ACCURACY ASSESSMENT OF REMOTELY SENSED DERIVED THEMATIC MAPS

A DISSERTATION

Submitted in partial fulfilment of the requirement for the award of the degree

of

MASTER OF ENGINEERING

in

CIVIL ENGINEERING

(With Specialization in Remote Sensing and Photogrammetric Engineering)

By

MOHAMAD ALA SHALAN

DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
ROORKEE–247 667 (INDIA)

DECEMBER, 2001

CANDIDATE'S DECLARATION

I hereby certify that the work is being presented in the dissertation entitled **"Accuracy Assessment of Remotely Sensed Derived Thematic Maps"** in partial fulfillment of the requirement for the award of the degree of **"MASTER OF ENGINEERING"** in **Civil Engineering** with specialization in **Remote Sensing and Photogrammetric Engineering**, submitted in the Department of Civil Engineering, IIT Roorkee is an authentic record of my own work carried out from September 2001 to December 2001 under the supervision of **Dr. Manoj K. Ar ora**, Asstt. Professor, Department of Civil Engineering, **Dr. S. K. Ghosh,** Associate. Professor Department of Civil Engineering, IIT Roorkee, Roorkee, and **Dr. S. Sarkar**, Scientist, C.B.R.I., Roorkee.

I have not submitted the matter embodied in this thesis for the award of any other degree or diploma.

Place: Roorkee
Date : 10 December 2001

(Mohamad Ala SHALAN)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge and belief.

(Dr. S. Sarkar)
Scientist,
C.B.R.I.
Roorkee – 247 667

(Dr. S. K. Ghosh)
Asso. Professor,
Civil Engg. Department
IIT Roorkee
Roorkee – 247 667

(Dr. M. K. Arora)
Asst. Professor,
Civil Engg. Department
IIT Roorkee
Roorkee – 247 667

ACKNOWLEDGEMENTS

With great pleasure, I avail the opportunity of expressing my deep sense of gratitude and indebtedness to **Dr. Manoj K. Arora, Dr. S.K. Ghosh,** and **Dr. S. Sarkar**, for their intuitive and meticulous guidance, and perpetual inspiration in carrying out this dissertation. It is facile to say that without their invaluable supervision and constant encouragement this dissertation would not have seen the light of the day. I appreciate their discussions and enthusiasm.

I am grateful to my supervisors for the provision of the references, which helped me in the preparation for this work.

There is no adequate way to express my sincere gratitude to my parents for their blessings, moral support and encouragement.

Dated: 17 December, 2001 **(M. A. SHALAN)**

ABSTRACT

Accuracy assessment of remotely sensed derived thematic maps have lately become an integral part of remote sensing image classification. There are a number of measures for the evaluation of accuracy of thematic classifications both crisp and fuzzy that have been proposed in the remote sensing literature.

There may be lot of variation in the results of classification by the use of different accuracy measures. However, the currently available commercial image processing packages incorporate only a few of these measures.

An attempt has been made here to develop a software for assessing the accuracy of thematic maps. The package has been written in MATLAB script. In order to perform the classification in crisp and fuzzy modes, the algorithms for two classifiers namely, Maximum Likelihood and Fuzzy C-Mean have been included. All commonly used accuracy measures for crisp and fuzzy classification outputs have been considered.

The software has been named as RSICAA and contains five basic modules: Display, Training Data, Classification, Testing Data and Accuracy Assessment Module.

The performance of classifications has been evaluated using IRS 1C LISS III data. A thorough comparison between various accuracy measures has been made. It has been observed that for the data set considered, the MLC and FCM in supervised mode is significantly better than that of FCM in unsupervised mode. Further, for accuracy assessment of crisp classifications, the Kappa and Tau coefficients appear appropriate, whereas for fuzzy classifications, measures of closeness may be considered better than others.

---

eywords: Thematic Maps, Image Classification, and Accuracy Measures.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

INTRODUCTION

1.1 General

A thematic map represents the spatial distribution of some theme such as land use land cover, soil, and geology. Remote sensing data in digital form have now been used widely to produce thematic maps. These maps may be presented either as "Digital Raster Data" derived from image processing, or as "Cartographic Presentation" of digital data.

Classification or interpretation of remote sensing images may be performed either visually or digitally to produce thematic maps. Visual image interpretation is based on human vision and pattern recognition capacities. This technique is laborious and time consuming. Therefore, it is uneconomical for large area mapping. Moreover, no proper qualitative classification accuracy assessment may be done in visual analysis and thus the evaluation is largely subjective in nature. For example, from only the appearance of the classified image, the assessment of the quality is made and termed as 'good' or 'bad' (Congalton, 1991).

With the advancements in computer technology and the increased used of digital remote sensing data for a variety of applications, digital analysis has gained enormous importance. The classification may be carried out on per-pixel basis and sub-pixel basis.

A common question about maps prepared from digital satellite remote sensing may be asked as: ***"how accurate is the classification?"*** No classification is considered complete unless assessment of accuracy has been performed (Jensen, 1986). This gives rise to another question ***"how to measure the accuracy?"*** Fortunately, there has been significant research on classification accuracy assessment techniques over the last two decades.

1.2 Need for Thematic Maps Accuracy Assessment

Accuracy assessment of thematic maps is a critical step in any mapping process. Therefore, it must be considered as an essential component in order to allow a degree of confidence to be attached to these for their effective use. There are two primary motivations behind the assessment of the accuracy of a thematic map:

1) To understand the errors in the map. Both producers and users of thematic maps are interested in this kind of information. Producers can improve methods of making the maps and presenting these along with the information on accuracy and errors to the end user. Information about the errors in the map, in turn, can help map users to interpret and use the map more effectively.

2) To provide an overall assessment that can be used as an indicator of the general reliability of a map.

   (i) It may assist in comparing two maps in order to determine which one is better than the other (Gopal and Woodcock, 1994).

   (ii) It provides the means for the comparison of two thematic classifications of different analysts, of different dates, and from different data sources.

Accuracy assessment is thus a crucial step in the processing of remote sensing data, which is an important source of thematic map production. It determines the value of the resulting data to a particular user, i.e., the information value. So it is a valuable tool in judging the fitness of these data for a particular application.

1.3 Scope and Justification of the Problem

The difficulties associated with assessing the accuracy of thematic maps may be due to their nature. In thematic maps, each location on the ground has to be assigned to a class and this is the conventional way. In essence, the continuum of variation found in the landscape has to be divided into a finite set of classes. Typically, the classes are easily differentiable in their pure states, and become less readily separable near the dividing lines between the classes. For example, consider the difference between the vegetation classes, conifer forest and hardwood forest. At their extremes there is no question regarding the appropriate class (Gopal and Woodcock, 1994). However, all degrees of mixing of coniferous and hardwood trees may be found. When coniferous trees dominate, the appropriate label may still be coniferous forest, but what happens as the mix approaches 50 percent of each? At that point the decision becomes arbitrary and neither class is either entirely right or entirely wrong. One solution is to add another class to the map that is mixed forest. This new class solves the problem in one case (the 50-50 mix). Another approach is to use fuzzy set, which gives the proportion of each class.

Nevertheless, whether a thematic map is produced with pure classes (in crisp form) or it is produced with mixed classes (in fuzzy form), its quality need to be assessed with appropriate measures to make the map meaningful. A number of accuracy indices have been proposed in the remote sensing literature. Some of these measures are percent correct (overall accuracy), user's and producer's accuracy, and Kappa coefficient etc. These measures now form an important part of any image processing system. However, these are useful for the assessment of crisp classifications only. There is lack of software that account for the evaluation of accuracy of fuzzy classifications. Therefore, proper software needs to be developed for various accuracy measures to evaluate fuzzy classifications and also for additional measures to assess crisp classifications.

Keeping this in mind, a software for the assessment of classification accuracy using both crisp and fuzzy measures has been planned to be developed in this thesis. Thus, the objectives of the present work are:

(i) To develop a window based software package in MATLAB environment for accuracy assessment of remotely sensed derived maps.

(ii) To produce thematic classifications both in crisp and fuzzy forms using three algorithms, namely Maximum Likelihood Classifier (MLC), Fuzzy C-Means (FCM) unsupervised, and FCM supervised

(iii) To generate and display classified and fraction images for various classes depicting the visual quality of the classification

(iv) To perform extensive classification accuracy assessment of thematic classifications in both crisp and fuzzy form from IRS 1C, LISS III data.

(v) To compare various accuracy measures in terms of their suitability for crisp and fuzzy classifications.

1.4 Organization of the Thesis

The work presented in this dissertation has been organized into six chapters.

Chapter 1 provides an introduction to the problem. Its justification and scope along with the objectives have also been stated.

In Chapter 2, a brief literature review of the works carried out in the area of subject selected has been highlighted.

Chapter 3 provides various aspects and characteristics of the package developed, its hardware and software requirements along with the data input/output options.

Chapter 4 discusses the methodology used to perform the accuracy assessment of thematic maps produced from remotely sensed data.

The results obtained have been reported and discussed in Chapter 5.

Finally, the conclusions derived from the present study, and future scope have been highlighted in Chapter 6.

LITERATURE REVIEW

2.1 General

Thematic map is a representation of the real world that contains both a spatial component (coordinates) and an attribute component. Attribute accuracy refers to the non-positional characteristics of a spatial data entity. In remote sensing, this accuracy (also known as classification accuracy) refers to the correspondence between the class label assigned to a pixel and the "true" class. The true class can be observed in the field directly or indirectly, for example, from a reference map or aerial photograph etc.

In the following sections, some methods used for image classification along with some commonly used thematic maps accuracy measures are briefly reviewed.

2.2 Digital Image Classification Techniques

Digital image classification is the process to convert a remote sensing image into a map representing classes of interest such as urban, agriculture, forest etc. There are two approaches of image classification namely supervised and unsupervised classification. Supervised classification involves three distinct stages; training, allocation and testing (Foody, 1995a). In contrast to supervised classification, unsupervised classification require only a minimal amount of initial input from the analyst, once the data are classified, the analyst attempts, to assign these spectral classes to the information classes of interest (Robinove, 1981).

With regard to pixel allocation phenomenon, there are two ways of classification namely, crisp and fuzzy classification

2.2.1 Crisp classification

In this, each pixel is assumed to be homogenous and is, therefore, classified to a particular class. In reality, not all pixels may be pure. Therefore, this technique may lead to loss of information content of the pixel. Hence, the results obtained from crisp classification may not be accurate. All conventional classification algorithms produce crisp classification outputs. MLC, minimum distance to means, parallelepiped and Mahalanobis distance classifiers are some of the algorithms that provide crisp classification.

2.2.2 Fuzzy classification

It is a kind of sub pixel classification. Here each pixel is decomposed into those classes which may represented by assigning the membership grades to each of those classes within the pixel. These membership values or grades indicate the class proportions within the pixel. Some of the fuzzy classifiers are FCM, linear mixture modeling, and fuzzy artificial neural networks.

The MLC has generally been used as a technique of providing crisp classification output. However, the output of an MLC may also be fuzzified to obtain the partial and multiple class membership for each pixel (Wang, 1990b). Here, the measures of strength of class membership rather than the code of the most likely class of membership may be the output (Foody, 1996c). Thus, for instance, the a-posteriori probabilities from a maximum likelihood classification may reflect to some extent the class composition of a mixed pixel (Foody et al., 1992).

2.3 Accuracy Assessment

Typically, a thematic map is derived from remote sensing data through a digital image classification procedure. Once a classified image is obtained, its quality is judged on the basis of the accuracy. In order to evaluate accuracy, the true value must be known. This involves selecting a set of pixels from the classified image and comparing their identity with that to the reference data.

The accuracy of a classification may be based on the Euclidian or statistical distance derived from training data itself. However, the training stage in classification is very subjective and therefore, the distance may not be considered independent and useful for accuracy assessment, if determined from training data set.

According to the nature of classification (i.e. crisp or fuzzy), proper accuracy measures may have to be used to derive qualitative information from thematic maps. In the following sections, some of the commonly used accuracy measures have been briefly reviewed.

2.3.1 Accuracy of thematic maps considering crisp classification

A thematic map produced from remote sensing data using conventional image classification techniques is a crisp one as each pixel is classified to one and only one class. A typical strategy for accuracy assessment of a crisp thematic map is to use a statistically sound sampling design to select a sample of pixels (also known as testing samples) in the study region, and to determine if the class assigned to that pixel matches the true class represented by that pixel on ground (reference data) or not. The sample data are often summarized in an error matrix, from which various accuracy measures may be derived (Congalton et al., 1983).

An error matrix is a cross-tabulation of the thematic classes on the classified image and on the reference data. It is represented by a c×c matrix (where c is the number of classes). The elements of this matrix indicate the number of samples in the testing data. The columns of the matrix generally define the reference data, and the rows define the classified image, but they can be interchanged. A typical error matrix is shown in Table 2.1

Table 2.1 A Typical Error Matrix

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="4">Reference data</td><td rowspan="2">Row Total</td></tr>
<tr><td>Class 1</td><td>Class 2</td><td>...</td><td>Class c</td></tr>
<tr><td rowspan="4">Classified image</td><td>Class 1</td><td>n_{11}</td><td>n_{12}</td><td>...</td><td>n_{1c}</td><td>N_1</td></tr>
<tr><td>Class 2</td><td>n_{21}</td><td>n_{22}</td><td>...</td><td>n_{2c}</td><td>N_2</td></tr>
<tr><td>⋮</td><td>⋮</td><td>⋮</td><td>...</td><td>⋮</td><td>⋮</td></tr>
<tr><td>Class c</td><td>n_{c1}</td><td>n_{c2}</td><td>...</td><td>n_{cc}</td><td>N_c</td></tr>
<tr><td colspan="2">Col. Total</td><td>M_1</td><td>M_2</td><td>...</td><td>M_c</td><td>$N = \sum_{i=1}^{c} N_i$</td></tr>
</table>

In this matrix, the various terms have been defined as:

N= total number of testing samples.

c= number of classes.

n_{ii}= number of samples correctly classified.

