Out of the total of 47 features, four features, namely: mean near infrared, maximum difference, hue (from the hue, saturation intensity image transform calculation) and grey level co-occurrence matrix (GLCM) dissimilarity were identified as primary splitters and subsequently used for classification.

The class samples and segmentation were stored as Environmental Systems Research Institute (ESRI) shapefiles with the selected features as their attributes. These shapefiles were used as input during the classification and training phase.

3.4.3   Software development

A software system was developed using C++ and the Microsoft® Visual Studio® 2010 (Express edition) development environment to automate the process of classification and accuracy assessment. Additional open-source libraries were also required to complete the implementation of the system. Libsvm 3.0 (Chang & Lin 2011) was used to implement one-against-one multiclass SVM. The ML and NN classifiers were implemented using the

OpenCV 2.2 library (Bradski 2000) and the Geospatial Data Abstraction Library (GDAL) (GDAL Development Team 2010) was used for the manipulation of shapefiles and rasters.

The radial basis function kernel, as recommended by Hsu, Chang & Lin (2010), was selected for the SVM implementation. Appropriate values for the error parameter (*C*) and the kernel parameter (γ) were determined using a simple grid search and cross-validation approach. First, a coarse grid search was carried out on C = 2^{-5}, 2^{-3}, … , 2^{-15} and γ = 2^{-15}, 2^{-13}, … , 2^{3}, after which a finer grid search was performed based on the results of the first search (as recommended by Hsu, Chang & Lin (2010)). All data was scaled from -1 to 1 to prevent data with higher numerical ranges having greater effect than those with lower ranges (Hsu, Chang & Lin 2010)

3.4.4   Experiment workflow

The developed system requires the following input: a shapefile containing the selected object samples, a shapefile representing the unclassified objects (the same features need to be included in both shapefiles and they need to be in the same order), and a user-defined number of iterations. At the start of each experiment (program run), the object samples are randomly split into a training and a reference data set of equal size. The following steps are repeated for the number of defined iterations:

1. Randomly select four training samples per class (i.e. 20% of all training samples) from the training data set.

2. Train the SVM, NN and ML classifiers using the selected samples.

3. Use SVM, NN and ML to respectively classify the unclassified shapefile, and perform automated accuracy assessments using the reference data set.

4. Add two randomly selected training areas per class to the training set (10% increase) and repeat steps (2) to (4) until all the samples in the training data set have been selected (i.e. 100% of all training samples)

Multiple iterations are performed to account for the sensitivity of the classifiers to the random sample selection at each training set size. All the relevant statistics (confusion matrices and summaries of overall kappa, overall accuracy and producer's and user's accuracies) are written to multiple text files at the end of each set of iterations. A batch file was created to run sequential instances of the program and to organize the output files. A python script was also created to summarize the output generated from multiple program runs. The results presented

in this paper are averaged values based on 20 individual program runs each consisting of 100 iterations.

Confusion matrices were investigated at each sample size and used to compare the performance of the different classifiers concerning the specific land cover classes. The matrices were also used to calculate the producer's, user's and overall accuracies, as well as the kappa statistic, for each classifier and sample size combination.

3.5   RESULTS AND DISCUSSION

The results of the investigation into the effect of training-data size on classification performance, in an object-based supervised classification environment, are shown in Tables 2−10. The confusion matrices for only three training set sizes, namely 20% (Tables 3.2 to 3.4), 60% (Tables 3.5 to 3.7) and 100% (Tables 3.8 to 3.10), are provided due to limited space. The results are summarized in the overall accuracy and kappa graphs (Figures 3.4 & 3.5). An overall accuracy of more than 85% is generally regarded as acceptable, although the required accuracy of maps depends on many factors (Congalton & Green 2009).

Table 3.2: SVM confusion matrix at four samples per class (20%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	19442	8094	834	331	28700	67.7	32.3
Forbs/herb/gram	4848	19790	322	920	25880	76.5	23.5
Bare gr/built up	196	1571	6564	793	9124	71.9	28.1
Water & shadow	404	2034	2964	14239	19642	72.5	27.5
TOTALS	24890	31489	10685	16282	83346		
CA%	78.1	62.8	61.4	87.5			
EC%	21.9	37.2	38.6	12.5			
Overall accuracy:	72.0						
Overall kappa:	0.61						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 3.3: NN confusion matrix at four samples per class (20%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	15700	5723	7276	1	28700	54.7	45.3
Forbs/herb/gram	5840	17977	1686	377	25880	69.5	30.5
Bare gr/built up	2169	1237	4064	1654	9124	44.5	55.5
Water & shadow	27	618	1317	17680	19641	90.0	10.0
TOTALS	23737	25556	14342	19711	83346		
CA%	66.1	70.3	28.3	89.7			
EC%	33.9	29.7	71.7	10.3			
Overall accuracy:	66.5						
Overall kappa:	0.54						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 3.4: ML confusion matrix at four samples per class (20%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	10688	6841	6983	4188	28700	37.2	62.8
Forbs/herb/gram	6380	8832	6049	4618	25880	34.1	65.9
Bare gr/built up	1668	2282	3322	1852	9124	36.4	63.6
Water & shadow	3829	4685	5012	6115	19641	31.1	68.9
TOTALS	22565	22642	21366	16773	83346		
CA%	47.4	39.0	15.5	36.5			
EC%	52.6	61.0	84.5	63.5			
Overall accuracy:	34.7						
Overall kappa:	0.13						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 3.5: SVM confusion matrix at 12 samples per class (60%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	24045	4405	222	27	28700	83.8	16.2
Forbs/herb/gram	3142	22495	52	190	25880	86.9	13.1
Bare gr/built up	17	608	7985	514	9124	87.5	12.5
Water & shadow	95	667	786	18093	19641	92.1	7.9
TOTALS	27299	28176	9046	18825	83346		
CA%	88.1	79.8	88.3	96.1			
EC%	11.9	20.2	11.7	3.9			
Overall accuracy:	87.1						
Overall kappa:	0.82						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 3.6: NN confusion matrix at 12 samples per class (60%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	19222	4589	4889	0	28700	67.0	33.0
Forbs/herb/gram	6046	18178	867	790	25880	70.2	29.8
Bare gr/built up	1463	852	5353	1456	9124	58.7	41.3
Water & shadow	4	444	590	18604	19641	94.7	5.3
TOTALS	26735	24063	11698	20850	83346		
CA%	71.9	75.5	45.8	89.2			
EC%	28.1	24.5	54.2	10.8			
Overall accuracy:	73.6						
Overall kappa:	0.64						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 3.7: ML confusion matrix at 12 samples per class (60%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	24722	3430	534	15	28700	86.1	13.9
Forbs/herb/gram	2757	21501	1513	109	25880	83.1	16.9
Bare gr/built up	206	14	8036	868	9124	88.1	11.9
Water & shadow	785	912	1016	16929	19642	86.2	13.8
TOTALS	28469	25857	11099	17921	83346		
CA%	86.8	83.2	72.4	94.5			
EC%	13.2	16.8	27.6	5.5			
Overall accuracy:	85.4						
Overall kappa:	0.80						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 3.8: SVM confusion matrix at 20 samples per class (100%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	25152	3543	4	2	28700	87.6	12.4
Forbs/herb/gram	2163	23660	3	54	25880	91.4	8.6
Bare gr/built up	5	383	8271	466	9124	90.7	9.3
Water & shadow	59	535	101	18946	19641	96.5	3.5
TOTALS	27379	28121	8379	19467	83346		
CA%	91.9	84.1	98.7	97.3			
EC%	8.1	15.9	1.3	2.7			
Overall accuracy:	91.2						
Overall kappa:	0.88						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 3.9: NN confusion matrix at 20 samples per class (100%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	20655	5037	3008	0	28700	72.0	28.0
Forbs/herb/gram	6419	17845	599	1017	25880	69.0	31.0
Bare gr/built up	1043	639	6319	1123	9124	69.3	30.7
Water & shadow	0	449	286	18907	19642	96.3	3.7
TOTALS	28118	23969	10212	21048	83346		
CA%	73.5	74.5	61.9	89.8			
EC%	26.5	25.5	38.1	10.2			
Overall accuracy:	76.5						
Overall kappa:	0.67						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 3.10: ML confusion matrix at 20 samples per class (100%)

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%†	EO%†
Trees & shrubs	25867	2833	0	0	28700	90.1	9.9
Forbs/herb/gram	1765	22779	1336	0	25880	88.0	12.0
Bare gr/built up	213	0	8168	743	9124	89.5	10.5
Water & shadow	571	946	263	17862	19642	90.9	9.1
TOTALS	28416	26558	9766	18606	83346		
CA%	91.0	85.8	83.6	96.0			
EC%	9.0	14.2	16.4	4.0			
Overall accuracy:	89.6						
Overall kappa:	0.86						

†PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission
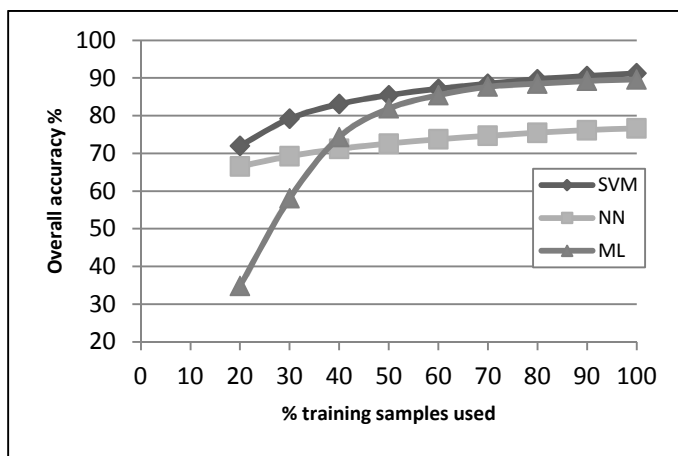

Figure 3.4: Average overall accuracy values for SVM, NN and ML at different training set sizes.
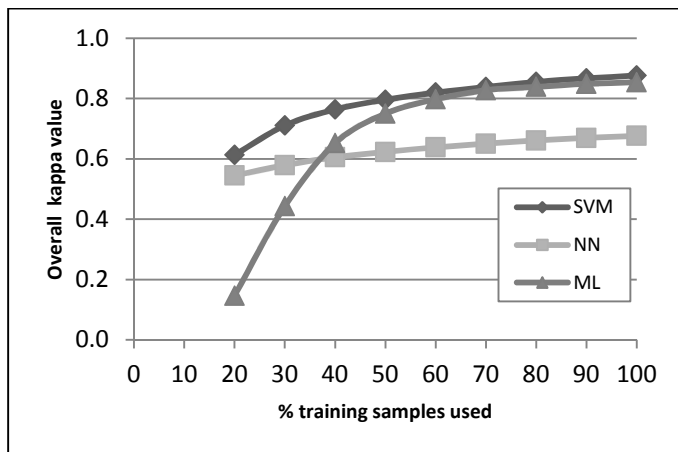

Figure 3.5: Average kappa values for SVM, NN and ML at different training set sizes.

From Tables 3.2 to 3.4 it is clear that SVM was more successful than the other classifiers in identifying all four land cover classes when only 16 (20%) of the training samples were used as input. SVM's superior performance at the minimum sample size is in agreement with observations that SVM generally outperforms traditional classifiers when few training samples are available (Foody & Mathur 2004b; Li *et al.* 2010; Lizarazo 2008; Pal and Mather 2005; Mountrakis, Im & Ogole 2011). The NN classifier's lower overall accuracy was most significantly influenced by its inability to correctly classify the *Bare ground & built* up class while also being less successful than SVM at identifying Trees & shrubs and Forbs, herbland and graminoids. The *Water & Shadow* class was, however, classified more accurately by NN than by SVM. The high level of confusion that existed between all classes and the significantly poorer overall performance of ML at 20% training samples indicate that ML fails when an insufficient number of training samples are used.