N_i= row total for class i.

M_i=column total for class i.

For an ideal classification, it is expected that all points lie on the diagonal of the matrix. This indicates that the same class has been observed both on the ground as well as on the map. An error of omission occurs when a class on the ground is incorrectly recorded in the map. An error of commission occurs when the class recorded in the map does not match on the ground.

Ideally, a single accuracy measure should express classification accuracy. However, a plethora of measures have been proposed in the remote sensing literature. Some of the commonly used accuracy measures for crisp classification are shown in (Table 2.2).

Table 2.2 Accuracy Measures for Crisp Classification

Measure	Abbreviation	Explanation	Formula	Base Reference(s)
Overall Accuracy	OA	Percent of samples correctly classified	$\frac{1}{N}\sum_{i=1}^{q} n_{ii}$	Story and Congalton (1986)
User's accuracy	UA	Index of individual class accuracy computed from row total	n_{ii}/N_i	Story and Congalton (1986)
Producer's accuracy	PA	Index of individual class accuracy computed from column total	n_{ii}/M_i	Story and Congalton (1986)
Average accuracy	AA_u	Average of all the individual user's accuracies.	$\frac{1}{q}\sum_{i=1}^{q} \frac{n_{ii}}{N_i}$	Fung and LeDrew (1988)
	AA_p	Average of all the individual producer's accuracies.	$\frac{1}{q}\sum_{i=1}^{q} \frac{n_{ii}}{M_i}$	
Combined accuracy	CA_u	Average of overall accuracy and average user's accuracy	$\frac{1}{2}\left[OA + AA_u\right]$	Fung and LeDrew (1988)
	CA_p	Average of overall accuracy and average user's accuracy	$\frac{1}{2}\left[OA + AA_P\right]$	

Continued

Table 2.2 (contd.) Accuracy Measures for Crisp Classification

Measure	Abbreviation	Explanation	Formula	Base Reference(s)
Kappa coefficient of agreement	K	Proportion of agreement after removing the proportion of agreement by chance	$\frac{P_o - P_e}{1 - P_e}$	Congalton et al. (1983)
Weighted Kappa	K_w	Proportion of weighted disagreement corrected for chance	$1 - \frac{\sum v_{ij} P_{oij}}{\sum v_{ij} P_{cij}}$	Rosenfield and Fitzpatrick-Lins (1986)
Conditional Kappa	K_{i+}	Conditional Kappa computed from the i^{th} row in error matrix (User's)	$\frac{P_{o(i+)} - P_{e(i+)}}{1 - P_{e(i+)}}$	Rosenfield and Fitzpatrick-Lins (1986)
	K_{+I}	Conditional Kappa computed from the i^{th} column in error matrix (Producer's)	$\frac{P_{o(+i)} - P_{e(+i)}}{1 - P_{e(+i)}}$	
Tau coefficient	T_e	Tau for classifications based on equal probabilities of class membership	$\frac{P_o - \frac{1}{q}}{1 - \frac{1}{q}}$	Foody (1992), Ma and Redmond (1995)
	T_p	Tau for classifications based on unequal probabilities of class membership	$\frac{P_o - P_r}{1 - P_r}$	
Conditional Tau	T_{i+}	Conditional Tau computed from the i^{th} row (User's)	$\frac{P_{o(i+)} - P_i}{1 - P_i}$	Naesset (1996)
	T_{+I}	Conditional Tau computed from the i^{th} column (Producer's)	$\frac{P_{o(+i)} - P_i}{1 - P_i}$	

In this table, most of the terms have been defined earlier, the rest may be defined as:

$P_o = \frac{1}{N}\sum_{i=1}^{q} n_{ii}$	The observed proportion of agreement
$P_e = \frac{1}{N^2}\sum_{i=1}^{q} N_i M_i$	The expected chance agreement
$P_r = \frac{1}{N}\sum_{i=1}^{q} n_{i+} x_i$	x_i are the unequal priori probabilities of class membership
V_{ij}	Agreement weight
$P_{Oij} = \frac{n_{ij}}{N}$	Observed cell proportion
Pe_{ij}	Expected cell proportion
$Po_{(i+)}$	Observed agreement according to user's approach
$Pe_{(i+)}$	Agreement expected by chance for i^{th} row
$Po_{(+i)}$	Observed agreement according to producer's approach
$Pe_{(+i)}$	Agreement expected by chance for i^{th} column
P_i	A priori probability of class membership

2.3.2 Accuracy of thematic maps considering fuzzy classification

The accuracy measures used in crisp classification assume that each testing sample is associated with one class in the classified image and one class in the reference data (Congalton, 1991). Frequently, the samples comprising the testing data set may contain mixed classes and thus, may not belong to only one class. As a consequence, the classification accuracy measures derived on the basis of the error matrix may result into under or over estimation of accuracy. Therefore, these measures may not be appropriate when either the classification output or the reference data or both are fuzzy. Under such circumstances, it is better to use some other measures. Some of the commonly used measures for fuzzy classification are listed in Table 2.3. The various terms used in Table 2.3 may be defined as follows:

${}^1p_i$ is the proportion of i^{th} class in a pixel from the fuzzy reference data.

${}^2p_i$ is the proportion of i^{th} class in a pixel from the fuzzy classification.

${}^1p$ is the probability distribution of fuzzy reference data.

${}^2p$ is the probability distribution of fuzzy classification output.

$Cov({}^1p, {}^2p)$ is the covariance between the two distributions.

σ_{1p} , σ_{2p} are the standard deviations of both the distributions.

Table 2.3 Accuracy Measures for Fuzzy Classification

Measure	Abbreviation	Explanation	Formula	Base Reference(s)
Entropy	H	Measure of uncertainty, when there is a finite set of alternative classes.	$-\sum_{i=1}^{c}({}^{2}p_i)\log_2({}^{2}p_i)$	Maselli et al. (1994)
Measures of distance	S	Measure the Euclidean distance between the fuzzy representation of land cover and reference data	$\sum_{i=1}^{c}({}^{1}p_i-{}^{2}p_i)^2 \Big/ c$	Kent and Mardia (1988) Foody (1996b)
L_1 (city block) distance	L_1	Absolute value of the difference between proportions in fuzzy reference and classified data	$\sum_{i=1}^{c}\left\lvert {}^{1}p_i-{}^{2}p_i \right\rvert \Big/ c$	Foody and Arora (1996)
Cross-entropy (direct divergence)	d	Distance measure to evaluate the degree of similarity between two data sets	$d({}^{1}p,{}^{2}p)=-\sum_{i=1}^{c}({}^{1}p_i)\log_2({}^{2}p_i)+\sum_{i=1}^{c}({}^{1}p_i)\log_2({}^{1}p_i)$	Foody (1995b)
Measure of information closeness	D	Measure of the closeness of pairs of probability distributions	$D({}^{1}p,{}^{2}p)=d\left({}^{1}p,\frac{{}^{1}p+{}^{2}p}{2}\right)+d\left({}^{2}p,\frac{{}^{1}p+{}^{2}p}{2}\right)$	Foody (1996b)
Correlation coefficients	r	Measures of the relationship between two data sets	$\frac{\mathrm{Cov}({}^{1}p,{}^{2}p)}{\sigma_{{}^{1}p}\sigma_{{}^{2}p}}$	Maselli et al. (1996) Foody and Cox (1994)

2.3.3 Comparison of classification accuracy measures

The overall accuracy (OA) is one of the most commonly adopted measure (Arora and Ghosh, 1998). OA is a measure of classification as a whole and not of individual classes. However, it has a tendency of bias towards the class having a large number of testing samples. This situation occurs when the testing samples are collected in a stratified random sampling scheme, in which some classes occupy a larger proportion of the area than others (Miguael-Ayanz et al., 1996). A way to resolve the problem of differences in sample size is to normalize the elements of the error matrix and then compute OA. The normalized value of OA has been called "normalized accuracy". Nevertheless, OA does not take into account the off-diagonal elements of the error matrix which represent misclassification errors. These errors may be grouped into two types, namely "error of omission" and "error of commission" (Story and Congalton, 1986). Complementary to these errors, a new set of accuracy measures has been derived: producer's accuracy (PA) and user's accuracy (UA). These measures determine the accuracy of individual classes. PA is so aptly called, since the producer of the classified image is interested in knowing how well the samples from the reference data can be mapped using remotely sensed data. In contrast, UA indicates the probability or reliability that a sample from the classified image represents an actual class on the ground. Although these measures may appear simple, it is critical that they both be considered when assessing the accuracy of a classified image on a per class basis.

While OA is biased towards the class with a large number of testing samples, Average Accuracy (AA) is biased towards the class having a small number of samples (Fung and LeDrew, 1988). Combined Accuracy (CA) may be used to reduce the biases of OA and AA. However, AA and CA do not take into account the agreement between the data sets (i.e., classified image and reference data) that arises due to chance alone. Thus, these measures tend to overestimate the classification accuracy (Ma and Redmond, 1995). The Kappa coefficient of agreement (K) has the ability to account for chance agreement (Foody, 1992). The proportion of agreement by chance is the result of the misclassifications represented by the off-diagonal elements of the error matrix.

Therefore, K uses all the elements of the error matrix, and not just the diagonal elements (as is the case with OA). Therefore, the Kappa coefficient of agreement may be used for the assessment of the accuracy of the classification as a whole and for individual classes using conditional Kappa after making some compensation for chance agreement. This may prove to be a desirable accuracy index. Hence, K has now become a commonly used accuracy index. Weighted Kappa (K_w) can be thought as a generalization of Kappa, as it does not treat all the misclassifications (disagreements) equally and tends to give more weight to some errors that are more serious than others.

Tau coefficient is superficially similar to Kappa. However, the critical difference between the two coefficients is that Tau is based on a ***priori*** probabilities of group membership, whereas Kappa uses the a ***posteriori*** probabilities. Tau is easier to understand and interpret than Kappa. Unlike Kappa, Tau compensates for unequal probabilities of groups or for difference in number of groups. In other words, T_p compensates for the influence of unequal probabilities of groups on random agreement, and T_e compensates for the influence of the number of groups (Ma and Redmond, 1995). A Conditional Tau may also be used to determine the accuracy of an individual class. However, Conditional Tau corresponds close to producer's accuracy.

Although, the percent correct and Kappa coefficient are the most widely used measures of accuracy, these may be appropriate for crisp classifications only, when each pixel is associated with only one class in the classification and only one class in the reference data. These measures may under or over estimate the accuracy of fuzzy classification. Therefore, some other measures may be used such as Entropy, which shows how the strength of class membership in the classification output is partitioned between the classes for each pixel. Entropy is therefore attractive as an indicator of classification quality in situations where ambiguity exists as it indicates the degree to which the class membership probabilities are partitioned between the defined classes (Foody, 1996b). Entropy is maximized in the situation when the probability of class membership is partitioned evenly between all classes in the thematic map and minimized when it is associated entirely with one class. Its value as an indicator of classification accuracy is therefore based implicitly on the assumption that in an accurate classification each pixel will have a high probability of membership with only one class. This is, however, only appropriate for situations in which the output of the classification is fuzzy and the reference data are "crisp". Therefore, entropy may not be a good indicator of thematic quality if multiple and partial class membership is a feature of both the classification output and the reference data. To accommodate fuzziness in both the classification output and the reference data, other measures are required such as simple measures of distances namely Euclidian distance (S), L_1 distance (L_1) etc, which measures the separation of two data sets and may be based on the relative extent or proportion of each class in the pixel (Foody and Arora, 1996). Another approach, which may be used to express the information closeness, is to calculate the directed divergence or cross-entropy (d), where the distance between two data sets may be assessed. A small distance indicates that the classification is an accurate representation of the thematic data (Wang, 1990b). This distance measure is applicable when the probability distributions to be compared are compatible. However, to make it applicable to any pair of probability distribution, the generalized measure of information closeness (D) may be used. Correlation coefficient (r) may also used to indicate the accuracy on per-class basis estimated from a fuzzy classification output and fuzzy reference data. The higher the correlation coefficient, higher is the classification accuracy of a class.

In this chapter, the various accuracy measures have been briefly discussed. It has been found that a plethora of accuracy measures has been proposed in the remote sensing literature. These measures may be divided into two categories depending on the nature of the classification (i.e., crisp or fuzzy). Under a given situation, a particular measure may be used. So far as the crisp classification accuracy measures are concerned, some of them have been incorporated in commercial image processing systems. However, for others and the fuzzy measures, no proper software is available. The next chapter demonstrates the various aspects of the software package developed for this purpose.