Increasing the sample size to 12 samples per class (60%) yielded improved results for all three classifiers (Tables 3.5 to 3.7). Although SVM remained the superior classifier, ML improved significantly to achieve competitive results. NN, however, performed significantly weaker than the other classifiers. Tables 3.8 to 3.10 similarly reveal higher accuracies for all three classifiers at the maximum sample size (100%), but that SVM outperformed ML and NN with the latter achieving considerably lower accuracies. From the confusion matrices at 12 (60%) and 20 (100%) samples per class, it can be seen that SVM was, compared to ML, far more effective in identifying the *Bare ground & built up* areas. ML's lower overall accuracies compared to SVM was mainly due to the over classification (high percentage of commission errors) of the *Bare ground & built up* class. The misclassification of *Bare ground & built up* was also the most significant factor in NN's comparatively weak performance, although it was also significantly less successful in classifying the *Trees & shrubs* and *Forbs herbland & graminoids* classes.

Generally, the most confused classes were *Trees & shrubs* and *Forbs, herbland & graminoid*. These classes are spectrally similar and are consequently difficult to discriminate. However, the ability of SVM and ML to distinguish between these classes, improved significantly as the training sets became larger. At the maximum sample size (100%), ML was slightly better

(9.5% confusion[1]) at distinguishing between these classes than SVM (11.7% confusion) while NN (29.8% confusion) was far less successful (see Tables 8 to 10). These findings is in contrast with those of Dixon & Candade (2008) who noted that SVM provided better discriminatory power than ML for classes with a close spectral proximity. Szuster, Chen & Borger (2011) found that SVM performed substantially better than ML and ANN in differentiating between their spectrally-similar *Coastal rock* and *Developed* classes, while Kavzoglu & Colkesen (2009) showed that SVM performed significantly better than ML for extracting their *Deciduous* class which had a high spectral resemblance to several other classes. It should be noted, however, that these pixel-based studies used only spectral band values as classification input. The object-based nature of this study, as well as the object features selected through CTA, might have contributed to ML achieving slightly better discrimination between the spectrally similar *Trees & shrubs* and *Forbs, herbland & graminoid* classes than SVM. The use of textural measures (such as the selected GLCM Dissimilarity measure) has been shown to improve the performance of the ML classifier (Berberoglu *et al.* 2007, Heinl *et al.* 2009). Further research is required to determine the impact of specific features on the performance of each of the tested classifiers for GEOBIA.

Water & shadow was consistently classified more accurately than any class (with the exception of ML at very small sample sizes). This finding was not unexpected as *Water & shadow* is generally an easily distinguishable spectral class and thus relatively easy to classify. This was also noted by Kavzoglu & Colkesen (2009) who achieved nearly 100% accuracy for water classifications irrespective of the classifier used.

The *Bare ground & built-up* class is, in contrast to *Water & shadow*, a complex class comprising many different man-made structures and some natural features which make it hard to classify successfully. Compared to the other classifiers, SVM's ability to differentiate this class from other classes improved significantly as the training set size increased. ML tended to over classify the *Bare ground & built-up* class. Dixon & Candade (2008), using a more sophisticated seven-class classification scheme, also found that ML significantly over classified their *Urban* class (which is similar to the *Bare ground & built-up* class used in this study) when compared to SVM and ANN.

---

[1]The percentage confusion between the two classes was calculated by adding the number of misclassifications between them and dividing by the sum of the reference pixels for the two classes, e.g. the percentage confusion for SVM at 20 samples (Table 3.8) was calculated as follows: (3543+2163)/(25152+23660)*100 = 9.5%.

As one would expect, the overall accuracy of all the classifiers improved as the training-sample size increased (Figures 3.4 & 3.5). At small training set sizes (i.e. less than eight per land cover class), the ML classifier performs poorly compared to the other two classifiers. A sharp increase in performance is observed when the training set increases from four (20%) to 10 (50%) samples per land cover class. The relatively weak performance of the ML classifier at very small training set sizes, as well as the rapid initial increase in accuracy can likely be attributed to insufficient training data for estimating the representative mean vector and variance–covariance matrix used by the classifier to calculate the likelihood values for each class (Oommen *et al*. 2008). SVM and NN do not require the calculation of such statistical measures and therefore perform better when few samples are available. They both show a steady increase in accuracy when the training samples per class increase from four (20%) to 20 (100%). SVM consistently outperformed the other classifiers and achieved significantly higher accuracies than ML and NN when using few (less than 10 samples per class) training samples. ML produced competitive results when 12 or more samples per class (60%) were used while NN remained comparatively weak.

While SVM produced superior results when using small training sets as compared to ML and NN, significant increases in classification accuracy are obtained as the training set size was increased. The pixel-based comparative studies conducted by Foody & Mathur (2004b), Huang, Davis & Townshend, (2002) and Oommen *et al.* (2008) revealed a similar improvement in SVM accuracy as the training set size increased with SVM outperforming other classifiers at the maximum training set size. Our findings support those of Foody & Mathur (2004a) who concluded that, while only a small number of training samples are required to produce the support vectors, a larger training set is required to ensure that the optimal support vectors are identified. It is consequently important to use a specific training data collection approach (Foody & Mathur 2004b, 2006) when SVM is used under very small training set conditions.

The higher accuracies achieved by SVM throughout the experiment indicate that SVM holds much potential for object-based supervised classification of high-resolution remotely sensed imagery. It is also clear that the parametric nature of ML makes it unsuitable for situations where few training samples are available, while competitive results can be expected when sufficient training data is available. The non-parametric NN classifier, while achieving higher accuracies than ML at very small training set sizes, was the weakest of the three classifiers tested. This result should be of particular interest to eCognition users, as the latest version of

the software (8.7) allows users to choose between SVM, CART and Bayes classifiers as alternatives to the commonly applied NN classifier.

3.6   CONCLUSIONS

Supervised classification, as a technique for mapping land cover, is highly dependent on the available training data. In this study, the performances of SVM, NN and ML were compared for object-based land cover classification when different training set sizes are used. The results show that all the classifiers performed better when more training samples were used. The statistical ML classifier was found to be unsuitable for use when very small training sets are available while the NN classifier performed comparatively poorly throughout the experiment. While SVM performed considerably better than the other classifiers at very small training set sizes, an intelligent sample selection approach, specifically tailored for SVM classification (Foody & Mathur 2004b, 2006), might be required to improve classification accuracies under such conditions.

In this study a basic four-class land cover classification scheme was used. Future work will focus on the implementation of more complex classification schemes to see if comparable results are obtained. The research will also be extended to investigate the influence of various object features and increased feature dimensionality on the performance of the classifiers for GEOBIA.

CHAPTER 4: EFFECT OF FEATURE DIMENSIONALITY ON OBJECT-BASED IMAGE CLASSIFICATION: A COMPARISON OF THREE CLASSIFIERS[*]

4.1 ABSTRACT

Supervised classifiers are commonly employed in remote sensing to extract land cover information. While spectral information is typically used as the main discriminating features for such classifiers, additional features such as vegetation indices, transformed images, textural information, contextual information and ancillary data may also significantly influence the accuracy of classification. Geographic object-based image analysis (GEOBIA) allows the easy integration of such additional features into the classification process. This paper compares the performance of three supervised classifiers in a GEOBIA environment as an increasing number of object features are included as classification input. Classification tree analysis (CTA) was employed for feature selection and importance ranking. Object features were considered in the order of their obtained rank. The support vector machine (SVM) produced superior classification accuracies when compared to those of nearest neighbour (NN) and maximum likelihood (ML) classifiers. Both SVM and NN produced stable results as the feature set size was increased towards the maximum (22 features). Although the ability of ML to perform under conditions of large feature dimensionalities and few training areas was limited, it outperformed NN under certain conditions.

4.2 INTRODUCTION

Detailed, accurate and up-to-date land cover information is essential for environmental and socio-economic research (Heinl *et al.* 2009; Lu and Weng 2007). A large number of operational satellite platforms exist that have the capability to provide remotely sensed imagery at various spatial and temporal scales (Foody 2002). This abundance of available data offers great potential for generating frequently updated thematic maps as remotely sensed images cover large areas, are acquired at regular intervals and are less costly than traditional ground-survey methods (Foody 2009; Gao 2009; Pal & Mather 2004; Szuster, Chen & Borger

---

[*]This chapter was submitted for publication to the International Journal of Applied Earth Observation & Geoinformation and consequently conforms to the prescribed structure of that journal. Some of the text, tables and figures used in this chapter are identical to those of the previous chapter as the same data and methods were used.

2011). Current image processing techniques are limited in their ability to extract accurate land cover features automatically (Baraldi *et al.* 2010). Many factors affect the accuracy of image classification (Lu & Weng 2007) and the quality of land cover maps is often perceived as being insufficient for operational use (Foody 2002).

Supervised classification, an approach commonly used for remote sensing image classification, requires samples of known identity (training samples) to construct a model capable of classifying unknown samples. Apart from selecting a suitable classifier, the number and quality of training samples are key to successful classification (Hubert-Moy *et al.* 2001; Lillesand, Kiefer & Chipman 2008; Lu & Weng 2007). A sufficient number of training samples is generally required to perform a successful classification and the samples need to be well distributed and sufficiently representative of the land cover classes being evaluated (Campbell 2006; Mather 2004; Lu & Weng 2007; Gao 2009). In remote sensing applications, the availability of labelled training samples is often limited (Gehler & Shölkopf 2009; Mountrakis, Im & Ogole 2011) as their collection is time-consuming, expensive and tedious, often requiring the study of maps and aerial photographs and carrying out field visits (Campbell 2006).

While the selection of an appropriate classifier and the delineation of the training set are crucial, the addition of variables other than the original spectral bands can significantly influence the performance of image classification (Heinl *et al.* 2009; Lu & Weng 2007). In particular, vegetation indices, transformed images, textural information, contextual information and ancillary data are often incorporated into image classification (Lu & Weng 2007). Heinl *et al.* (2009) have compared the performance of maximum likelihood (ML), artificial neural network (ANN) and discriminant analysis (DA) classifiers when topographic measures, normalized difference vegetation index (NDVI), and texture measures are incrementally added to Landsat 7 ETM+ spectral data as input variables. The addition of such variables generally leads to an increase in classification accuracy implying that the addition of such variables could potentially be as significant as classifier selection. Berberoglu *et al.* (2007) found that the addition of texture features improved the performance of ML and ANN classifiers for certain land covers. However, for some classifiers an increase in input dimensionality decreases the reliability of statistical parameter estimations and may consequently result in a decrease in classification accuracy (Oommen *et al*. 2008; Pal & Mather 2005). This is known as the Hughes effect (Hughes 1968) – the so-called curse of dimensionality – which postulates that the classification accuracy will decrease after a certain feature set size is reached unless the number of training samples is proportionally increased

(Chen & Ho 2008). The Hughes effect is therefore more likely to be encountered when small training sets are used and the input dimensionality is increased.

Geographic object-based image analysis (GEOBIA) has emerged as an alternative to pixel-based image processing (Blaschke 2010; Blaschke & Lang 2006; Hay & Castilla 2006, 2008). GEOBIA involves a segmentation step during which image pixels are grouped into homogeneous interlocking regions as determined by a specific segmentation algorithm (Campbell 2006). These image segments contain additional spectral and spatial information when compared to single pixels (Blaschke 2010). Its ability to incorporate contextual information and ancillary data makes GEOBIA suitable for the integration of various additional features for image classification. Usually, the mean values of the pixels within an object are used to train an object-based supervised classifier. Because this effectively reduces the number of training samples available to the classifier (Tzotsos & Argialas 2008), GEOBIA is generally more sensitive to the Hughes effect when statistical classifiers are used.

Support vector machines (SVMs) have been shown to improve the reliability and accuracy of supervised classifications (Oommen *et al.* 2008). SVMs are known for their good generalizing ability even when few training samples are available (Foody & Mathur 2004b; Li *et al.* 2010; Lizarazo 2008; Mountrakis, Im & Ogole 2011; Pal & Mather, 2005) and they are robust to increases in input dimensionality compared to other statistical classifiers (Camps-Valls & Bruzzone 2005; Camps-Valls *et al.* 2004, 2006; Melgani & Bruzzone 2004; Mercier & Lennon 2003; Oommen *et al.* 2008; Pal & Mather 2004, 2005). Comparative studies have shown that SVMs produce superior, or at least comparable, results for multispectral and hyperspectral image classification opposed to more commonly used methods such as ML, NN, ANN and decision trees (Camps-Valls & Bruzzone 2005; Camps-Valls *et al.* 2004; Dixon & Candade 2008; Foody & Mathur 2004a; Gualtieri & Cromp 1998; Huang, Davis & Townshend 2002; Kavzoglu & Colkesen 2009; Keuchel *et al.* 2003; Melgani & Bruzzone 2002, 2004; Mercier & Lennon 2003; Oommen *et al.* 2008; Pal & Mather 2004, 2005; Szuster, Chen & Borger 2011; Tzotsos & Argialas 2008; Watanachaturaporn Arora & Varshney 2008).

Very few studies have compared the performance of different supervised classifiers in an object-based environment. A notable exception is Tzotsos & Argialas (2008), who reported that SVM outperformed NN classifiers for mapping land cover when using Landsat TM spectral bands as input. Other recent studies have also implemented object-based SVM classification (Duro, Franklin & Dubé 2012; Li *et al.* 2010; Liu & Xia, 2010; Lizarazo 2008; Tzotsos, Karantzalos & Argialas 2011; Wu *et al.* 2009). However, no studies have been

published that investigate the comparative performance of SVM as feature space is increased through the use of additional object features for GEOBIA land cover classification. Although it is expected, due to its non-parametric nature, that SVM would be more effective than statistical classifiers for incorporating additional features, it has not been demonstrated with land cover mapping in an object-based environment.

This paper aims to investigate the performance of object-based SVM for land cover classification compared to NN and ML classifiers, with a focus on feature dimensionality. The NN and ML classifiers were chosen for benchmarking as the latter is the most commonly used supervised classification method in remote sensing (Albert 2002; Stephenson 2010; Waske *et al.* 2009) and NN is the supervised classifier most commonly employed for object-based supervised classification (Campbell 2006).

The rest of the paper is structured into four sections, the first overviews the NN, ML and SVM classifiers and the next section describes the data used, the experimental design and the development of the software that automated the assessments. Thereafter the results are discussed and finally the findings are summarized and suggestions are made for further research.

4.3   OVERVIEW OF THE TESTED SUPERVISED CLASSIFIERS

Supervised classifiers are typically categorized as either parametric or non-parametric. Parametric classifiers assume that the data follows a known distribution and they involve the estimation of certain statistical parameters from training data (Jain, Duin & Mao 2000). In contrast, non-parametric classifiers make no assumptions about the distribution of the data and do not require the calculation of statistical parameters. Supervised classifiers can also be categorized as density estimation (DE) or boundary estimation (BE) classifiers. DE classifiers obtain decision boundaries indirectly from density functions, while BE classifiers obtain decision boundaries directly from training data by solving some optimization problem (Fourie 2011; Jain, Duin & Mao 2000). The three classifiers applied in this paper have different properties regarding these definitions.

4.3.1   Maximum likelihood

ML is the most commonly used supervised classification method in remote sensing and is derived from the Bayesian decision rule for the case where all classes have equal prior probabilities (Albert 2002; Waske *et al.* 2009). It is a parametric, DE method that calculates the mean vector and the variance–covariance matrix from the training data and uses it to

construct an estimated distribution for each class. The probability ($P(x)$) of an unknown sample, represented by a vector x of p features, belonging to a certain class (i) is calculated from this distribution by

$$P(x) = 2\pi^{-0.5p}|S_i|^{-0.5}\mathrm{e}^{-0.5y} \qquad\qquad (4.1)$$

where S_i is the variance–covariance matrix for class i and y is the Mahalanobis distance. The sample is assigned to the class for which the highest probability is calculated. ML has the advantage of accounting for variability within each class by using the covariance matrix. Although ML produces high classification accuracies (Albert 2002; Gao 2009; Stephenson 2010; Szuster, Chen & Borger 2011; Pal & Mather 2003; Waske *et al.* 2009,), its assumption that all input data is normally distributed (Albert 2002; Oommen *et al*. 2008) and its sensitivity to training-data quality (Campbell 2006) have been cited as reasons for the poor performance in certain remote sensing applications (Fourie 2011).

4.3.2   Nearest neighbour

NN is a simple distance-based, non-parametric decision procedure and is also a DE method. The NN decision rule is a special case of the k-nearest neighbour (kNN) rule with the free parameter k set to one. The kNN rule assigns an unknown sample to the class that occurs most frequently among its k-nearest neighbours (Campbell 2006; Cover & Hart 1967). NN simply assigns a sample to the class of the training sample it is closest to in multidimensional feature space. In other words, if the training instances are represented by vectors x_1, $x_{2, \ldots}$, x_n and the Euclidean distance between two vectors is $d(x_1, x_2)$, the nearest neighbour x_{nn} of an unknown sample x is defined as:

$$\mathbf{x}_{nn\,=}\,\mathbf{x}_j \quad \text{where} \quad \mathbf{x}_j \in \{\,\mathbf{x}_1, \mathbf{x}_{2,\ldots}, \mathbf{x}_n\} \qquad\qquad (4.2)$$

if

$$d(x_j,\,x) = \min d(x_i, x) \quad \text{for} \quad i = 1, \ldots, \mathrm{n} \qquad\qquad (4.3)$$

The unknown sample is classified to the same class as x_{nn}. The NN and kNN rules are often employed in remote sensing as their simplicity and effectiveness in classifying heterogeneous classes provide a practical advantage (Campbell 2006). In this paper the NN rule was employed because small training sets were used during the experiment. Using larger values of k is more appropriate when the number of samples is large (Cover & Hart 1967).

4.3.3   Support vector machines

SVM is a non-parametric, BE-supervised classification technique based on statistical learning theory (Vapnik 1995). In its simplest form, SVM is a binary classifier capable of separating two linearly separable classes. SVM relies on identifying the optimal separating hyperplane between two sets of classes in a multidimensional feature space. Once identified, the hyperplane is used as a decision surface to classify unknown samples. As shown in Figure 4.1(a), many hyperplanes may exist that are capable of separating two classes, but there can only be one optimal hyperplane that maximizes the margin between the hyperplane and the closest training samples (called the support vectors) (see Figure 4.1(b)). The support vectors lie on two hyperplanes parallel to the optimal hyperplane (Oommen *et al.* 2008) and are defined as:

$$w\,\mathbf{x}_i + b = \pm\,1 \tag{4.4}$$

where $\mathbf{x}$ is a point on the hyperplane, w is a vector normal to the hyperplane and b is the bias. Rescaling the parameters w and b allows the margin between these support planes to be expressed as $2/\|w\|$ (Fourie 2011; Oommen *et al.* 2008).   The optimization problem for maximizing this margin is given by:

$$\text{Minimize}\quad \frac{1}{2}\,\|w\|^2 \tag{4.5}$$

subject to the constraints

$$y_i(w\,\mathbf{x}_i + b) - 1 \geq 0 \quad \text{for}\ \ i = 1, 2, \dots N \tag{4.6}$$

where $y_i \in \{1,-1\}$ is the class label and N is the number of training samples. This optimization problem is the basic definition of SVM for the case of two linearly separable classes. Target classes are often spectrally similar and the training set may include outliers that reduce the likelihood that classes will be linearly separable for the majority of remote sensing classification problems.
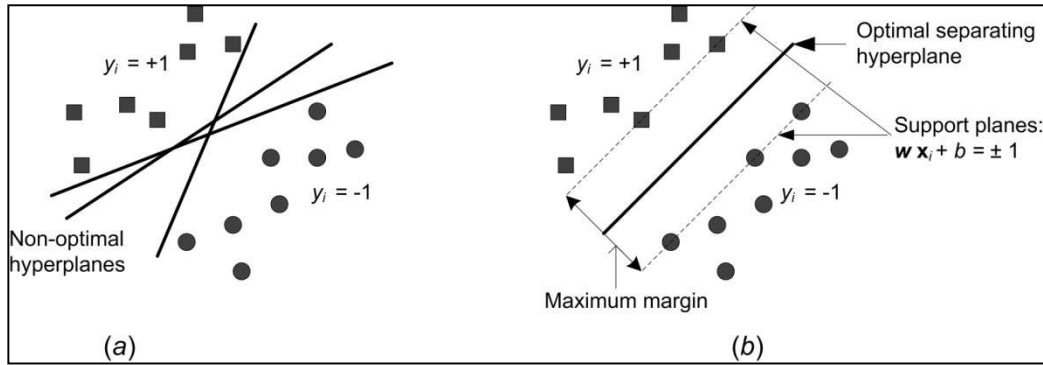
Figure 4.1: Conceptual view of SVM showing how (a) multiple hyperplanes may separate the classes and how (b) SVM relies on identifying the optimal separating hyperplane.

SVM can be extended for the linearly inseparable case through the introduction of slack variables and by applying a kernel function (Mountrakis, Im & Ogole 2011). Equations (4.5) and (4.6) are then reformulated as:

Minimize $\quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i$ (4.7)

subject to the constraints

$y_i(w\ \Phi(x_i) + b) - 1 \geq 1 - \xi_i$ and $\xi_i \geq 0$, i = 1, 2, ..., N (4.8)

where the ξ_i's are the slack variables, C is a user-defined penalty parameter and Φ is a mapping function used to map the data (x_i) to a higher-dimensional feature space and where a linear hyperplane can be fitted that is non-linear in the original input space (Waske *et al*. 2009). To solve this optimization problem, it is typically reformulated as a Lagrange function from which the following decision rule can be obtained:

$f(x) = sign(\ \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b)$ for $i = 1, 2, ... N$ (4.9)

where the α_i's are non-zero Lagrange multipliers and $K(\mathbf{x}_i, \mathbf{x})$ is a kernel function which satisfies Mercer's theorem and is equal to the dot product $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})$. Using a kernel function allows for the training of a classifier without explicit knowledge of the mapping function Φ (Huang, Davis & Townshend 2002). Some commonly used kernel functions that satisfy Mercer's condition include the linear, polynomial, radial basis functions as well as sigmoid kernels (Tzotsos & Argialas 2008).

The above formulation of SVM is applicable to binary classification problems which are not common in remote sensing (Waske *et al.* 2009). SVM is typically extended to the multiclass problem through either the one-against-all or the one-against-one approach. In the one-against-all approach, a set of binary classifiers is trained to separate each class from all others. This results in *N* binary classifiers being trained for an *N*-class problem. A decision value is calculated for each class from these classifiers and data objects are classified to the class for which the largest decision value was determined (Tzotsos & Argialas 2008). The one-against-one approach applies a series of binary classifiers to each pair of classes resulting in *N*(*N* – 1)/2 classifiers. An unknown sample receives a vote to the winning class from each binary classifier and is eventually assigned to the class having the most votes (Huang, Davis & Townshend 2002).

The full mathematical details of the theory and concepts of SVM are given by Vapnik (1995). Other useful sources are a SVM tutorial by Burges (1998) and the comprehensive introductory texts by Christianini & Shawe-Taylor (2000) and Hamel (2009). Steinwart & Christman (2008) provide an in-depth look at the successes and recent developments of SVM and a good introduction to the concepts of kernel learning algorithms can be found in Gehler & Schölkopf (2009).