DETAILS OF SOFTWARE DEVELOPED

3.1 General

The main objective of the work presented in this thesis is to focus on various accuracy measures to evaluate the accuracy of remotely sensed derived thematic maps. Since there are a number of accuracy measures both for crisp and fuzzy, it was imperative to develop a user-friendly software for classification accuracy assessment. An attempt has been made in this direction to develop a comprehensive package, though not up to the professional level. The following sections outline the structure and the capability of the package.

3.2 Hardware and Software Requirement

The package has been developed in MATLAB environment. MATLAB is software, developed by Math Works Inc. (U.S.A), basically for easy matrix computations. Generally, a script file with extension "m" is written to execute a sequence of MATLAB statements. The syntax of the script is based on C language. To facilitate various operations, the MATLAB contains a number of toolboxes such as Image Processing, Fuzzy Logic, Neural Network, Signal Processing etc. In the present work, help from some of these toolboxes has been derived. Thus, various MATLAB routines and functions have been used to develop this package. The basic Graphical User Interface (GUI) resource of the MATLAB has also been used in order to make the package user-friendly.

The minimum requirements of this package are Windows '95' Operating system or its later versions, and 16 MB RAM.

3.3 Salient Features of the Package

The package has been named as RSICAA (Remote Sensing Image Classification Accuracy Assessment). It consists of five basic modules:

1) Display Module

2) Training Data Module

3) Classification Module

4) Testing Data Module

5) Accuracy Assessment Module

GUI based Main Menu is shown in Plate 3.1.

The various options available on menu bar attached to the Main Menu are shown in Table 3.1

Table 3.1 Options on Menu Bar

Main Options	Popup Menus
File	Open Exit
Display	Display Input images Display Crisp classified images Display Fraction images
Training Data	Generate Training data
Classification	Maximum Likelihood Classifier Fuzzy C-Mean Classifier
Testing Data	Generate Testing data Proportions for whole image Proportions for only testing pixels
Accuracy Assessment	Crisp measures Fuzzy measures

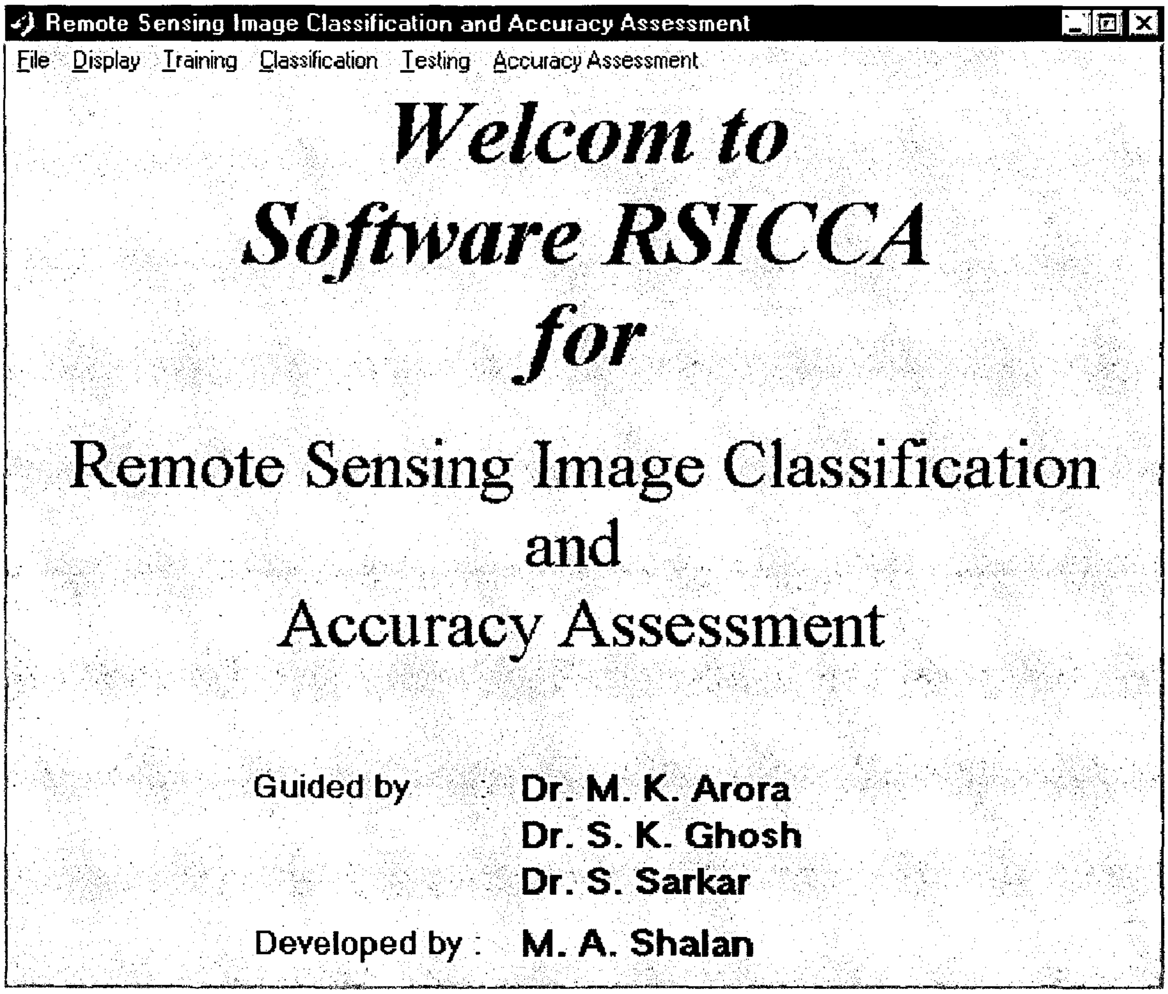

Plate 3.1 Main Menu of Software RSICAA

3.3.1 Display module

This module displays the input and output images stored as ASCII or text files. The ASCII file consists of the information of a pixel in each row, the columns indicating the X and Y coordinates and the Digital Numbers (DN) of the pixel in various bands (b1, b2, b3, b4...). A sample of the image data file is shown in Table 3.2. This format matches with the ASCII format corresponding to ERDAS Imagine software.

Table 3.2 Format of Image Data File

X (m)	Y (m)	b1	b2	b3	b4
44541.00	2962276.00	119	85	89	193
44566.00	2962276.00	116	86	92	191
44591.00	2962276.00	113	82	95	189
44616.00	2962276.00	109	76	96	186
44641.00	2962276.00	109	73	96	180
⋮	⋮	⋮	⋮	⋮	⋮

Each single band image is displayed as B&W image in shades of gray. The multi-spectral image is displayed as False Color Composite (FCC) where the user has the option of choosing any three bands.

Similarly, the classified images generated from crisp and fuzzy classifications in ASCII form can also be displayed. In Table 3.3, the first two columns indicate the X and Y coordinates for each pixel in the classified image and the last column indicates the class identity for each pixel.

Table 3.3 Format of Crisp Classification Output File

X (m)	Y (m)	Class Identity
44541.00	2962276.00	2
44566.00	2962276.00	2
44591.00	2962276.00	1
44616.00	2962276.00	2
44641.00	2962276.00	5
44666.00	2962276.00	2
44691.00	2962276.00	2
44716.00	2962276.00	4
⋮	⋮	⋮

In order to display crisp classification, the user has the option of choosing a particular colour for each class from the color selection dialog box.

In fuzzy classification, a set of fraction images shall be generated for each class where the membership of the class may be represented in gray shades. The lighter the shade, higher is the class membership for the corresponding class in that pixel.

A sample of fuzzy output classification file is shown in Table 3.4. Here, the first two columns represent the X and Y coordinates for each pixel, the rest of the columns represent the proportion or class membership for each pixel in each class (class1, class2, class3, class4, class5) of interest.

Table 3.4 Format of Fuzzy Classification Output File

X (m)	Y (m)	class1	class2	class3	class4	class5
44541	2962276	0.292	0.048	0.426	0.090	0.144
44566	2962276	0.315	0.056	0.363	0.103	0.163
44591	2962276	0.366	0.067	0.244	0.126	0.199
44616	2962276	0.389	0.076	0.139	0.153	0.244
44641	2962276	0.316	0.101	0.108	0.198	0.277
44666	2962276	0.264	0.127	0.090	0.235	0.285
44691	2962276	0.172	0.173	0.051	0.323	0.281
44716	2962276	0.080	0.360	0.027	0.365	0.167
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Further, if the user desires to export the images to any other packages such as ERDAS Imagine, the same can be done by saving it in JPG format, after the image has been displayed on screen. Table 3.5 list the various MATLAB functions used in this module.

Table 3.5 MATLAB Functions for Display Module

Matlab Functions	Activity
image( )	Create and display image object
imshow ( )	Display image
imwrite( )	Write image file
uisetcolor( )	Color selection dialog box
figure( )	Create figure window
axis( )	Control axis scaling and appearance
title ( )	Adds text at the top of the figure
xlabel( )	Adds text along the X-axis
ylabel( )	Adds text along the Y-axis

3.3.2 Training data module

In this module, data required in training stage of a supervised classification may be generated. To generate training data, the user can define select the training areas by interactively displaying the input image on the screen. The selection may be polygon based or per pixel basis. Subsequently, the selected training areas can also be plotted to view their spatial location. All the training areas for a particular class are merged and stored in an ASCII file. This file shall consist of the information of a training pixel in each row arranged as per each class, while the columns indicate the X and Y coordinates the DN value of each training pixel in different bands (b1, b2, b3...). The last column contains the class identity number (1,2,3...). A sample of training data file is shown in Table 3.6. The user may also enter the training data through an existing file. This file may not contain the last column of class identity, as the program will prompt the user to enter the number of classes and the number of training pixels in each class: On this basis the last column can be generated automatically.

This training data file either created in the package or the existing one shall then be used in the classification module.

Table 3.6 Format of Training Data File

X (m)	Y (m)	b1	b2	b3	b4	Class Identity
45016	2962276	102	77	83	182	1
45841	2961976	103	70	67	152	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
46241	2961976	102	76	82	186	2
46291	2961976	101	75	80	195	2
46341	2961976	101	75	81	199	2
46216	2961951	103	76	81	188	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
46316	2961951	102	75	81	200	3
44816	2961926	103	71	67	156	3
45641	2961926	102	75	78	180	3
46191	2961926	103	76	81	186	3
46216	2961926	102	77	82	194	3
46266	2961926	102	76	81	200	3

3.3.3 Classification module

In this module, the classification of remote sensing image is carried out. Both crisp and fuzzy classifications can be performed using the following classifiers:

(i) Maximum Likelihood Classifier (MLC)

(ii) Fuzzy c-Means Algorithm (FCM): Both Supervised and Unsupervised.

The menu of MLC is shown in Plate 3.2. It consists of two main stages, training and allocation stage. In the training stage, the user may select the training areas (if not done in training module) or may input the existing training data file created in the training module. The training file may also be imported from other packages in the format described in Section 3.3.2.

The user may view the spatial location of these training areas at this stage also. An example of the plot is shown in Plate 3.3.

Sometime it is necessary to examine the quality of training areas of a class by examining the histogram. A uni-modal histogram is an indication of the homogeneity of training data for a class. The package has an option to display the histogram of the training data selected for a class. Plate 3.4 shows a sample plot of the histogram for a class.

In the allocation stage, the classification for testing pixels only and/or whole image are performed using MLC. For this the probabilities that each pixel belongs to a particular class are computed using Equation 3.1.

$$p_i(x) = \frac{1}{(2\pi)^{b/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right] \qquad (3.1)$$

Where:

b is the dimension of a pixel vector (i.e., the number of bands).

μ_i is The mean vector of training data, and may be computed using Equation 3.2

$$\mu_i = \frac{\sum_{j=1}^{n} x_j}{n} \qquad (3.2)$$

Σ_i is the variance-covariance matrix and is computed using Equation 3.3

$$\Sigma_i = \frac{\sum_{j=1}^{n}(x_j-\mu_i)(x_j-\mu_i)^T}{n-1} \qquad (3.3)$$

where x_j is the pixel value of the j^{th} pixel in different bands, n is the number of training pixels in the i^{th} class.