4.4   DATA AND EXPERIMENTAL DESIGN

4.4.1   Data and pre-processing

The study area is located near Paarl in the Western Cape province of South Africa (see Figure 4.2). The boundaries of the study area were chosen to match those of a Chief Directorate National GeoSpatial Information (CDNGI), 1:10 000 orthophoto map (3318DD5) and they extend from 33°44´55˝ to 33°48´05˝S and from 18°56´54˝ to 19°00˝06˝E (see Figure 4.3(*a*)). The area, measuring 4.9 km × 5.9 km, was chosen because it is easily accessible by road and is suitably located for field visits. SPOT 5 multispectral and panchromatic scenes (dated 29 March 2010) were acquired for the area. The scenes were orthorectified and subjected to atmospheric and topographic correction (ATCOR 3). The multispectral and panchromatic scenes were then fused using a statistical fusion algorithm (PCI *PanSharp*) to create a single 2.5-m-resolution multispectral image consisting of four spectral bands (green, red, near infrared and shortwave infrared). Fusion was required as the higher spatial resolution would improve discrimination of urban features (Amarsaikhan *et al.* 2010; Pohl & Van Genderen

1998). The PCI *PanSharp* algorithm has been shown to produce superior fusion quality compared to other image fusion techniques regardless of sensor, image and spectral band variations (Zhang & Mishra 2012). The image was subsetted to match the extents of the study area.
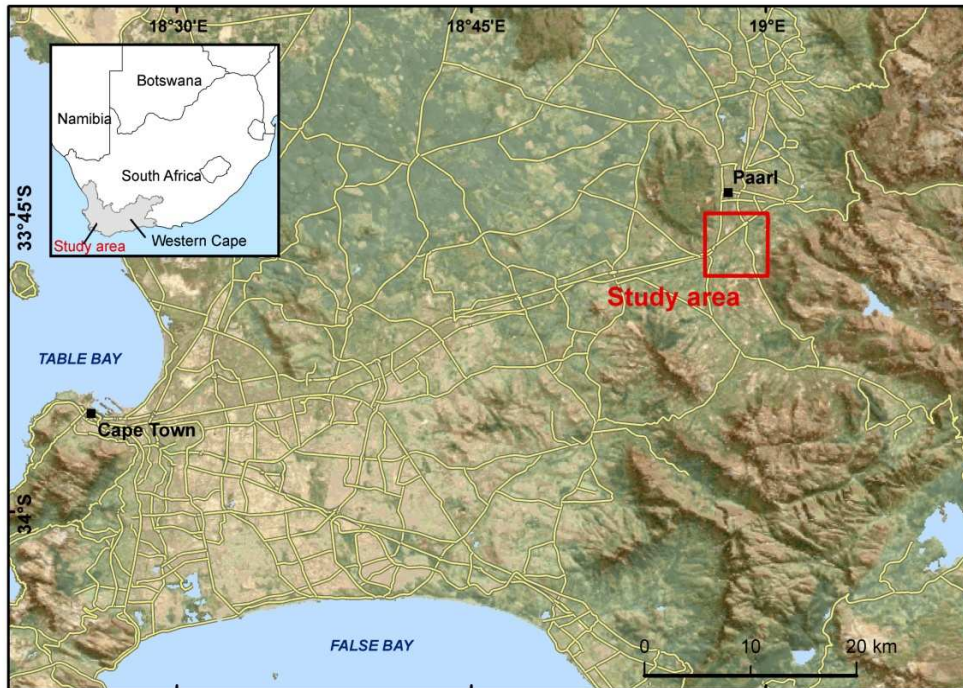
Figure 4.2: Location of the study area near Paarl in the Western Cape province of South Africa.
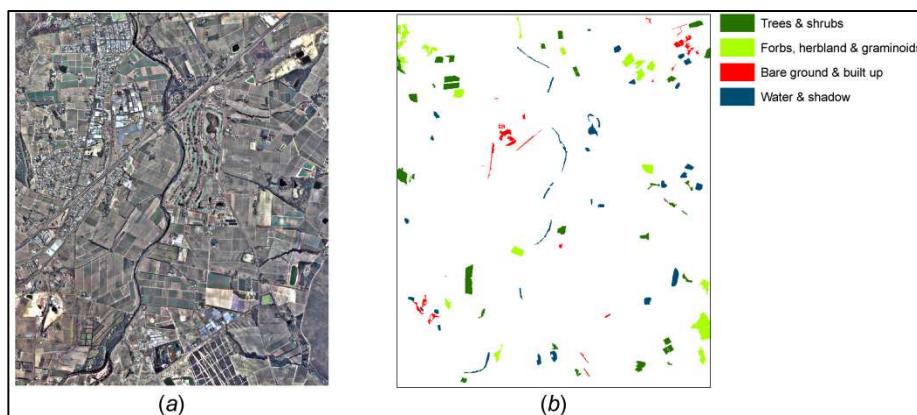
Figure 4.3: Aerial photograph of the study area (a) and the location of the selected land cover class samples (b).

4.4.2   Image segmentation, training data selection and feature ranking

It is well known that poor image segmentation can negatively affect the results of an object-based classification (Baatz, Hoffmann & Willhauk 2008; Fourie 2011; Hofmann, Strobl & Blaschke 2008; Tzotsos, Karantzalos & Argialas 2011). An optimal segmentation was, however, considered less critical for this study as it was assumed that all classifiers would be

equally affected by a poor segmentation. Nevertheless, a good segmentation was attempted using the multiresolution segmentation (MRS) algorithm in eCognition 8.0. Various segmentation parameters were sequentially tested until a segmentation was obtained that, based on visual inspection, adequately represented all land cover features. A scale parameter of 30, a shape parameter of 0.2 and a compactness value of 0.3 produced the best results and provided a total of 6439 image objects with a high level of homogeneity.

A broad four-class classification scheme was adopted to limit subjectivity during the generation of training sets (see Table 4.1). Some field visits were made and class samples were selected by visual interpretation of a high-resolution (0.5m) colour aerial photograph. Forty object samples per class were selected for use as training and reference data. Figure 3(*b*) shows the location of the object samples for each class.

Table 4.1: Land cover class description

No	Class name	Description
1	Trees & shrubs	Mainly woody vegetation consisting of various trees shrubs and bushes. Includes natural vegetation as well as plantations and agriculture.
2	Forbs, herbland & graminoid	Less woody vegetation consisting of forbs, herbs and grass areas. Includes mainly natural vegetation and recreational areas such as parks and fields.
3	Bare ground & built-up	Man-made, built-up features (buildings and roads) and artificial and natural bare areas.
4	Water & shadow	All dams, rivers and artificial water bodies and shadow areas that cannot be classified otherwise.

A total of 47 object features, roughly based on the features used by Yu *et al.* (2006) and Laliberte, Browning & Rango (2010, 2012), were considered in this study (see Table 4.2). Classification tree analysis (CTA) was used for selecting the most significant features for the particular application. CTA has been shown to be a an effective feature selection method and has been successfully applied in GEOBIA (Addink *et al.* 2010; Chubey, Franklin & Wulder 2006; Laliberte, Browning & Rango 2010, 2012; Laliberte, Fredrickson & Rango 2007; Yu *et al.* 2006). CART® software (by Salford Systems) was employed to perform a CTA on the 160 samples and to statistically rank the importance of the features. CART® calculates a variable importance score for each feature based on the frequency and significance of its use as either a primary or surrogate splitter in the decision tree (Yu *et al.* 2006). Twenty-two of the initial 47 features were identified as primary or surrogate splitters and were subsequently considered for classification. The resulting feature ranking is shown in Table 4.2.

48

Table 4.2: Object features and importance ranks as derived from CTA

Type	Feature	Rank
Spectral:	Mean green	
	Mean red	
	Mean NIR	5
	Mean SWIR	11
	Mean brightness	9
	Std dev green	
	Std dev red	
	Std dev NIR	
	Std dev SWIR	
	Ratio green	8
	Ratio red	1
	Ratio NIR	14
	Ratio SWIR	21
	Maximum difference	7
VIs:	NDVI	3
	OSAVI	2
Texture:	GLCM Homogeneity	15
	GLCM Contrast	
	GLCM Dissimilarity	17
	GLCM Entropy	16
	GLCM Angular 2nd Movement	
	GLCM Correlation	
	GLCM Mean	10
	GLCM Std dev	
	GLDV Angular 2nd Movement	19
	GLDV Mean	18
	GLDV Contrast	
	GLDV Entropy	20
Geometric:	Area	
	Asymmetry	
	Border Length	
	Compactness	
	Density	
	Length	
	Length/Width	22
	Main Direction	
	Rectangular Fit	
	Roundness	
	Shape Index	
	Width	
Contextual:	Mean diff to neighbour green	
	Mean diff to neighbour red	
	Mean diff to neighbour NIR	12
	Mean diff to neighbour SWIR	
Transforms	HSI Hue	13
	HSI Saturation	3
	HSI Intensity	6

The class samples and segmentation were stored as ESRI shapefiles with the mean values of the 22 selected variables as attributes (ordered according their importance ranking). These shapefiles were inputted to the classification and accuracy assessment software.

4.4.3 Software development

A software system was developed using C++ and the Microsoft® Visual Studio® 2010 (Express edition) development environment to automate the processes of classification and accuracy assessment. Additional open-source libraries were acquired to complete the

implementation of the system. Libsvm 3.0 (Chang & Lin 2011) was used to implement one-against-one multiclass SVM. The ML and NN classifiers were implemented using the OpenCV 2.2 library (Bradski 2000) and the geospatial data abstraction library (GDAL) (GDAL Development Team 2010) was used for the manipulation of shapefiles and rasters.

The radial basis function kernel, as recommended by Hsu, Chang & Lin (2010), was selected for the SVM implementation. Appropriate values for the error parameter (C) and the kernel parameter (γ) were determined using a simple grid search and cross-validation approach. A coarse grid search was carried out on C = 2^{-5}, 2^{-3}, … , 2^{-15} and γ = 2^{-15}, 2^{-13}, … , 2^{3}, after which a finer grid search was performed based on the results of the first search (as recommended by Hsu, Chang & Lin (2010)). All data was scaled from -1 to 1 to prevent data with higher numerical ranges having greater effect than those with lower ranges (Hsu, Chang & Lin 2010).

4.4.4   Experiment workflow

The developed system was designed to test the performance of SVM, NN, ML classifiers as the number of features is increased. At the start of each experiment (program run), the object samples are randomly split into a training and a reference data set of equal size. The following steps are then repeated:

1. Select only the first feature in the shapefiles as input for classification.

2. Train the SVM, NN and ML classifiers using the training data set and the currently selected input feature space.

3. Use each of these classifiers to classify the unclassified shapefile, and perform automated accuracy assessments using the reference data set.

4. Add the next object feature to the current input feature space and repeat Steps 2 to 4 until all the object features (22) have been incorporated.

As mentioned in Section 4.4.2, the features in the shapefiles were ordered according to the importance scores obtained by the CTA as performed on the 160 class samples. Features are therefore incorporated into the experiment in the order of their importance (see Table 4.2). Results from 50 individual program runs were averaged, thus adopting a 50-fold repeated random sub-sampling validation with a 20/20 samples per class training/validation split. A second set of 50 program runs were also performed using a 10/30 samples per class training/validation split to investigate classifier specific relationships between feature dimensionality and training set size. Confusion matrices were investigated at each feature set

size iteration and used to compare the performance of the different classifiers concerning the specific land cover classes. The matrices were also used to calculate the producer's, user's and overall accuracies, as well as the kappa statistic, for each classifier and feature set size combination.