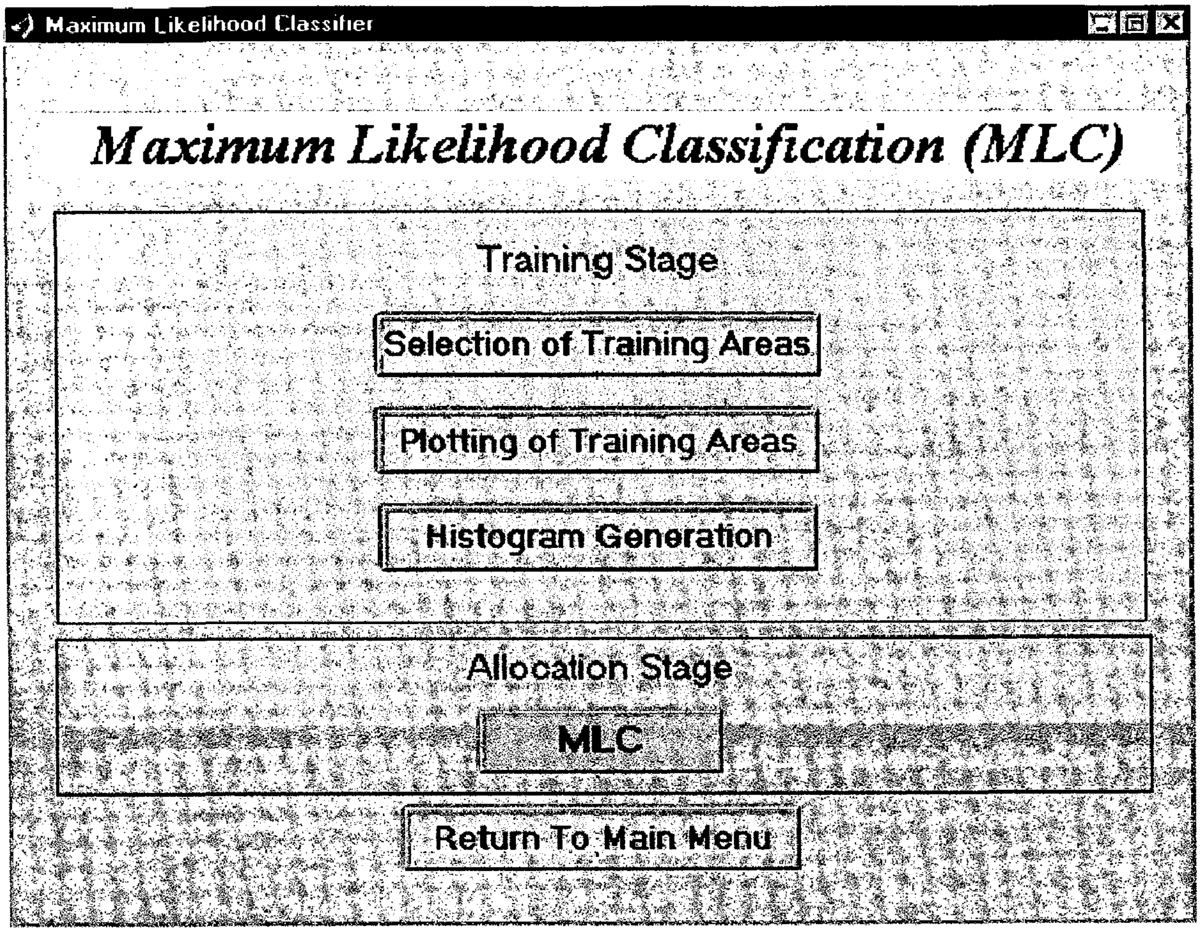

Plate 3.2 Menu for Maximum Likelihood Classification

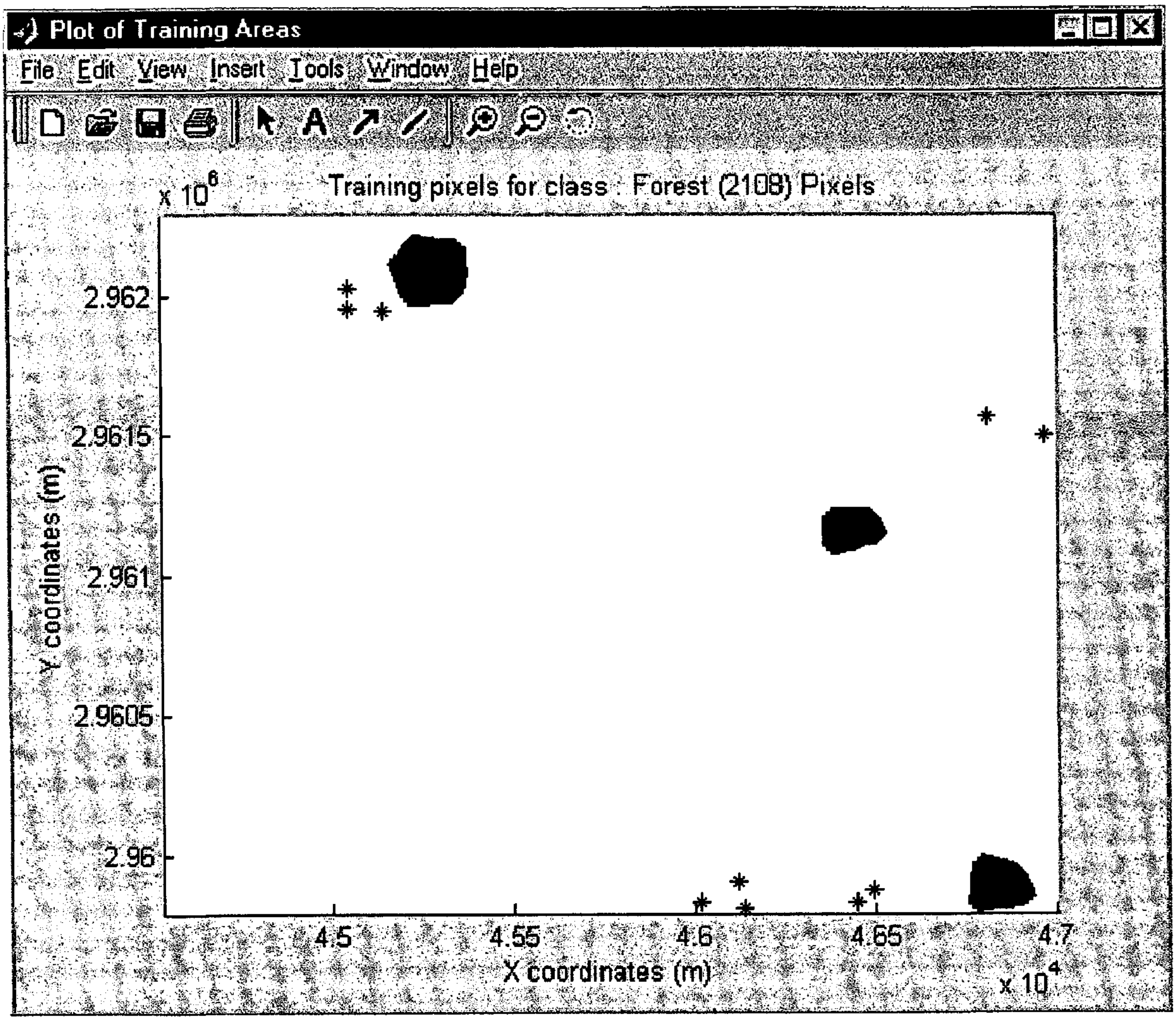

Plate 3.3 Sample Plot for Training Areas

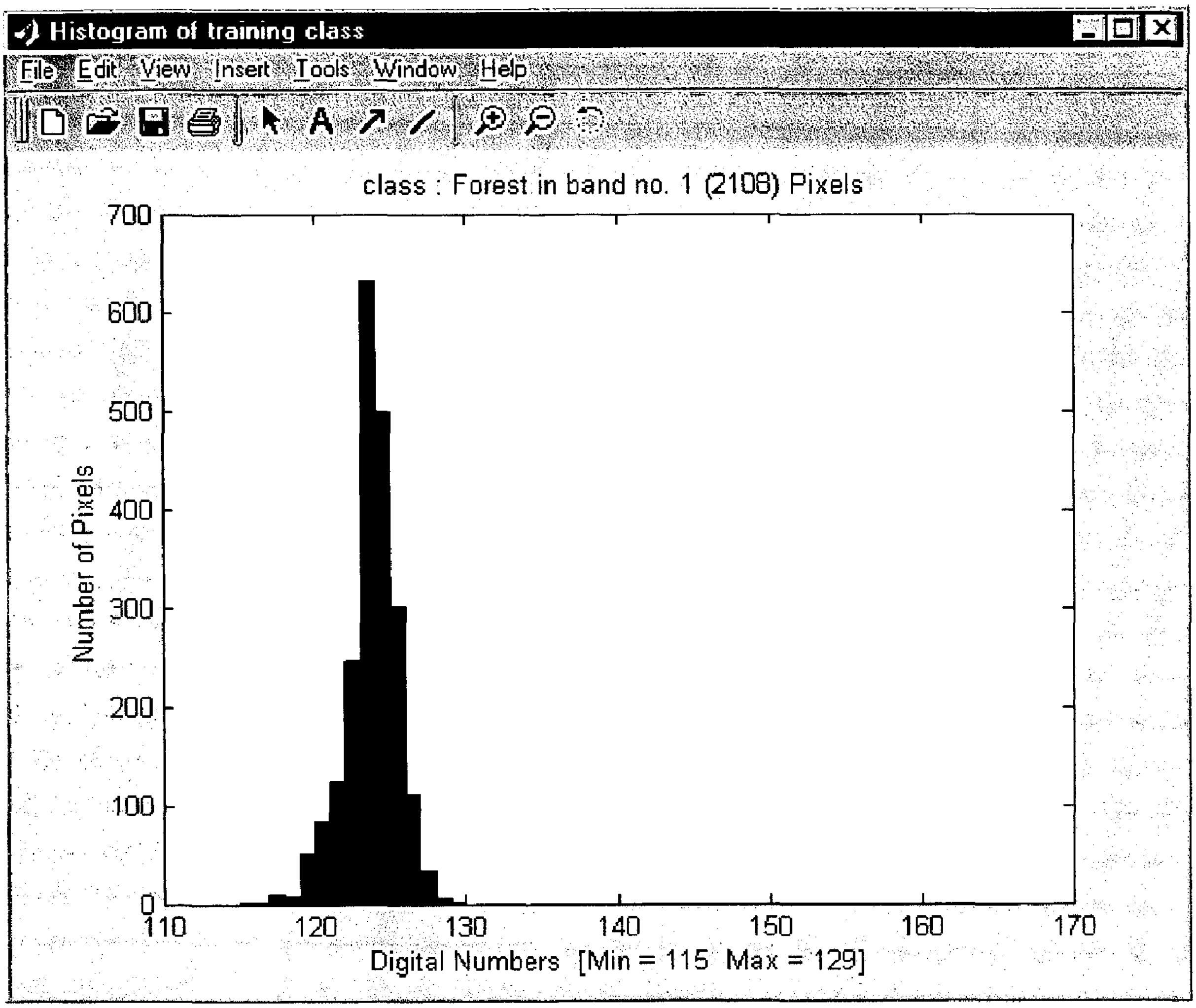

Plate 3.4 Sample Plot for Histogram of a class in a band

In crisp classification, a pixel is assigned to that class whose probability is the highest amongst all other classes, while in fuzzy classification, the membership value of a pixel to class i, can be computed from Equation 3.4

$$f_i = \frac{p_i(x)}{\sum_{j=1}^{m} p_j(x)} \tag{3.4}$$

where $p_i(x)$ is given by Equation 3.1 except that μ_i and Σ_i are replaced by μ_i^* and Σ_i^* which may be computed using Equations 3.5 and 3.6 respectively.

$$\mu_i^* = \frac{\sum_{j=1}^{n} f_i(x_j)x_j}{\sum_{j=1}^{n} f_i(x_j)} \tag{3.5}$$

$$\Sigma_i^* = \frac{\sum_{j=1}^{n} f_i(x_j)(x_j - \mu_i^*)(x_j - \mu_i^*)^T}{\sum_{j=1}^{n} f_i(x_j)} \tag{3.6}$$

Where $f_i(x_j)$ is the membership values of class i in a pixel and x_j is the pixel value vector ($1 \leq j \leq n$) in 'b' bands.

Figure 3.1 depicts the flow chart for MLC classification Procedure.