4.5   RESULTS AND DISCUSSION

The results of the investigation into the effect of feature dimensionality on object-based supervised classification performance using 20 training samples per class are summarized in the overall accuracy and kappa graphs (Figures 4.4 and 4.5). The accuracy and kappa graphs show a high degree of correlation and it is clear that, overall, SVM produced more accurate results compared to those of NN and ML. This finding supports those of other comparative studies that have found SVM to produce superior classification results (Dixon & Candade 2008; Foody & Mathur 2004a; Huang, Davis & Townshend 2002; Keuchel *et al.* 2003; Kavzoglu & Colkesen 2009; Oommen *et al.* 2008; Pal & Mather 2005; Szuster Chen & Borger 2011; Tzotsos & Argialas 2008).
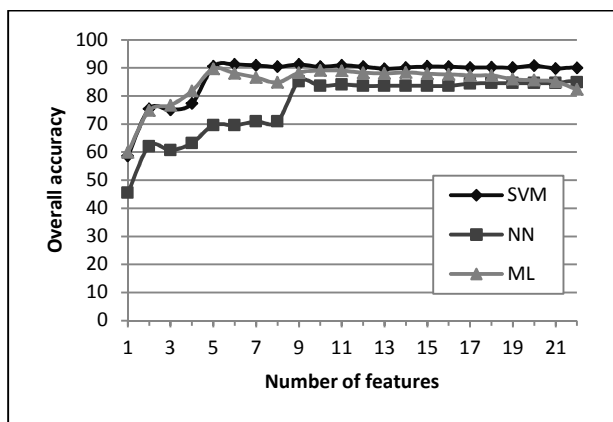

Figure 4.4: Average overall accuracy values for SVM, NN and ML with an increasing number of features (20 training samples per class).
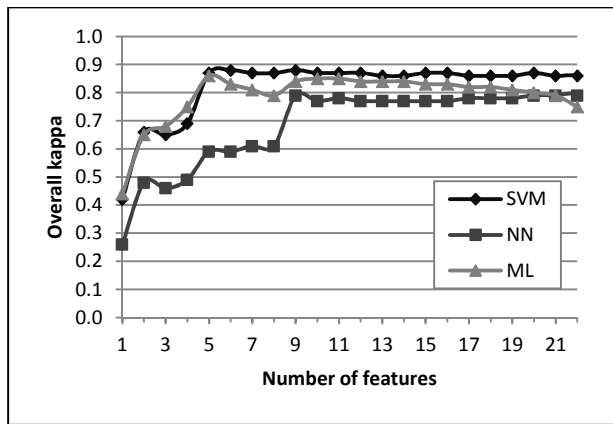
Figure 4.5: Average kappa values for SVM, NN and ML with an increasing number of features (20 training samples per class).

All three classifiers performed poorly (< 0.75 overall kappa) until the addition of the fifth feature (mean NIR). At this point the performance of SVM and ML increased dramatically – achieving overall kappa values of 0.87 and 0.86 respectively. The performance of NN, while also receiving a boost from the addition of the fifth feature (mean NIR), remained comparatively weak (0.59 overall kappa). NN's overall kappa improved significantly (to 0.79) after the addition of ninth feature (mean brightness). The graphs also indicate an improvement in performance for ML at this point (after ML's accuracies had dropped after the addition of features six through eight). The sudden increases in accuracy after the fifth and ninth features were added suggests that the importance of mean NIR and mean brightness was underestimated during the CTA feature selection process. The difference in response of the classifiers after incorporating these features (particularly mean brightness) indicates that some features are more important than others for specific classifiers and that CTA-based feature selection is not necessarily optimal for all classifiers.

After the inclusion of the mean NIR band, the overall performance of SVM is not significantly influenced by an increase in feature dimensionality. As the number of features was increased from five to 22, SVM's overall kappa remained between 0.86 and 0.88. NN's overall performance remained largely unaffected by the increase in feature dimensionality (from nine features) with overall kappa values ranging from 0.77 to 0.79. Conversely, ML's performance was negatively affected by the increase in dimensionality. While it performed consistently between nine and sixteen features (overall kappa ranging from 0.83 to 0.85), a gradual decrease in accuracy is observed when more features were used. ML's overall kappa dropped to 0.75 at 22 features – lower than NN's (0.79) at the same feature set size. The drop in accuracy is most likely due to the susceptibility of ML to the Hughes effect which has been well documented (Oommen *et al.* 2008; Pal & Mather 2005).

Confusion matrices were investigated to compare the performance of the different classifiers concerning specific land cover classes. Only one set of confusion matrices are provided due to space limitations. Confusion matrices for the classifiers at a feature set size of five are shown (Tables 4.3 to 4.5) as the addition of the fifth feature (mean NIR) proved significant for all classifiers. From Tables 4.3 to 4.5, it is clear that the *Water & shadow* class was the most accurately mapped by all the classifiers as it is the most distinct class and it is relatively easy to discern (Kavzoglu & Colkesen 2009). The more complex *Bare ground & built-up* class was mapped more accurately by SVM than the other classifiers. The very weak overall performance of NN at five features is mostly due its inability to correctly classify this class. This indicates that, for NN, the first 5 features as ranked by the CTA are not sufficient for identifying *Bare ground & built-up* areas. Only after the inclusion of the mean brightness feature (nine features) could NN classify this class more accurately. Compared to SVM, ML produced more commission errors for this class, indicating slight over classification. This is consistent with the findings of Dixon & Candade (2008) that ML significantly over classified their *Urban* class (which would be similar to the *Bare ground & built-up* class used in this study) compared to SVM and ANN.

Table 4.3: SVM confusion matrix for five features

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%[a]	EO%[a]
Trees & shrubs	24676	2341	149	337	27502	89.7	10.3
Forbs/herb/gram	2668	24185	160	0	27014	89.5	10.5
Bare gr/built up	285	211	8091	960	9548	84.7	15.3
Water & shadow	179	113	507	18019	18818	95.8	4.2
TOTALS	27808	26850	8907	19317	82881		
CA%[a]	88.7	90.1	90.8	93.3			
EC%[a]	11.3	9.9	9.2	6.7			
Overall accuracy:	90.5						
Overall kappa:	0.87						

[a] PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 4.4: NN confusion matrix for five features

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%[a]	EO%[a]
Trees & shrubs	15485	3297	8719	0	27502	56.3	43.7
Forbs/herb/gram	4417	19685	1716	1195	27014	72.9	27.1
Bare gr/built up	2239	1026	5340	943	9548	55.9	44.1
Water & shadow	0	341	1282	17195	18818	91.4	8.6
TOTALS	22141	24349	17057	19334	82881		
CA%[a]	69.9	80.8	31.3	88.9			
EC%[a]	30.1	19.2	68.7	11.1			
Overall accuracy:	69.6						
Overall kappa:	0.59						

[a] PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

Table 4.5: ML confusion matrix for five features

	Tr & shr	frb/hrb/gr	Bare/built	Wt & sh	TOTALS	PA%[a]	EO%[a]
Trees & shrubs	24581	2921	0	0	27502	89.4	10.6
Forbs/herb/gram	1649	24584	780	0	27014	91.0	9.0
Bare gr/built up	293	38	8050	1167	9548	84.3	15.7
Water & shadow	1191	0	537	17090	18818	90.8	9.2
TOTALS	27715	27542	9367	18257	82881		
CA%[a]	88.7	89.3	85.9	93.6			
EC%[a]	11.3	10.7	14.1	6.4			
Overall accuracy:	89.7						
Overall kappa:	0.86						

[a] PA = Producer's accuracy; EO = Errors of omission; CA = Consumer's accuracy; EC = Errors of commission

The *Trees & shrubs* and the *Forbs, herbland & graminoid* classes are spectrally similar and resulted in much classification confusion throughout the experiment. When the first five features were used, ML was slightly better (9.3% confusion[1]) than SVM (10.3% confusion) at distinguishing between these classes while NN (21.9% confusion) was far less successful (see Tables 4.3 to 4.5). These findings are in contrast with those of Dixon & Candade (2008), Kavzoglu & Colkesen (2009) and Szuster, Chen & Borger (2011) who have shown SVM to be superior at discerning spectrally similar classes. It should be noted, however, that these pixel-based studies used only spectral band values as classification input. The object-based nature of this study, as well as the object features selected through CTA, might have

---

[1] The percentage confusion between the two classes was calculated by adding the number of misclassifications between them and dividing by the sum of the reference pixels for the two classes, e.g. the percentage confusion for SVM at five features (Table 4.3) was calculated as follows: (2341+2668)/(24676+24185)*100 = 10.3%.

contributed to ML achieving slightly better discrimination between the spectrally similar *Trees & shrubs* and *Forbs, herbland & graminoid* classes than SVM.

The general findings regarding specific class accuracies held true for most feature set sizes after 5 features, however, some variations were notable. The kappa and accuracy graphs revealed a decline in ML's performance when features six to eight (HSI intensity, maximum difference and ration green) were included. Inspection of the corresponding confusion matrices showed this decline to be caused by increased over classification of the *Trees & shrubs* and *Bare ground & built up* classes. Since ML's best results (0.86 overall kappa) were obtained before the inclusion of these features (despite the increase in accuracy that occurs after mean brightness is included at nine features used), it is likely that the HSI intensity, maximum difference and ratio green features negatively affected parameter estimation and were not suitable for ML classification despite being ranked as relatively important by the CTA. These features did, however, not negatively affect the SVM and NN classifiers. This indicates that the influence of certain features on supervised classification may be classifier specific. Furthermore, SVM showed a significant improvement in identifying *Bare ground & built up* and *Water & shadow areas* after the inclusion of GLCM Homogeneity at 15 features. This was, however, at the expense of SVM's ability to discern *Trees & shrubs* and *Forbs, herbland & graminoids.* ML's and NN's results remained largely unchanged by the addition of the 15th feature (GLCM Homogeneity).

The experiment was repeated with a smaller training set size (10 samples per class) to gain insights into classifier specific relationships between feature dimensionality and training set size. The overall accuracy and kappa results are summarised in Figures 4.6 and 4.7. As expected, all three classifiers produced less accurate results when fewer training samples were used. The shapes of the SVM and NN graphs in Figures 4.6 and 4.7 are not much different when compared to the 20 samples per class graphs (Figures 4.4 and 4.5). Again the SVM stabilized after the addition of the fifth feature (mean NIR), while NN stabilized after the ninth feature (mean brightness) was added. This indicates that, although the smaller training set size influenced the overall performance of the classifiers, the classifier specific influence of certain features was consistent regardless of the number of training samples used. The shape of the ML graphs in Figures 4.6 and 4.7 are initially similar to those generated from the 20 sample per class experiment, but became unstable when more than five features were used, resulting in a general decline in overall accuracies. This is likely due to poor parameter estimation often associated with small training set size and increased dimensionality –

exposing ML's reliance on sufficient training data and its susceptibility to the Hughes effect (Oommen *et al*. 2008; Pal & Mather 2005).
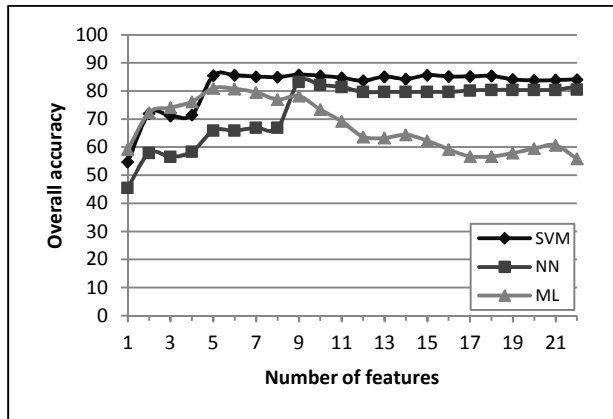
Figure 4.6: Average overall accuracy values for SVM, NN and ML with an increasing number of features (10 training samples per class).
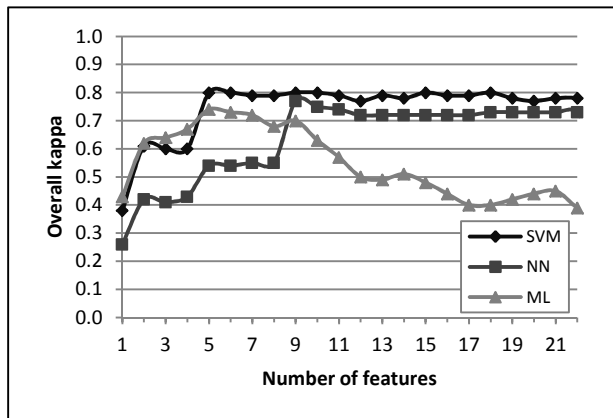
Figure 4.7: Average kappa values for SVM, NN and ML with an increasing number of features (10 training samples per class).