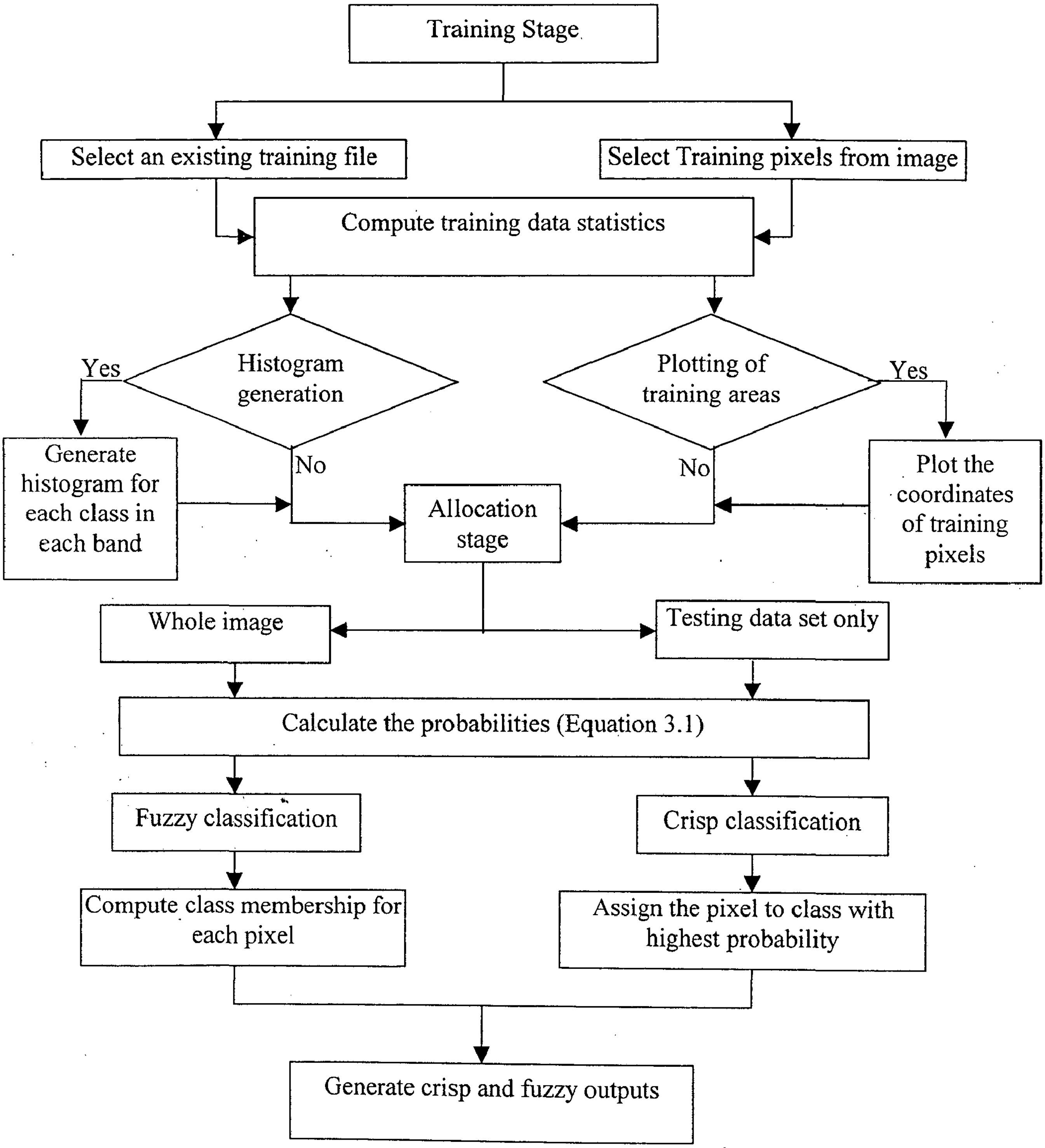

Figure 3.1 Flow Chart to Perform Maximum Likelihood Classification

In the software package, the Fuzzy C-Means (FCM) classifier has been incorporated both in supervised and unsupervised mode. The FCM algorithm is an iterative clustering method that is used to partition a data set. An optimal fuzzy c-partition is the one that minimizes the generalized least-squared errors function (Equation 3.7):

$$\text{Minimize}: Jm(U,v) = \sum_{k=1}^{N}\sum_{i=1}^{c}(u_{ki})^m \|y_k - v_i\|^2 A \qquad (3.7)$$

where: $Y=\{y1,y2,y3,\ldots,yN\} \subset R^n$ is the data set,

c is the number of clusters in Y:$2 \leq c < n$,

m is a weighting exponent: $1 \leq m < \infty$,

$U=\{u_{ki}\}$ is the fuzzy c-partition of Y,

$\|y_k - v_i\| A$ is an induced a-norm on R^n, and,

A is a positive-definite (nxn) weight matrix.

To perform FCM supervised classification, the fcm ( ) function from MATLAB toolbox has been used here.

In the fuzzy c-mean supervised approach, which is similar to MLC, the user has the choice of selecting training areas from the image or to provide an existing training data file created in training module. Here also, the function fcm ( ) from MATLAB toolbox has been used except that the mean of each training class is taken as the centers of the clusters.

The fuzzy classification outputs is presented as final fuzzy partition matrix (or membership function matrix), whereas for crisp classification outputs, the pixel is assigned that class which has the maximum value in the fuzzy partition matrix. A flow chart for FCM is shown in Figure 3.2

3.3.4 Testing data module

To evaluate the performance of a classification, a set of testing data is needed. In RSICAA package, the user provides to generate testing data, the name of the classified image and the number of testing pixels to be generated. The testing pixels are generated randomly and stored in ASCII format as described in Section 3.3.1. This file is subsequently used in the accuracy assessment module.

In order to evaluate the accuracy of fuzzy classifications, it is necessary to have the knowledge of actual proportions of classes within each pixel of the image or for a set of testing pixels. The proportions can be determined from field or reference data or from reference data such as existing maps, GPS surveys, aerial photographs and remote sensing data at finer resolution than that used for classification. The proportion module in this package is based on the last case (i.e., deriving proportions from the fine resolution image taken as reference data).

The proportions may be computed for each pixel of the image or for a set of testing pixels generated in the training module. The philosophy behind this is that each pixel in a coarse resolution image shall contain a constant number of pixels in the corresponding fine resolution image provided both the images are registered with each other. For example, a pixel having 20 m resolution (coarse resolution image) shall contain four pixels, each of 10 m resolution (fine resolution image). Assuming that each pixel in the fine resolution image contains one and only one class, the class proportions of a pixel in coarse image can be found out. The class identity of each pixel in the fine resolution image can be determined by performing a good quality crisp classification of this image. The job of proportion estimation option in this module is to perform this. The flow chart is shown in Figure 3.3.

3.3.5 Accuracy assessment module

The quality of outputs, whether crisp or fuzzy, produced by different classifiers is examined in this module. A number of measures as reported in Tables 2.2 and 2.3 (see Section 2.3) have been incorporated in this module. There are separate menus each for crisp and fuzzy classification accuracy assessment.

The menu for crisp classification accuracy assessment is shown in Plate 3.5. It consists of two major options, error matrix and accuracy measures. In the error matrix option, the user has the choice of generating an error matrix from the testing data files, selecting an existing error matrix file or by directly entering the elements of the matrix from the keyboard.

To generate error matrix from the package, the user has to specify two files, one for classified image and the other for reference data. These two files must have same format (ref. Table 3.3). The first file may be generated from training module (Section 3.3.2), whereas the second file may be obtained from testing module (Section 3.3.4). The format of existing error matrix file is shown in Table 2.1 (Section 2.3.1).

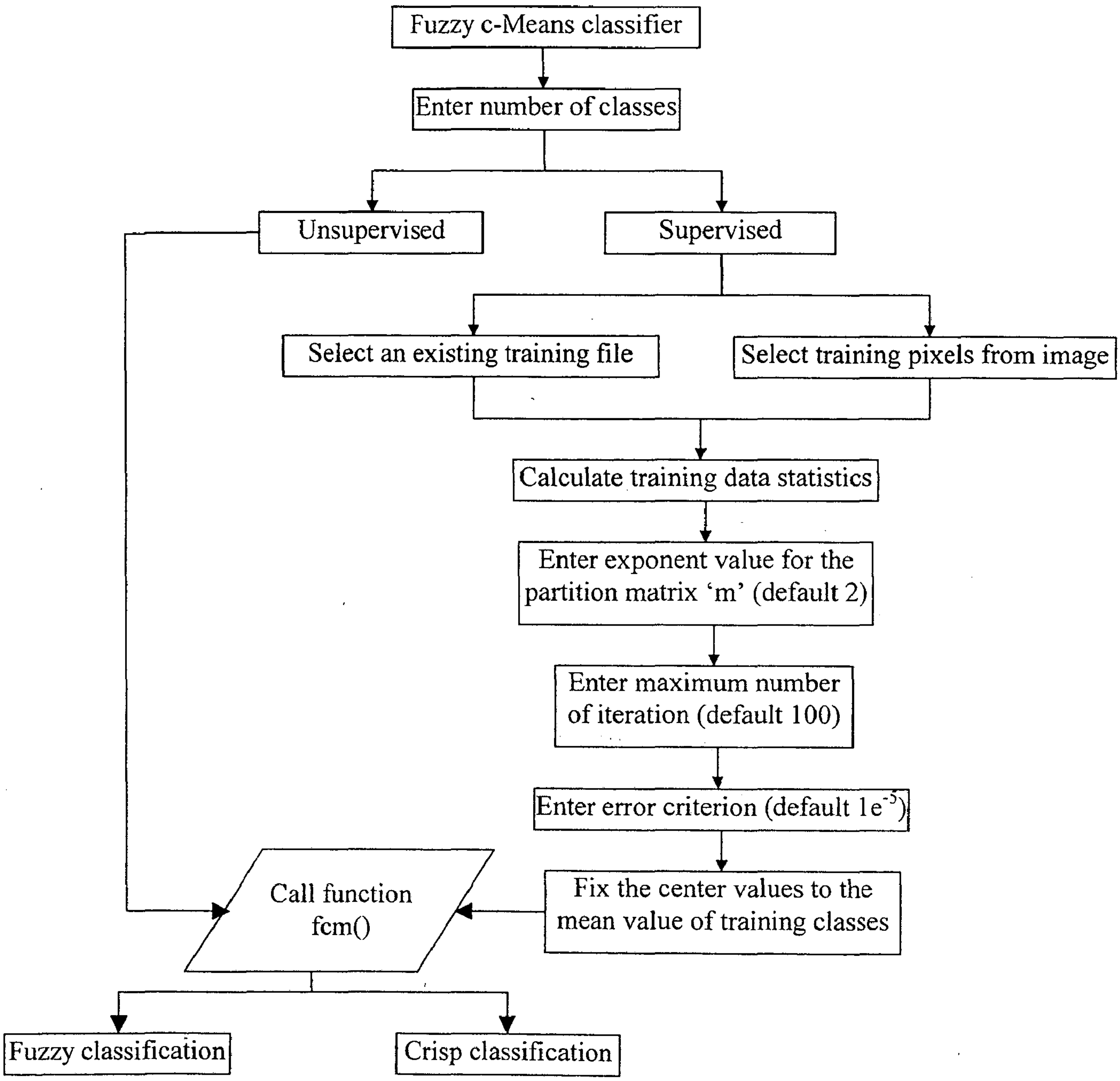

Figure: 3.2 Flow Chart for the Fuzzy C-Mean Classification

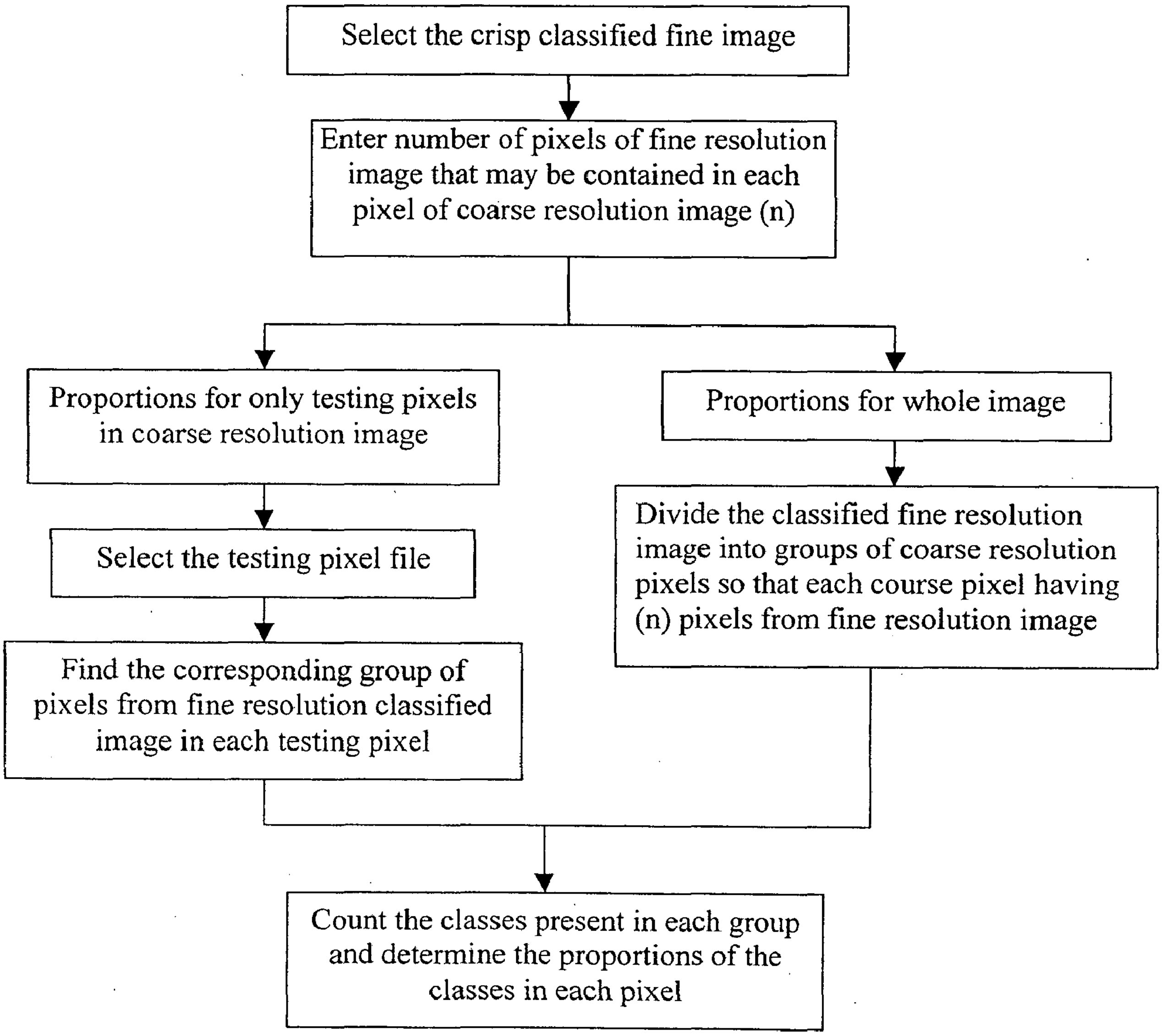

Figure 3.3 Flow Chart for Proportion Estimation Option in Testing Module

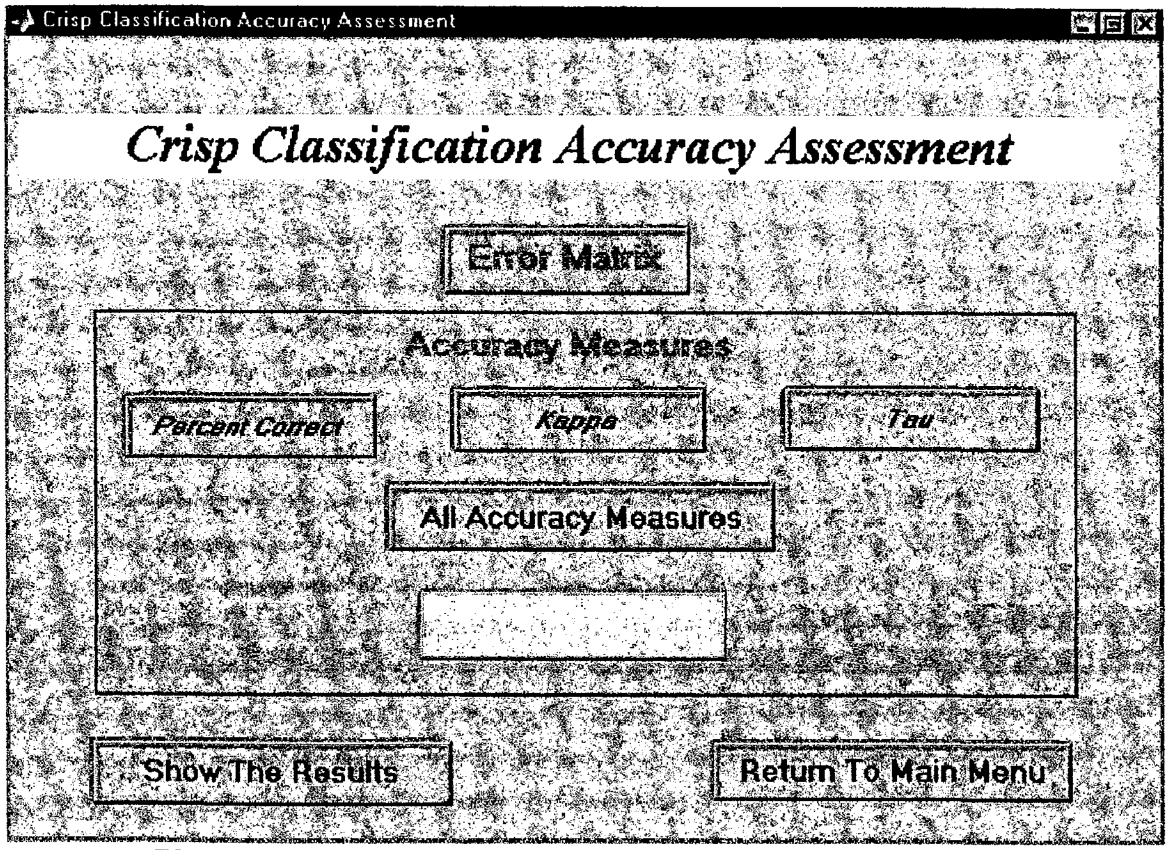

Plate 3.5 Menu for Crisp Classification Accuracy Assessment

After the error matrix has been obtained, the user may select the desired accuracy measures. All crisp measures have been divided into three categories which a user may select.

(i) Percent correct

(ii) Kappa

(iii) Tau

In percent correct category, five accuracy measures namely overall accuracy, user's accuracy, producer's accuracy, average accuracy and combined accuracy have been incorporated.

There are four accuracy measures namely Kappa coefficient, Weighted Kappa, Conditional Kappa (user's and producer's way) under the category Kappa. To obtain Weighted Kappa, a weight matrix has also to be provided by the user at the prompt.

In the Tau category, the Tau with equal and unequal probabilities and the conditional Tau (from user's and producer's perspective) can be computed. The unequal probabilities have to be supplied by the user at the prompt.

Finally, the user may also select the "All Accuracy Measures" option to compute the values of all the measures at one go.

After the desired measures have been selected, the user may click "Compute" to calculate their values. The output file, thus generated, will contain the error matrix along with the values of selected accuracy measures. A sample of output file is given in Appendix A.

For fuzzy classification accuracy assessment, the menu is shown in Plate 3.6. There are three options; entropy measures, measures of closeness and correlation coefficients. The first one is to be used if fuzziness is present only in classified output whereas the other two are used when fuzziness is present both classified outputs and reference data.

The user will be prompted to give the names of respective testing data files for classified and / or reference data.

Under the entropy measures category, entropy and cross entropy based on the mathematical formulations given in Table 2.3 (see Section 2.3.2) has been considered. Measures of distance and information closeness have been incorporated in the category "Measures of Closeness".

Finally, the correlation coefficient can also be obtained by clicking at the "correlation coefficient" option. The user may also select "All Accuracy Measures" option to compute the values of all the fuzzy measures at one go.

After selecting the desired measures, the user may click "Compute" to perform the computations. The values of the measures can be stored in an output file, which can be seen by clicking at the "Show The Results" button. A sample of fuzzy accuracy measures output file is given in Appendix B.

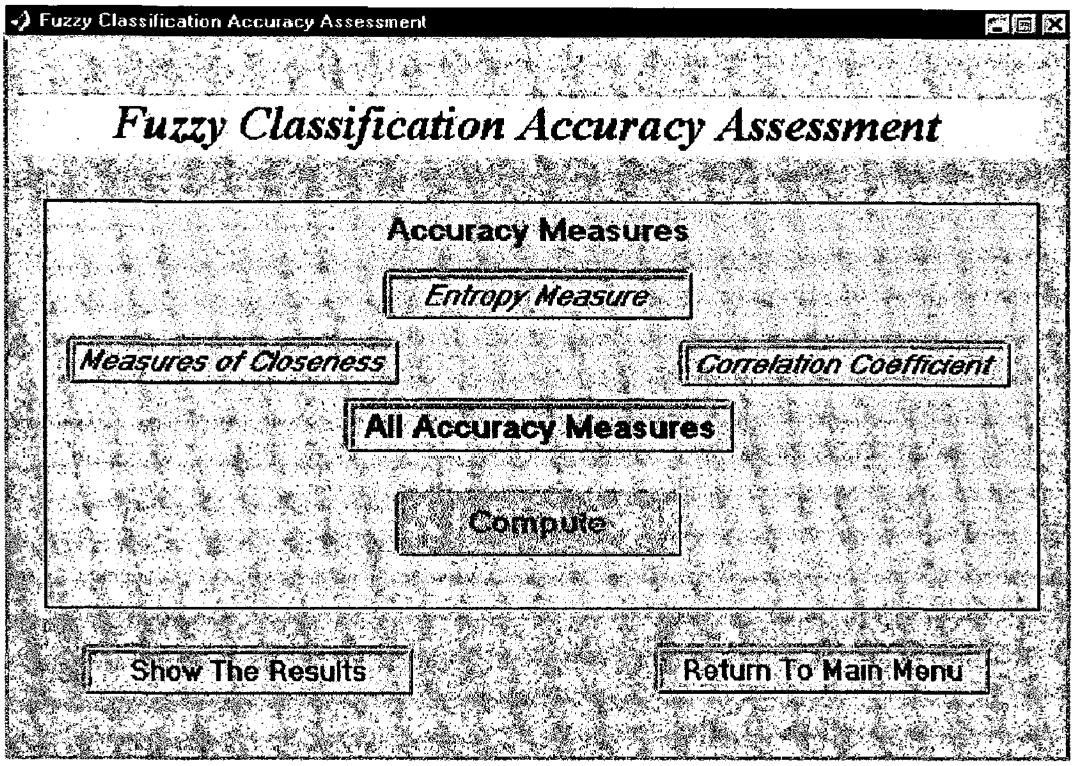

Plate 3.6 Menu for Fuzzy Classification Accuracy Assessment

DATA AND METHODOLOGY

4.1 General

In order to test the efficacy of the software developed, IRS 1C Linear Imaging Self Scanning Sensor (LISS III) and Panchromatic (PAN) data have been used. The LISS III image has been classified using the two classifiers and their accuracy evaluated using different accuracy measures. In this chapter, the details of the data and methodology adopted have been provided.

4.2 Study Area and Data

The study area lies between 88° 27′ E and 88° 28′ E longitudes and 26° 45′N and 26° 46′ N latitudes of Jalpaiguri district in West Bengal (Figure 4.1). The extent of the area considered is estimated to be 620 km^2. The area is primarily covered with agriculture, forest, grasslands, built up and sandy areas. In view of this, five land cover classes have been considered here to produce a thematic map in the form of land cover classification from remote sensing data.

4.2.1 Remote sensing data

Two remote sensing images have been used. The first one is LISS III image in four spectral bands (101 x 99 pixels). The FCC (Red: band 4, Blue: band 2, Green: band 1) of this image (date: 22.3.2000) is shown in Plate 4.1. This is the primary image that has been classified using the two classifiers in fuzzy and crisp modes.

The second image is the PAN image (505 x 495 pixels) from the same satellite taken at about the same time (date: 3.4.2000) and is shown in Plate 4.2. This image has been used as reference data for accuracy assessment and to derive proportions of various classes for the pixels in the LISS III image.

4.2.2 Topographical map

A topographical map at 1:25,000 scale (Survey of India Toposheet No. 78 B/5/6, 1976) of the area has also been used as reference data besides the PAN image.

4.3 Methodology

The main objective of this thesis work is to evaluate the accuracy of remotely sensed derived thematic maps in crisp and fuzzy form using various accuracy measures. Several steps are involved to achieve the objective. The broad methodology has been shown in the form of the flow chart in Figure 4.2

4.3.1 Registration of images

LISS III image has been registered to PAN image to sub-pixel accuracy. The size of each pixel in PAN image has been recomputed to (5x5m) whereas of LISS III image to (25x25m). Thus, each pixel in LISS III image contains 25 pixels of PAN image. The coordinates of the pixels for these two images have been plotted in Plate 4.3. This plot reveals that there is a need to shift the pixels for LISS III image to match with the corresponding pixels in PAN image. A routine in MATLAB has been written to perform the shifting and re-sampling process that makes each pixel in LISS III image to match with the corresponding 25 pixels of PAN image. The new plots for the coordinates after registration are shown in Plate 4.4.

4.3.2 Reference data: classified IRS IC PAN image

PAN image has been classified using Maximum Likelihood Classifier of this software. The training data for each class have been extracted from the image after cross checking with the topographical map. These training data are used to perform the classification for whole image. The output thematic map (Plate 4.5) is based on crisp classification and consists of five classes namely agriculture, forest, grassland, sandy, and built up areas. The classified image has been compared with the PAN image and the topographical map visually to ensure that the quality of this classification is good so that it can be used as reference data (ground truth).

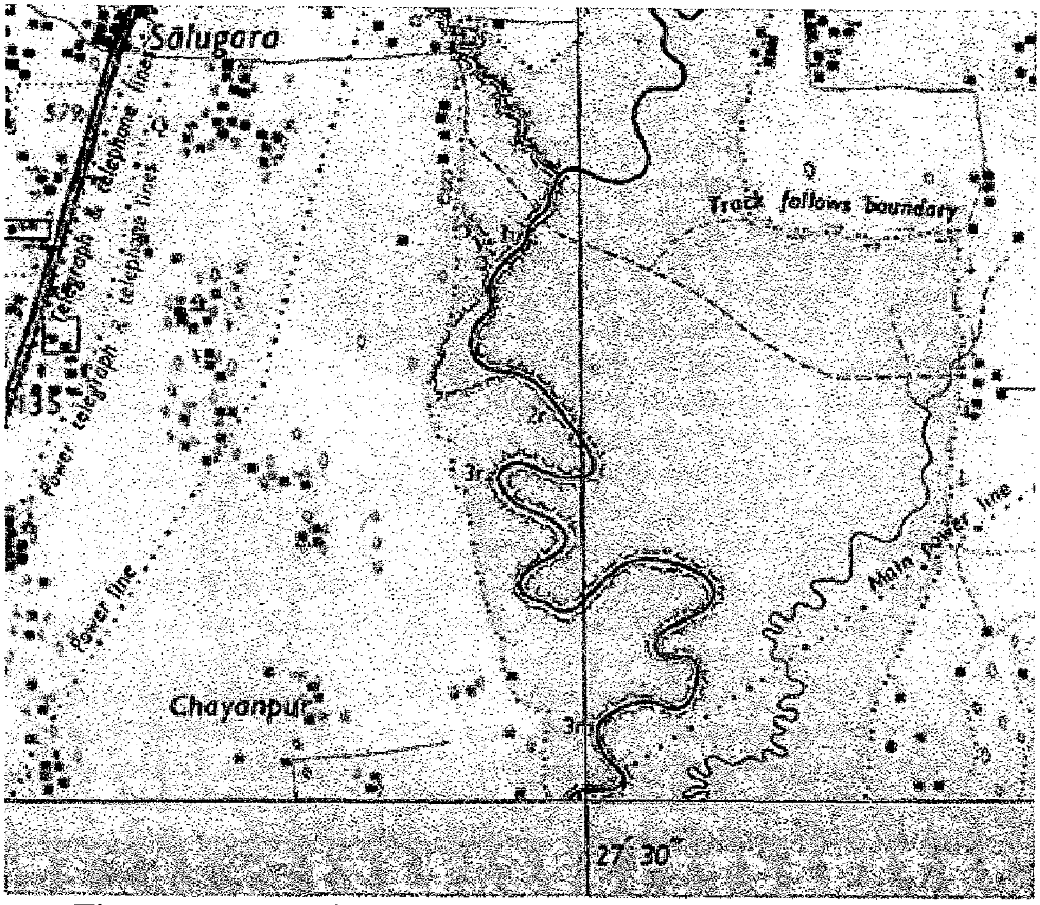

Figure 4.1 Location of Test Site (Toposheet No. 78 B/5/6)