The overall results indicate, for the data and the classification scheme used in this study, that SVM generally produces superior classification results when compared to ML and NN. For both the 20 and 10 training samples per class experiment, neither SVM's nor NN's performance was significantly (negatively) affected by an increase in feature dimensionality. ML's ability to perform under conditions of small training set sizes and large feature dimensionalities was shown to be limited. Given sufficient training data and using few selected features, ML outperformed NN. This finding suggests that NN as the weakest of the three classifiers for GEOBIA under such conditions. This should be of particular interest to eCognition users, as the latest version of the software (8.7) allows users to choose between SVM, CART and Bayes classifiers as alternatives to the commonly applied NN classifier.

4.6 CONCLUSIONS

It is well known that the incorporation of additional variables (e.g. vegetation indices, image transforms, textural information, contextual information and ancillary data) in the land cover classification workflow can improve the accuracy of object-based supervised classifiers (Heinl *et al.* 2009; Lu & Weng 2007). Although GEOBIA provides an ideal platform for the inclusion of such features, the number of available training samples is generally less for object-based problems than for traditional pixel-based approaches. This study compared the performance of SVM, NN and ML for object-based land cover classification, with particular attention to increasing the number of input features. SVM generally produced superior classification results. SVM and NN were not significantly (negatively) affected by an increase in feature dimensionality. In contrast ML's well-known susceptibility to the Hughes effect and its reliance on a sufficient number of training data was confirmed in a GEOBIA context. The results also revealed that some features are more important than others for specific classifiers and that CTA-based feature selection is not necessarily optimal for all classifiers. The nature of the data, the desired classification output and the specific classifier should therefore be considered carefully when additional features are incorporated.

This study adopted a very simple four-class land cover classification scheme. More research is needed to investigate the effect of feature dimensionality on the performance of SVM, NN and ML when more complex classification schemes are used. An analysis that probes which combinations of additional features provide the greatest improvements in accuracy is recommended. Such experiments should preferably be applied to a variety of imagery, additional features and other ancillary data to gain insight into the performance of supervised classifiers for specific types of classification problems. This will contribute significantly toward finding automatic (or at least cost-effective) solutions for mapping land cover over large areas using high and very-high resolution imagery.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

A summary and an assessment of the findings of the two experiments described in the two previous chapters are presented in this chapter. The research aim and objectives are revisited, conclusions are drawn and recommendations for future work are made.

5.1 ASSESSMENT OF THE IMPACT OF TRAINING SET SIZE AND FEATURE DIMENSIONALITY

Two experiments were designed to evaluate the impact of training set size and feature dimensionality on object-based land cover classification respectively. The salient findings of the experiments are presented in this section.

5.1.1 Number of training samples

Supervised classifiers as a rule produce superior classifications when using larger training sets (Foody & Mathur 2004a; 2004b; Mather 2004; Pal & Mather 2003). This held true in the first experiment (Chapter 3) in which all three classifiers achieved higher accuracies when using more training samples. SVM consistently produced the best classification results. The ML classifier performed significantly weaker than SVM and NN when very small training sets were used, most likely due to insufficient data being available for the accurate estimation of statistical parameters. ML's overall accuracy results became comparable to those of SVM when 12 or more samples per class were used while the NN classifier performed comparatively poorly throughout the experiment. SVM performed significantly better under small training set size conditions. Adopting a training data collection approach optimized for SVM classification might further improve the classifier's performance under such conditions. The ability of SVM and ML to distinguish between the spectrally similar *Trees & shrubs* and *Forbs, herbland & graminoids* classes and to successfully classify the complex *Bare ground & built up* class improved significantly as the training sets grew larger. SVM was, however, much better at identifying *Bare ground & built up* areas than ML while NN's poor overall performance was mainly due to its inability to correctly classify this class.

5.1.2 Number of features

Section 2.2.4.2 reported on the use of additional features improving classification results. It is known that increasing feature dimensionality, while maintaining a constant training set size, may cause a drop in classification accuracy – a phenomenon known as the Hughes effect. In the experiment discussed in Chapter 4, fixed training sets (20 and 10 samples per class) were

used while incrementally adding object features as input to the classification. Features were incorporated in the order of their importance rank as determined by CTA.

SVM generally produced superior classification results when compared to ML and NN. Neither SVM nor NN were significantly negatively affected by an increase in feature dimensionality. ML, however, suffered from the Hughes effect – much more so when fewer training samples (10 per class) were used. ML's ability to perform under conditions of small training set sizes and large feature dimensionalities was therefore shown to be limited. However, given sufficient training data and using a small set of carefully selected features, ML outperformed NN and produced results comparable to those of SVM.

Classifier-specific influences of certain features were also observed. Considering the use of 20 training samples per class, all classifiers performed poorly (< 0.75 overall kappa) until the fifth feature (mean NIR) was added. SVM's and ML's overall kappa values increased drastically at this point (to 0.87 and 0.86 respectively). NN's overall kappa remained comparatively low (0.59) mostly due to its inability to correctly classify the *Bare ground & built up* class. Only after the inclusion of the mean brightness feature (nine features) could NN classify this class more accurately, which significantly improved its overall performance (0.79 overall kappa). ML's performance declined when features six to eight (HSI intensity, maximum difference and ration green) were added and SVM showed a significant improvement in identifying *Bare ground & built up* and *Water & shadow areas* after the inclusion of GLCM Homogeneity at 15 features. The classifier-specific influence of certain features remained consistent when fewer training samples (10 per class) were used (apart from the rapid decline in ML's performance when more than five features were used). These results indicate that some features are more important than others for specific classifiers and that the CTA-based feature ranking is not necessarily optimal for all classifiers.

5.2 EVALUATION OF THE CLASSIFIERS FOR OBJECT-BASED CLASSIFICATION

The investigation into the influence of training set size and feature dimensionality on object-based SVM, NN and ML classifiers was limited to a single SPOT 5 subscene, fixed ranges of training samples (4 to 20 per class) and features (1 to 22), and a basic four-class classification scheme. Despite these limitations the results of the experiments shed light on the potential of the classifiers for object-based land cover classification. ML's suitability for GEOBIA was shown to be comparatively limited. The first experiment showed that ML is unsuitable for use when very small training sets are available and the second experiment revealed the classifier's

inability to handle large feature dimensionalities. ML did compete with SVM, and outperformed NN, when a sufficient number of training samples and few features were used. However, small training sets and the use of additional features are common in GEOBIA (Hay & Castilla 2006, 2008; Tzotsos & Argialas 2008). Ensuring that sufficient samples are available and implementing some form of feature selection or data reduction techniques is consequently crucial should ML be used for object-based land cover classification.

One can conclude that, of the three classifiers evaluated in this study, SVM holds the most potential for object-based classification. SVM consistently outperformed the other classifiers, was not negatively affected by increases in feature dimensionality and was significantly better at identifying the spectrally-complex *Bare ground & built up* class. Such advantages could be even more telling should more complex classification schemes be adopted. SVM's good performance using few training areas and its robustness to large input dimensionalities makes it ideal for detailed GEOBIA classification. NN, an option most commonly chosen for object-based supervised classification (Campbell 2006), performed poorly compared to SVM. NN was more accurate than ML only under conditions of small training set sizes and large feature dimensionalities. The nature of the data, the desired classification output, available features and the feasibility of adopting a particular feature selection technique should therefore be carefully considered when choosing a classifier.

5.3   REVISITING THE RESEARCH AIM AND OBJECTIVES

The aim of the reported research was to evaluate SVM, ML and NN classifiers for object-based land cover classification and to investigate the effects of the training set size and feature dimensionality on classification accuracies. Such comparative investigations are driven by the need for finding more cost-effective ways of producing sufficiently accurate land cover information, hence the goal of this study to gain insights into the potential of each of the three classifiers for GEOBIA. Individual objectives were set to achieve the overarching aim and this section discusses the degree to which each was achieved.

A literature review (the first objective) focusing on the different approaches to RS image classification and the various factors that influence classification accuracies was presented in Chapter 2. This exercise revealed that SVMs hold considerable potential for RS image classification given their good performance with limited training samples and their robustness to high feature dimensionalities. However, most of the existing literature focused on traditional pixel-based classification, and investigations into its potential for GEOBIA were scarce.

Acquiring and preparing suitable satellite imagery constituted the second objective. Chapters 3 and 4 documented the properties of the acquired SPOT 5 image and the preprocessing to which it was subjected to before to commencing the classification experiments.

The third objective was to develop a software system to perform object-based SVM, NN and ML classification and automated accuracy assessment. The system was developed using C++ and the Microsoft Visual Studio 2010 development environment. Additional open-source libraries such as Libsvm 3.0 (Chang & Lin 2011), OpenCV 2.2 (Bradski 2000) and GDAL 1.7.0 (GDAL Development Team 2010) were acquired and used to complete the system.

The developed program was extended to implement an experiment to investigate the effect of the size of training sets on SVM, NN and ML classifiers (Chapter 3). An additional python script was developed to summarize the results obtained from multiple program runs and fully complete the experiment (the fourth objective). The python script and modifications to the program were used to address the fifth objective, namely to implement an experiment to investigate the influence of feature dimensionality on object-based SVM, NN and ML classifiers (Chapter 4). More details of the developed software components are presented in Appendix A and the source code (in digital format) has been made available online (see Appendix B). The sixth objective, the presentation, discussion and interpretation of the results of each experiment is addressed in Chapters 3 and 4 respectively and complemented in Chapter 5 with a summary and evaluation of the core findings.

5.4   CONCLUSION

Two experiments were conducted with the goal of gaining insights into the potential of SVM, NN and ML for object-based land cover classification. The first evaluated the impact of training set size on classification accuracies and the second investigated the capacity of each classifier to effectively incorporate additional object features. The results demonstrate that SVM holds the most potential for object-based classification in that it consistently outperformed NN and ML. Increases in training set size produced higher overall accuracies for all three classifiers. It was also revealed that certain features are more important than others for specific classifiers and that CTA-based feature selection is not necessarily optimal for all classifiers. ML's inability to perform under conditions of small training sets and large feature dimensionalities was also exposed in a GEOBIA context.

The availability of increasingly large volumes of remotely sensed imagery at various spatial and temporal scales is an attractive source of detailed and up-to-date land cover information. Automatic and semi-automatic processing of RS imagery is currently limited and finding

cost-effective ways to derive accurate land cover information are needed to make effective use of the large volumes of available RS data. The comparison of image classifiers and the investigation into the factors that influence classification accuracies, as presented in this thesis, contribute toward finding optimal solutions for classification tasks. SVM was found to be a robuster and more effective classifier than ML and NN for GEOBIA by consistently outperforming the latter classifiers under varying conditions of training set size and feature dimensionality. The conditions under which each classifier performed best were also established for the given data and classification scheme used. The developed software was designed with the capability to process new data with little or no modification. This presents opportunities for conducting case studies that compare the capabilities of these classifiers using a variety of imagery, classification schemes, additional features and other ancillary data. It is recommended that the presented experiments be extended to more thoroughly investigate which combinations of additional features produce the highest accuracies for each classifier under specific conditions. Knowledge of the optimal number of training samples and combinations of features for specialized GEOBIA classification problems will be invaluable to increase the level of automation, while improving the cost-effectiveness of mapping land cover over large areas using high- and very-high-resolution imagery.

REFERENCES

Addink EA, De Jong SM, Davis SA, Dubyanskiy V, Burdelov LA & Leirs H 2010. The use of high-resolution remote sensing for plague surveillance in Kazakhstan. *Remote Sensing of Environment* 114: 674-681.

Addink EA, De Jong SM & Pebesma EJ 2007. The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing* 73: 905-912.