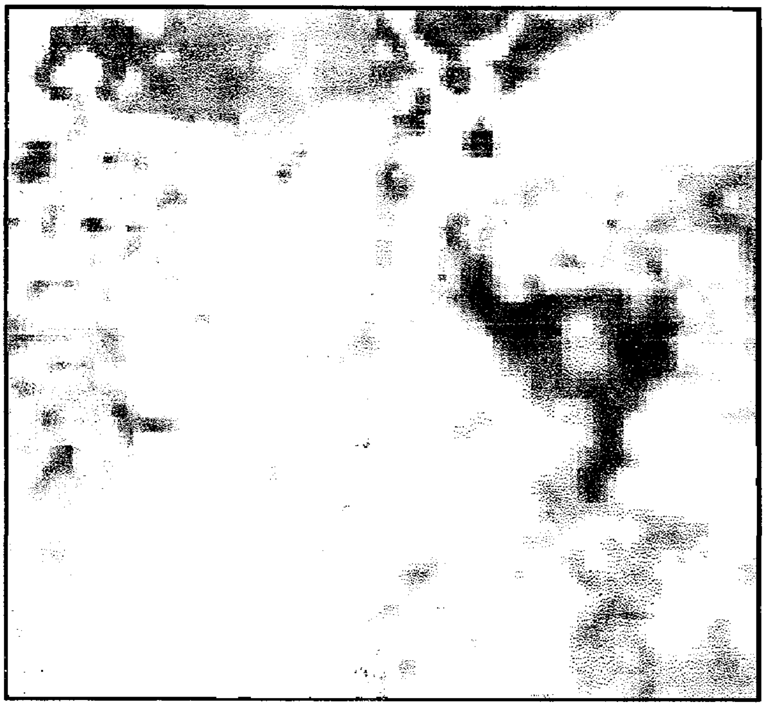

Plate 4.1 IRS 1C LISS III FCC (Red: band 4, Blue: band 2, Green: band 1)

Plate 4.2 IRS 1C PAN image

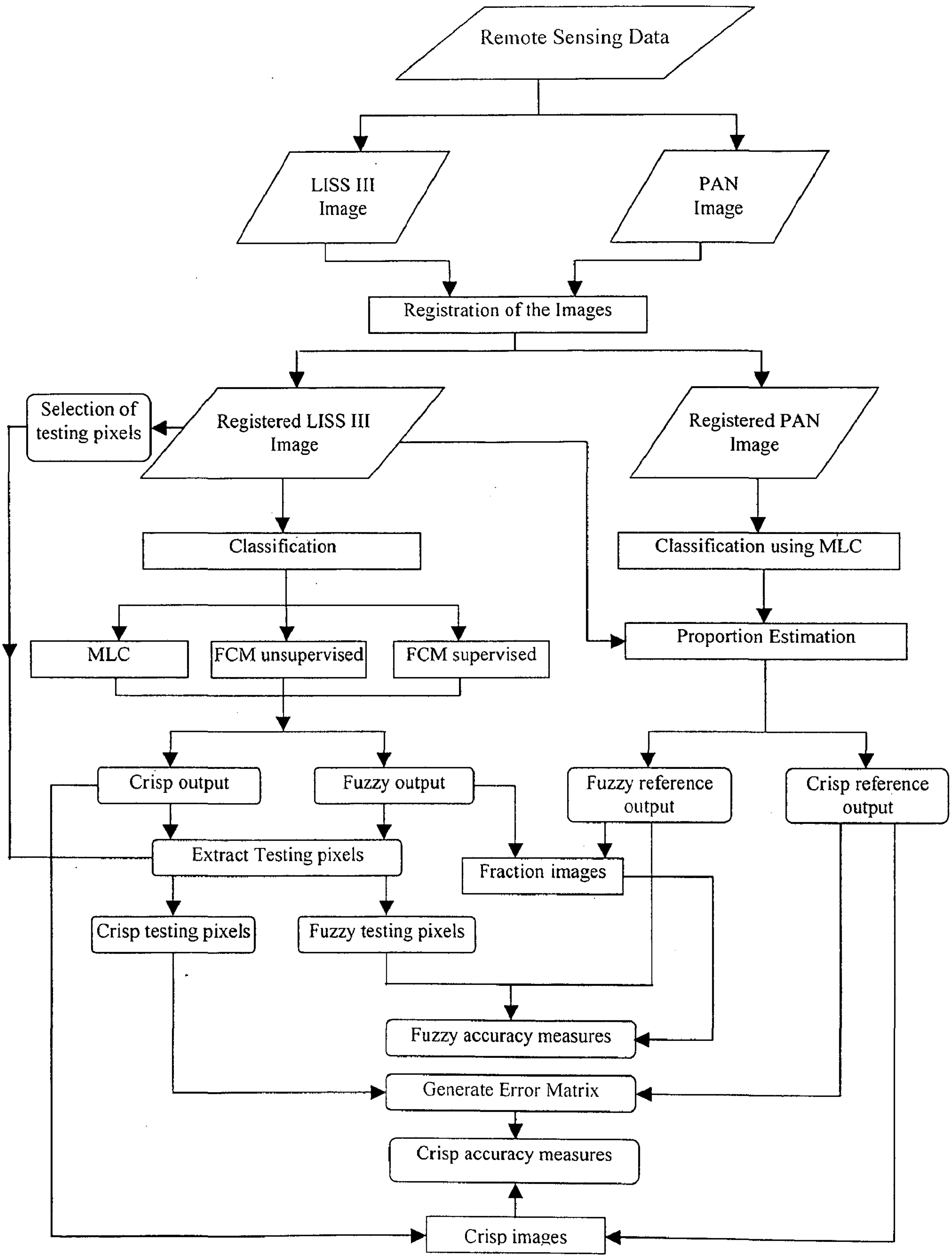

Figure 4.2 Flow Chart of Methodology Adopted

4.3.3 Proportion estimation

Since each pixel in LISS III image matches with the corresponding 25 pixels in the thematic map produced from PAN image, the LISS III image has been degraded to 5 m resolution so that each pixel in this image gets compatible with its corresponding pixel. The proportions module incorporated in the developed software (Section 3.3.4) has been used to perform this degradation. A sample of final proportions obtained for each pixel in LISS III image is shown in Table 4.1

Table 4.1 A sample of Proportions Obtained for each Pixel in LISS III Image

X (m)	Y(m)	class1	class2	class3	class4	class5
44541	2962276	0.000	0.880	0.120	0.000	0.000
44566	2962276	0.000	0.480	0.520	0.000	0.000
44591	2962276	0.320	0.480	0.200	0.000	0.000
44616	2962276	0.240	0.760	0.000	0.000	0.000
44641	2962276	0.040	0.880	0.080	0.000	0.000
44666	2962276	0.760	0.240	0.000	0.000	0.000
44691	2962276	0.120	0.000	0.000	0.200	0.680
44716	2962276	0.440	0.000	0.000	0.000	0.560
44741	2962276	0.280	0.120	0.000	0.000	0.600

4.3.4 Classification of IRS 1C LISS III image

The LISS III image has been classified using supervised MLC and FCM (unsupervised and supervised). Both crisp and fuzzy classifications have been performed. The image has been classified into five land cover classes as have been considered for the classification of PAN image. Table 4.2 shows the number of training pixels used for supervised classifications.

Table 4.2 Number of Training Pixels used for Supervised Classifications

Agriculture	Built up areas	Sandy areas	Forest	Grassland
997	286	279	1596	805

The quality of the training data has been checked by examining the histogram generated for each class in each band. A typical histogram for the class grassland in band 4 is shown in Plate 4.6. All the histograms are not uni-modal in shape, this demonstrating the existance of mixed pixels in the image. Therefore, it is anticipated that fuzzy classification would be more appropriate than crisp ones.

Thus a total of six image classifications were performed using various combinations.

4.3.5 Accuracy assessment

The final stage in classification is to evaluate the accuracy of remotely sensed derived thematic maps. The crisp and fuzzy classifications produced from LISS III image have been evaluated using appropriate accuracy measures as described earlier. For effective evaluation, the sample size and locations of testing pixels in each classification have been kept same. A total of 650 testing pixels have been selected randomly for this purpose.

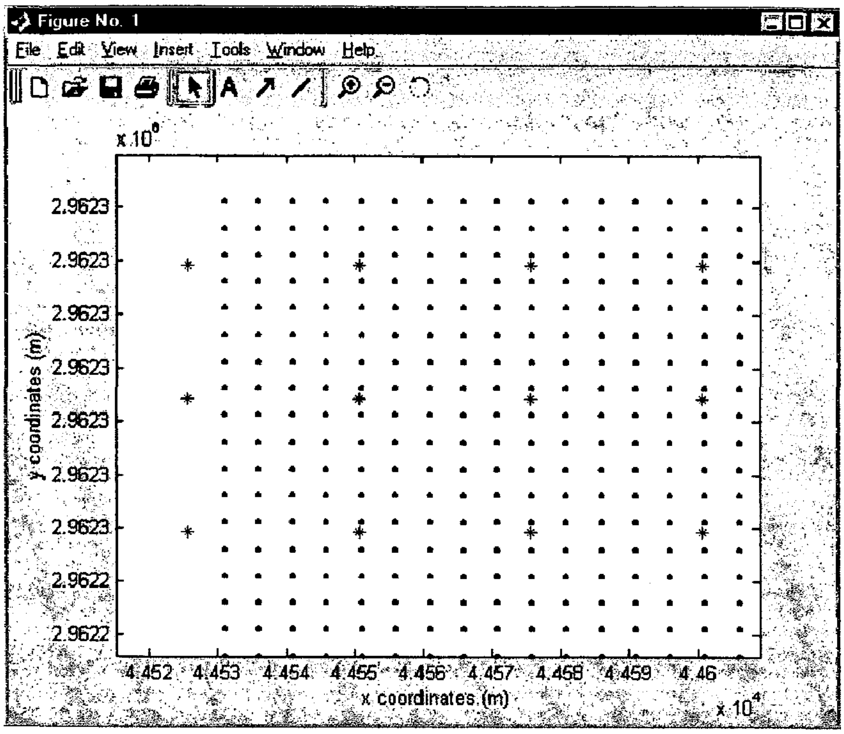

Plate 4.3 Coordinate Plots for Pixels in LISS III and PAN Image before Registration
(* LISS III / • PAN)

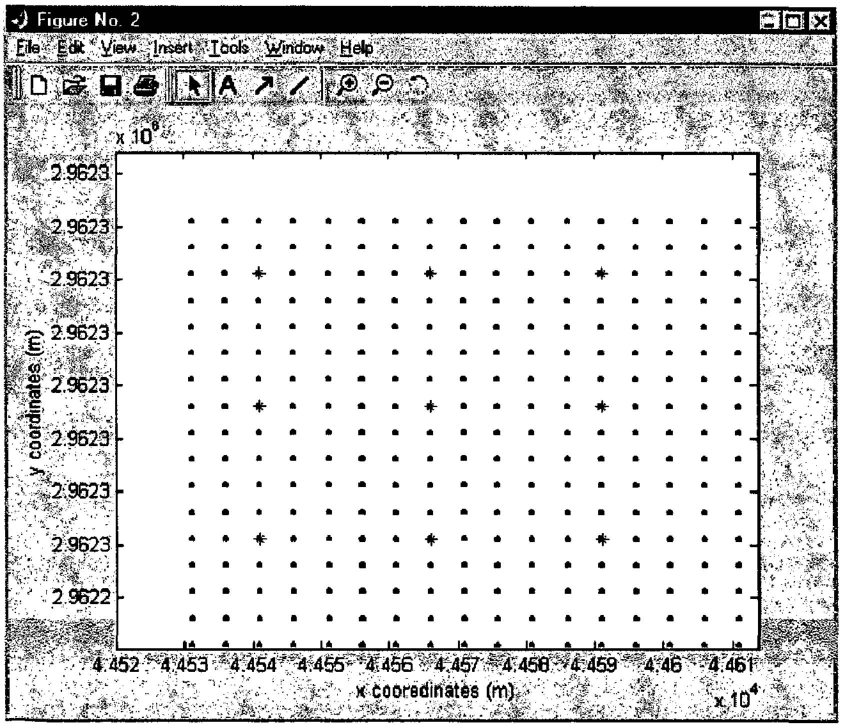

Plate 4.4 Coordinate Plots for Pixels in LISS III and PAN Image after Registration
(* LISS III/ • PAN)