Aitkenhead MJ & Aalders IH 2011. Automating land cover mapping of Scotland using expert system and knowledge integration methods. *Remote Sensing of Environment* 115: 1285-1295.

Albert TH 2002. Evaluation of remote sensing techniques for ice-area classification applied to the tropical Quelccaya ice cap, Peru. *Polar Geography* 26: 210-226.

Amarsaikhan D, Blotevogel HH, van Genderen JL, Ganzorig M, Gantuya R & Nergui B 2010. Fusing high-resolution SAR and optical imagery for improved urban land cover study and classification. *International Journal of Image and Data Fusion* 1: 83-97.

Baatz M, Hoffmann C & Willhauk G 2008. Progressing from object-based to object-oriented image analysis. In Blaschke T, Lang S & Hay GJ (eds) *Object-Based Image Analysis: Spatial Concepts for Knowledge-driven Remote Sensing Applications,* 29-42. Berlin: Springer.

Bauer T & Steinnocher K 2001. Per-parcel land use classification in urban areas applying a rule-based technique. *GeoBIT/GIS* 6: 24-27.

Baraldi A, Durieux L, Simonetti D, Conchedda G, Holecz F & Blonda P 2010. Automatic Spectral-Rule-Based Preliminary Classification of Radiometrically Calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 Imagery—Part I: System Design and Implementation. *IEEE Transactions on Geoscience and Remote Sensing* 48: 1299-1325.

Benz UC, Hofmann P, Willhauck G, Lingenfelder I & Heynen M 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing* 58: 239-258.

Berberoglu S, Curran PJ, Lloyd CD & Atkinson PM 2007. Texture classification of Mediterranean land cover. *International Journal of Applied Earth Observation and Geoinformation* 9: 322-334.

Berberoglu S, Lloyd CD, Atkinson PM & Curran PJ 2000. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. *Computers & Geosciences* 26: 385-396.

Blaschke T 2010. Object-based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2-16.

Blaschke T & Lang S 2006. Object-based image analysis for automated information extraction - A synthesis. Paper delivered at the MAPPS/ASPRS Speciality Conference, San Antonio.

Blaschke T, Lang S, Lorup E, Strobl J & Zeil P 2000. Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications. In Cremers A & Greve K (eds) *Environmental information for planning politics and the public*, vol. 2, 555-570. Marburg: Metropolis Verlag.

Blaschke T, Lang S & Hay GJ (eds) 2008. *Object-based Image Analysis: Spatial Concepts for Knowledge-driven Remote Sensing Applications*. Berlin: Springer.

Blaschke T & Strobl J 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS - Zeitschrift Für Geoinformationssysteme* 14: 12-17.

Bradski G 2000. The OpenCV library. *Dr.Dobb's Journal of Software Tools* 25: 120, 122-125.

Brown de Colstoun CE, Story MH, Thompson C, Commisso K, Smith TG & Irons JR 2003. National park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. *Remote Sensing of Environment* 85: 316-327.

Burges CJC 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2: 121-167.

Calvo S, Ciraolo G & Loggia GL 2003. Monitoring *Posidonia oceanica* meadows in a Mediterranean coastal lagoon (Stagnone, Italy) by means of neural network and ISODATA classification methods. *International Journal of Remote Sensing* 24: 2703-2716.

Campbell J 2006. *Introduction to Remote Sensing*. London: Taylor & Francis.

Camps-Valls G & Bruzzone L 2005. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 43: 1351-1362.

Camps-Valls G, Gómez-Chova L, Calpe-Maravilla J, Martin-Guerrero JD, Soria-Olivas E, Alonso-Chorda L & Moreno J 2004. Robust support vector method for hyperspectral data

classification and knowledge discovery. *IEEE Transactions on Geoscience and Remote Sensing* 42: 1530-1542.

Camps-Valls G, Gómez-Chova L, Muñoz-Marí J, Vila-Francés J & Calpe-Maravilla J 2006. Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* 3: 93-97.

Candade N & Dixon B 2004. Multispectral classification of Landsat images: a comparison of support vector machine and neural network classifiers. *ASPRS Annual Conference Procedings, May 2004, Denver, Colarado* [online]. Available from: http://www.usfsp.edu/bdixon/presentations/2004/Dixon_paper_ASPRS_Denver2004.pdf [Accessed 14 September 2011].

Carleer AP & Wolff E 2006. Urban land cover multi-level region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing* 27: 1035-1051.

Chang C & Lin C 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems Technology* 2: 27:1-27:27.

Chen CH & Ho P 2008. Statistical pattern recognition in remote sensing. *Pattern Recognition* 41: 2731-2741.

Chen Y, Su W, Li J & Sun Z 2009. Hierarchical object-oriented classification using very high resolution imagery and LIDAR data over urban areas. *Advances in Space Research* 43: 1101-1110.

Christianini N & Shawe-Taylor J 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press.

Chubey MS, Franklin SE & Wulder MA 2006. Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering & Remote Sensing* 72: 383-394.

Chuvieco E & Huete A 2010. *Fundamentals of Satellite Remote Sensing*. Boca Raton: CRC Press.

Congalton G & Green K 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton: CRC Press.

Cover T & Hart P 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13: 21-27.

Cracknell AP 1998. Synergy in remote sensing–what's in a pixel? *International Journal of Remote Sensing* 19: 2025-2047.

Dixon B & Candade N 2008. Multispectral landuse classification using neural networks and support vector machines: One or the other, or both? *International Journal of Remote Sensing* 29: 1185-1206.

Duda T & Canty M 2002. Unsupervised classification of satellite imagery: Choosing a good algorithm. *International Journal of Remote Sensing* 23: 2193-2212.

Duro DC, Franklin SE & Dubé MG 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment* 118: 259-272.

Fisher P 1997. The pixel: A snare and a delusion. *International Journal of Remote Sensing* 18: 679-685.

Foody GM 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80: 185-201.

Foody GM 2009. On training and evaluation of SVM for remote sensing applications. In Camps-Valls G & Bruzzone L (eds) *Kernel Methods for Remote Sensing Data Analysis,* 85-104. Chichester: Wiley.

Foody GM & Mathur A 2004a. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42: 1335-1343.

Foody GM & Mathur A 2004b. Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. *Remote Sensing of Environment* 93: 107-117.

Foody GM & Mathur A 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment* 103: 179-189.

Fourie C 2011. A one-class object-based system for sparse geographic feature identification. Master's thesis. Stellenbosch: Stellenbosch University, Department of Geography and Environmental Studies.

Fourie C, Van Niekerk A & Mucina L 2011. Optimising a one-class SVM for geographic object-based novelty detection. Paper delivered at the AfricaGeo conference, Cape town.

Fourie C, Van Niekerk A & Mucina L 2012. Semi-automated segment generation for geographic novelty detection using edge and area metrics. *South African Journal of Geomatics* 1: 113-148.

Friedl MA & Brodley CE 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* 61: 399-409.

Gao J 2009. *Digital Analysis of Remotely Sensed Imagery.* New York: McGraw Hill.

GDAL Development Team 2010. *GDAL - Geospatial Data Abstraction Library, Version 1.7.0* [online]. Open Source Geospatial Foundation. Available from: http://www.gdal.org/ [Accessed 4 August 2011].

Gehler PV & Schölkopf B 2009. An introduction to Kernel learning algorithms. In Camps-Valls G & Bruzzone L (eds) *Kernel Methods for Remote Sensing Data Analysis*, 25-45. Chichester: Wiley.

Gómez-Chova L, Calpe J, Soria E, Camps-Valls G, Martin JD & Moreno J 2003. CART-based feature selection of hyperspectral images for crop cover classification. Paper delivered at the International Conference of Image Processing (ICIP), Barcelona.

Gualtieri JA & Cromp RF 1998. Support vector machines for hyperspectral remote sensing classification. Paper delivered at the 27th AIPR Workshop: Advances in Computer Assisted Recognition, Washington DC.

Hamel L, 2009. *Knowledge Discovery with Support Vector Machines*. Hoboken: Wiley.

Hansen M, Dubayah R & Defries R 1996. Classification trees: An alternative to traditional land cover classifiers. *International Journal of Remote Sensing* 17: 1075-1081.

Hay GJ & Castilla GC 2006. Object-based image analysis: Strengths, weaknesses opportunities and threats (SWOT). Paper delivered at the International Conference on Object-based Image Analysis, Salzburg.

Hay GJ & Castilla GC 2008. Geographic object-based image analysis (GEOBIA): A new name for a new discipline? In Blaschke T, Lang S & Hay GJ (eds) *Object-based Image Analysis: Spatial Concepts for Knowledge-driven Remote Sensing Applications,* 75-89. Berlin: Springer.

Heinl M, Walde J, Tappeiner G & Tappeiner U 2009. Classifiers vs input variables—The drivers in image classification for land cover mapping. *International Journal of Applied Earth Observation and Geoinformation* 11: 423-430.

Hermes L, Frieauff D, Puzicha J & Buhmann JM 1999. Support vector machines for land usage classification in Landsat TM imagery. *Proceedings of IGARSS'99* 1: 348-350.

Herold M, Liu X & Clarke KC 2003. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering & Remote Sensing* 69: 991-1001.

Hofmann P, Strobl J & Blaschke T 2008. A method for adapting global image segmentation methods to images of different resolutions. Paper delivered at the GEOBIA 2008 conference, Calgary.

Hsu C, Chang C & Lin C 2010. A practical guide to support vector classification [online]. Available from: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf [accessed 21 July 2011].

Huang C, Davis LS & Townshend JRG 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* 23: 725-749.

Hughes G 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory* 14: 55-63.

Hubert-Moy L, Cotonnec A, Le Du L, Chardin A & Perez P 2001. A comparison of parametric classification procedures of remotely sensed data applied on different landscape units. *Remote Sensing of Environment* 75: 174-187.

Jain AK, Duin RPW & Mao J 2000. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22: 4-37.

Kavzoglu T & Colkesen I 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation* 11: 352-359.

Kavzoglu T & Mather PM 2003. The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing* 24: 4907-4938.

Keuchel J, Naumann S, Heiler M & Siegmund A 2003. Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data. *Remote Sensing of Environment* 86: 530-541.

Kim M, Madden M & Warner TA 2009. Forest type mapping using object-specific texture measures from multispectral Ikonos imagery: Segmentation quality and image classification issues. *Photogrammetric Engineering & Remote Sensing* 75: 819-829.

Kotsiantis SB 2007. Supervised machine learning: a review of classification techniques. *Informatica* 31: 249-268.

Laliberte AS, Browning DM & Rango A 2010. Feature selection methods for object-based classification of sub-decimeter digital aerial imagery. The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences: GEOBIA 2010, 29 Jun – July 2, 2010, Ghent, Belgium, ISPRS, Vol. No. XXXVIII-4/C7.

Laliberte AS, Browning DM & Rango A 2012. A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L imagery. *International Journal of Applied Earth Observation and Geoinformation* 15: 70-78.

Laliberte AS, Fredrickson EL & Rango A 2007. Combining decision trees with hierachical object-oriented image analysis for mapping arid rangelands. *Photogrammetric Engineering & Remote Sensing* 73: 197-207.

Laliberte AS, Koppa J, Fredrickson EL & Rango A 2006. Comparison of nearest neighbor and rule-based decision tree classification in an object-oriented environment. Paper delivered at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Denver.

Lang MW, Kasischke ES, Prince SD & Pittman KW 2008. Assessment of C-band synthetic aperture radar data for mapping and monitoring Coastal Plain forested wetlands in the Mid-Atlantic Region, U.S.A. *Remote Sensing of Environment* 112: 4120-4130.

Li H, Gu H, Han Y & Yang J 2008. Object-oriented classification of polarimetric SAR imagery based on statistical region merging and support vector machine. Paper delivered at the International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Beijing.

Li H, Gu H, Han Y & Yang J 2010. Object-oriented classification of high-resolution remote sensing imagery based on an improved colour structure code and a support vector machine. *International Journal of Remote Sensing* 31: 1453-1470.

Lillesand TM, Kiefer RW & Chipman JW 2008. *Remote Sensing and Image Interpretation.* Hoboken, NJ: John Wiley & Sons.

Liu D & Xia F 2010. Assessing object-based classification: Advantages and limitations. *Remote Sensing Letters* 1: 187-194.

Liu X, Skidmore AK & Van Oosten H 2002. Integration of classification methods for improvement of land-cover map accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing* 56: 257-268.

Lizarazo I 2008. SVM-based segmentation and classification of remotely sensed data. *International Journal of Remote Sensing* 29: 7277-7283.

Lu D & Weng Q 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 28: 823-870.

Mallinis G, Koutsias N, Tsakiri-Strati M & Karteris M 2008. Object-based classification using QuickBird imagery for delineating forest vegetation polygons in a Mediterranean test site. *ISPRS Journal of Photogrammetry and Remote Sensing* 63: 237-250.

Mansor S, Hong WT & Shariff ARM 2002. Object oriented classification for land cover mapping. *Proceedings of the 23rd Asian Conference on Remote Sensing, 25-29 November, Kathmandu, Nepal* [online]. Available at: http://www.gisdevelopment.net/application/environment/overview/envo0010pf.htm [Accessed 21 March 2012].

Marpu PR, Niemeyer I, Nussbaum S & Gloaguen R 2008. A procedure for automatic object-based classification. In Blaschke T, Lang S & Hay GJ (eds) *Object-based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications,* 169-184. Berlin: Springer.

Mas JF & Flores JJ 2008. The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing* 29: 617-663.

Mather PM 2004. *Computer Processing of Remotely-Sensed Images: An Introduction.* Chichester: Wiley.

Melgani F & Bruzzone L 2002. Support vector machines for classification of hyperspectral remote-sensing images. *IEEE Geoscience and Remote Sensing Symposium 2002* 1: 506-508.

Melgani F & Bruzzone L 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42: 1778-1790.

Meng Y & Peng S 2009. Object-oriented building extraction from high-resolution imagery based on fuzzy SVM. Paper delivered at the International Conference on Information Engineering and Computer Science (ICIECS), Wuhan, China.

Mercier G & Lennon M 2003. Support vector machines for hyperspectral image classification with spectral-based kernels. Paper delivered at the IEEE Geoscience and Remote Sensing Symposium (IGARSS), Toulouse.

Mountrakis G, Im J & Ogole C 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66: 247-259.

Nolin AW & Payne MC 2007. Classification of glacier zones in western Greenland using albedo and surface roughness from the multi-angle imaging spectroradiometer (MISR). *Remote Sensing of Environment* 107: 264-275.

Oommen T, Misra D, Twarakavi N, Prakash A, Sahoo B & Bandopadhyay S 2008. An objective analysis of support vector machine based classification for remote sensing. *Mathematical Geosciences* 40: 409.

Otukei JR & Blaschke T 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation* 12, Supplement 1: S27-S31.

Pal M & Mather PM 2003. An assessment of effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment* 86: 554-565.

Pal M & Mather PM 2004. Assessment of the effectiveness of support vector machines for hyperspectral data. *Future Generation Computer Systems* 20: 1215-1225.

Pal M & Mather PM 2005. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing* 26: 1007-1011.

Pohl C & Van Genderen JL 1998. Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing* 19: 823-854.

Roli F & Fumera G 2000. Support vector machines for remote-sensing image classification. *Proceedings of SPIE* 4170: 160 -166.

Shaban MA & Dikshit O 2001. Improvement of classification in urban areas by the use of textural features: The case study of Lucknow city, Uttar Pradesh. *International Journal of Remote Sensing* 22: 565-593.

Skidmore AK, Watford F, Luckananurug P & Ryan PJ 1996. An operational GIS expert system for mapping forest soils. *Photogrammetric Engineering & Remote Sensing* 62: 501-511.

Smith AMS, Wooster MJ, Powell AK & Usher D 2002. Texture-based feature extraction: Application to burn scar detection in Earth observation satellite sensor imagery. *International Journal of Remote Sensing* 23: 1733-1739.

Steinwart I & Christmann A 2008. *Support Vector Machines*. New York: Springer.

Stephenson G 2010. A comparison of supervised and rule-based object-orientated classification for forest mapping. Master's thesis. Stellenbosch: Stellenbosch University, Department of Geography and Environmental Studies.

Stephenson G & Van Niekerk A 2009. A cost-effective, rule-based technique to improve forestry inventory on a national scale. Paper delivered at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Cape Town.

Szuster BW, Chen Q & Borger M 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography* 31: 525-532.

Tadjudin S & Landgrebe D 1998. Classification of high dimensional data with limited training samples. Doctoral dissertation. West Lafayette, Indiana: Purdue University, School of Electrical and Computer Engineering.

Tansey K, Chambers I, Anstee A, Denniss A & Lamb A 2009. Object-oriented classification of very high resolution airborne imagery for the extraction of hedgerows and field margin cover in agricultural areas. *Applied Geography* 29: 145-157.

Tapia R, Stein A & Bijker W 2005. Optimization of sampling schemes for vegetation mapping using fuzzy classification. *Remote Sensing of Environment* 99: 425-433.

Tseng M, Chen S, Hwang G & Shen M 2008. A genetic algorithm rule-based approach for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 63: 202-212.

Tzotsos A & Argialas D 2008. Support vector machine classification for object-based image analysis. In Blaschke T, Lang S & Hay GJ (eds) *Object-based Image Analysis: Spatial Concepts for Knowledge-driven Remote Sensing Applications,* 663-677. Berlin: Springer.

Tzotsos A, Karantzalos K & Argialas D 2011. Object-based image analysis through nonlinear scale-space filtering. *ISPRS Journal of Photogrammetry and Remote Sensing* 66: 2-16.

Van Coillie FMB, Verbeke LPC & De Wulf RR 2007. Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sensing of Environment* 110: 476-487.

Vapnik VN 1995. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Waske B, Fauvel M, Benediktsson JA & Chanussot J 2009. Machine learning techiques in remote sensing data analysis. In Camps-Valls G & Bruzzone L (eds) *Kernel Methods for Remote Sensing Data Analysis,* 3-24. Chichester: Wiley.

Watanachaturaporn P, Arora MK & Varshney PK 2008. Multisource classification using support vector machines: An empirical comparison with decision tree and neural network classifiers. *Photogrammetric Engineering and Remote Sensing* 74: 293-246.

Wu B, Xiong Z, Chen Y & Zhao Y 2009. Classification of QuickBird image with maximal mutual information feature selection and support vector machine. *Procedia Earth and Planetary Science* 1: 1165-1172.

Yu Q, Gong P, Clinton N, Biging G, Kelly M & Schirokauer D 2006. Object-based detalied vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering & Remote Sensing* 72: 799-811.

Zhang X, Feng X & Jiang H 2010. Object-oriented method for urban vegetation mapping using IKONOS imagery. *International Journal of Remote Sensing* 31: 177-196.

Zhang Y & Mishra RK 2012. A review and comparison of commercially available pan-sharpening techniques for high resolution satellite image fusion. Paper delivered at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich.

APPENDIX A: SOFTWARE COMPONENTS

The development of software capable of performing the two experiments was a major component of the research. Figure A1 gives a simplified view of the components and workflow of the C++ script created for the first experiment (*samplesTest.cpp*). A modified version of this script (*featuresTest.cpp*) was used to implement the second experiment. An additional C++ object class was created (*ErrorMatrix.cpp* and *ErrorMatrix.h*) for the storage and manipulation of the error matrices calculated during each experiment. A python script (*summarize.py*) was created to summarize the results of multiple program runs by reading the output files generated by either *samplesTest.cpp* or *featuresTest.cpp*.
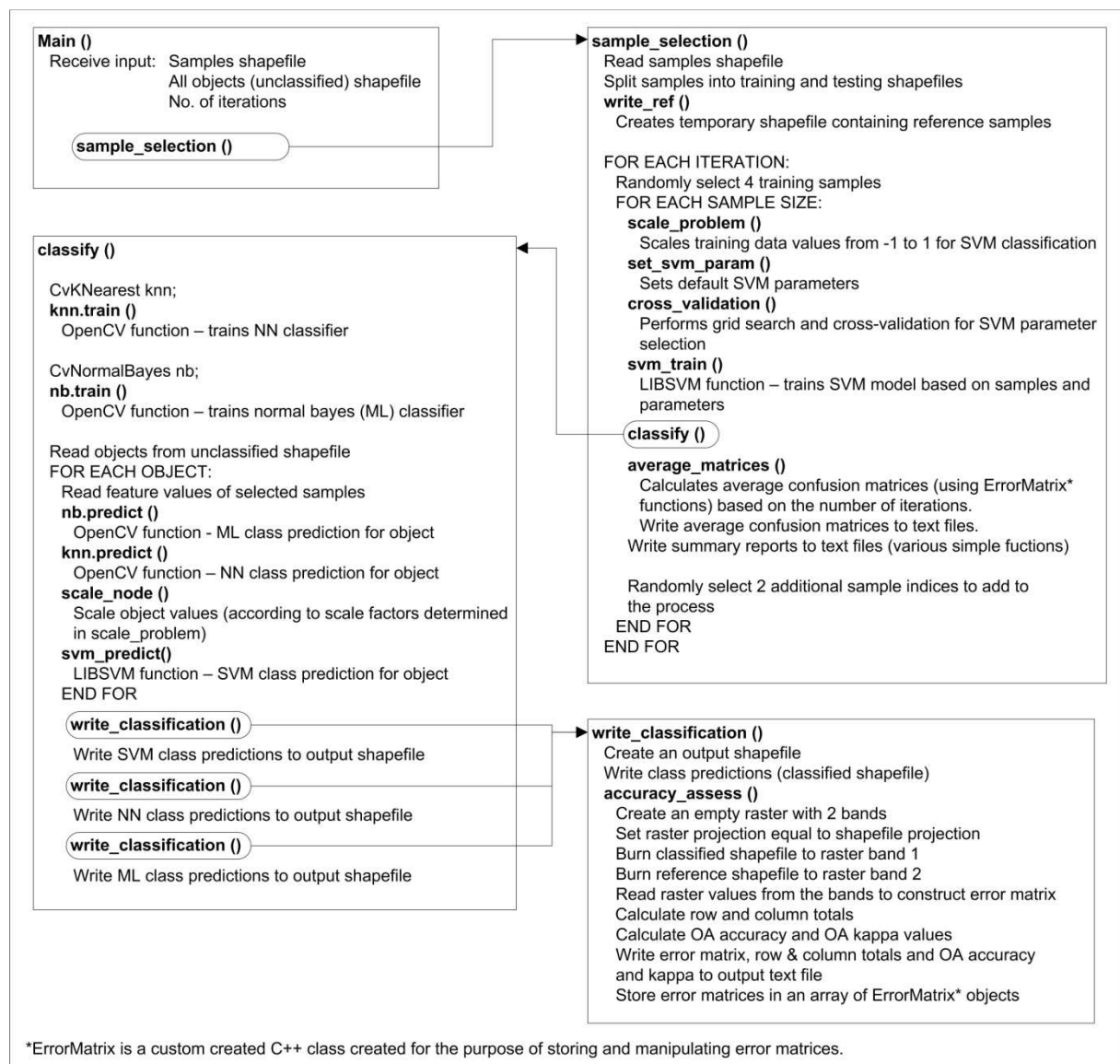

Figure A1: The components and flow of the main C++ script for the first experiment (*samplesTest.cpp*).

APPENDIX B: SOURCE CODE (ONLINE)

The source code of the developed software components (*samplesTest.cpp, featuresTest.cpp, ErrorMatrix.cpp, ErrorMatrix.h* and *summarize.py*) is available for download online at: http://www.sun.ac.za/cga/downloads/scripts.zip.