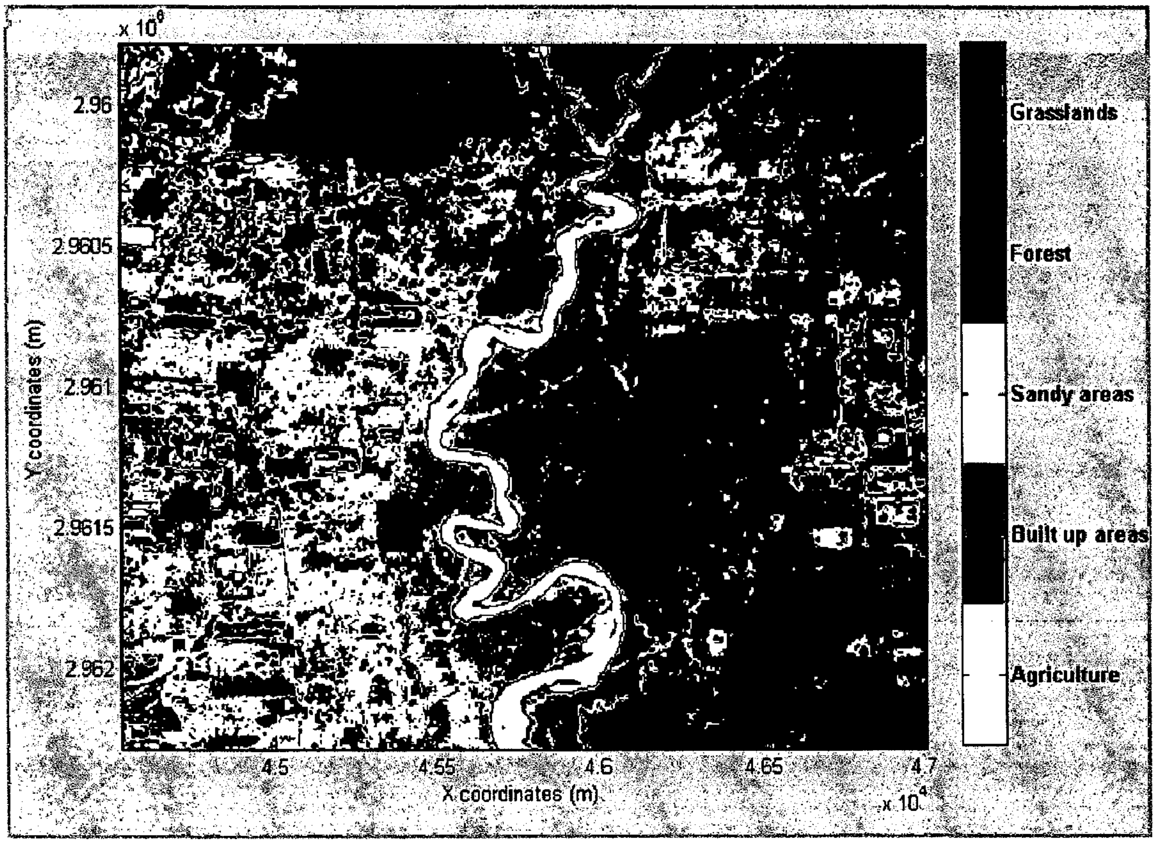

Plate 4.5 Classified PAN image used as Reference Data

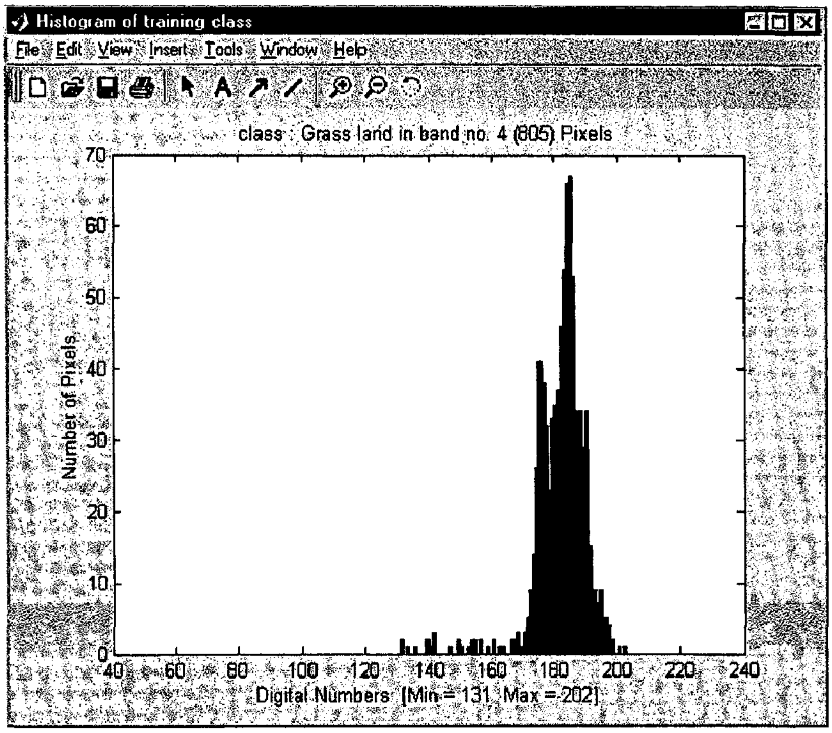

Plate 4.6 A Typical Histogram for Class Grassland in Band 4

RESULTS AND DISCUSSIONS

5.1 General

The LISS III image has been classified using three classifiers namely MLC, FCM supervised, and FCM unsupervised. The outputs of classifications are in crisp and fuzzy modes, and the reference data is also in crisp and fuzzy modes. Hence, the classifications for LISS III image have been assessed using appropriate accuracy measures. In this chapter, a brief discussion on the results obtained has been reported.

5.2 Evaluation of Classifiers in term of their Accuracy

5.2.1 Accuracy evaluation of crisp classifications

The error matrices generated for each crisp classification produced from three classifiers are shown in Tables 5.1 to 5.3.

For each of these error matrices, various crisp accuracy measures for whole classification have been computed from the software and are reported in Table 5.4

Table 5.1 Error Matrix Generated from MLC Classification

Ground Data (Refrence)

		Agriculture	Built up areas	Sandy areas	Forest	grassland	Row total
Classified Data	Agriculture	14	5	0	0	7	26
	Built up	37	39	5	1	17	99
	Sandy areas	1	7	17	0	4	29
	Forest	14	2	0	143	49	208
	grass land	70	13	0	57	148	288
	column total	136	66	22	201	225	650

Table 5.2 Error Matrix Generated from FCM Supervised Classification

Ground Data (Refrence)

		Agriculture	Built up areas	Sandy areas	Forest	grassland	Row total
Classified Data	Agriculture	76	18	0	6	70	170
	Built up	6	28	4	1	7	46
	Sandy areas	1	4	18	0	2	25
	Forest	26	13	0	131	53	223
	grass land	27	3	0	63	93	186
	column total	136	66	22	201	225	650

Table 5.3 Error Matrix Generated from FCM Unsupervised Classification

Ground Data (Refrence)

		Agriculture	Built up areas	Sandy areas	Forest	grassland	Row total
Classified Data	Agriculture	71	26	3	4	70	174
	Built up	14	8	0	46	14	82
	Sandy areas	2	5	19	0	2	28
	Forest	16	11	0	90	49	166
	grass land	33	16	0	61	90	200
	column total	136	66	22	201	225	650

Table 5.4 Crisp Accuracy Measures for Whole Classification

		MLC	FCM Supervised	FCM Unsupervised
Overall Accuracy		0.555	0.532	0.428
Average Accuracy	User's	0.544	0.573	0.435
	Producer's	0.567	0.573	0.471
Combined Accuracy	User's	0.550	0.552	0.431
	Producer's	0.561	0.553	0.450
Kappa coefficient		0.384	0.361	0.231
Tau coefficient (equal probabilities)		0.444	0.415	0.285
Tau coefficient (unequal probabilities)		0.391	0.360	0.217

It can be seen that although the accuracies as reported by each classifier are on a lower side, MLC and FCM (supervised) produced significantly higher accuracies than those produced by FCM (unsupervised). The difference in accuracies of MLC and FCM (supervised) are very marginal. This demonstrates the superiority of the supervised classifiers over unsupervised one for the data set considered.

The individual accuracies of each class for each classifier have also been determined and are shown in Table 5.5

Table 5.5 Crisp accuracy Measures for Individual Class

	MLC		FCM Supervised		FCM Unsupervised	
	User's	Producer's	User's	Producer's	User's	Producer's
Agriculture	0.538	0.103	0.447	0.559	0.408	0.522
Built up	0.394	0.591	0.609	0.424	0.098	0.121
Sandy	0.586	0.773	0.720	0.818	0.679	0.864
Forest	0.688	0.711	0.587	0.652	0.542	0.448
Grassy	0.514	0.658	0.500	0.413	0.450	0.400

The user's accuracy determines the accuracy of individual classes. For example, a user's accuracy of 53.8% for the agriculture in classified image using MLC classifier represents the actual agriculture on the reference data. In contrast, the producer's accuracy of 10.3% for the agriculture in the reference data is represented correctly as agriculture in the classified image. Similar conclusions can be drawn for other classes also. Since the sandy areas give the highest values, therefore, it may be inferred that in the classified image, sandy areas are good representation of actual areas on the ground.

On further examining Table 5.5, another important conclusion can be drawn. The least user's and producer's accuracies has been reported by the FCM (unsupervised) for the built up class. It clearly shows the poor classification of this class by FCM (unsupervised). This indicates that built up class is highly mixed or in confusion with other classes and its performance may become better than other classes in fuzzy classification. However, the values of individual class accuracies in general do not show any trend as to which classifier has performed better than the other in assessing the accuracy of an individual class.

Plate 5.1 shows the outputs of the crisp classifications generated by the software. These have been compared with the reference data as obtained from PAN image in crisp form. On examining these displays, the user is able to make visual evaluation of the classification. It is clear that the visual quality of the MLC crisp classification is more close to the reference data than the other classifiers. This confirms with the quantitative evaluation done earlier.

Plate 5.1 Crisp Classification Outputs

5.2.2 Accuracy evaluation of fuzzy classifications

The fuzzy classifications produced from different classifiers have been evaluated using various fuzzy accuracy measures obtained from the software and are given in Table 5.6

Table 5.6 Fuzzy Accuracy Measures for Whole Image

	MLC	FCM Supervised	FCM Unsupervised
Entropy	0.526	0.565	0.397
Cross-entropy	0.262	0.287	0.419
Distance measure	0.057	0.060	0.092
Information closeness	0.145	0.160	0.193

It can be inferred from the entropy values that the MLC and FCM (supervised) classifiers produced classifications with higher fuzziness as compared to FCM (unsupervised). Higher entropy measure indicates that the membership values of a pixel are well distributed among the classes. This is the situation with MLC and FCM (supervised). This proves the capability of these two classifiers to produce fuzzy classification. In other words, FCM (unsupervised) classification has not been able to portray the fuzziness in the image.

However, entropy value gives no indication of whether the proportions obtained are close to the fuzzy reference data. The cross-entropy, distance measures and information closeness may be used to express the quality of the fuzzy classification. A small value from these measures indicates that the classification is an accurate representation of the thematic data. Looking at the values of these measures for MLC and FCM (supervised), it can be stated that these classifiers have significantly lower values than the FCM (unsupervised). This again demonstrates the superiority of these classifiers over the FCM (unsupervised) even for fuzzy classifications.

To evaluate the performance of each class by a particular classifier, the values of cross-entropy (d), distance measure (S), and correlation coefficient (r) have been also computed and reported in Table 5.7

Table 5.7 Fuzzy accuracy Measuress for individual class

	MLC			FCM Supervised			FCM Unsupervised		
	d	S	r	d	S	r	d	S	r
Agriculture	0.101	0.067	0.590	0.048	0.063	0.495	0.064	0.094	0.473
Built up	0.021	0.045	0.507	0.027	0.033	0.626	0.101	0.099	-0.030
Sandy	0.008	0.007	0.854	0.007	0.007	0.860	0.004	0.008	0.845
Forest	0.056	0.077	0.708	0.109	0.100	0.583	0.126	0.131	0.402
Grassy	0.076	0.089	0.402	0.095	0.095	0.366	0.125	0.127	0.262
Average	0.052	0.057	0.612	0.057	0.060	0.586	0.084	0.092	0.390

From Table 5.7, it can be observed that the class sandy area has been classified as the most accurate class by all the measures for the three classifiers. The negative correlation does not provide any information on built up area but the other two measures suggest that the class is highly mixed and more fuzzier than the other classes. This is an outcome that supports the earlier conclusion derived while discussing crisp classifications. Though no specific trend can be seen for different classes by all the measures, in general, MLC has produced more actual proportions than others.

In order to evaluate the classifications visually, the fuzzy classification outputs have been used to generate fraction images to portray the spatial distribution of the five land cover classes. In the fraction images, the bright areas denote higher proportion of a class. To evaluate these images with the reference data, a fraction image for each class representing actual class compositions obtained from PAN reference image has also been generated (Plate 5.2). From this, it can be observed that for all the cases, MLC has provided the best relationship of the proportions with the reference data.

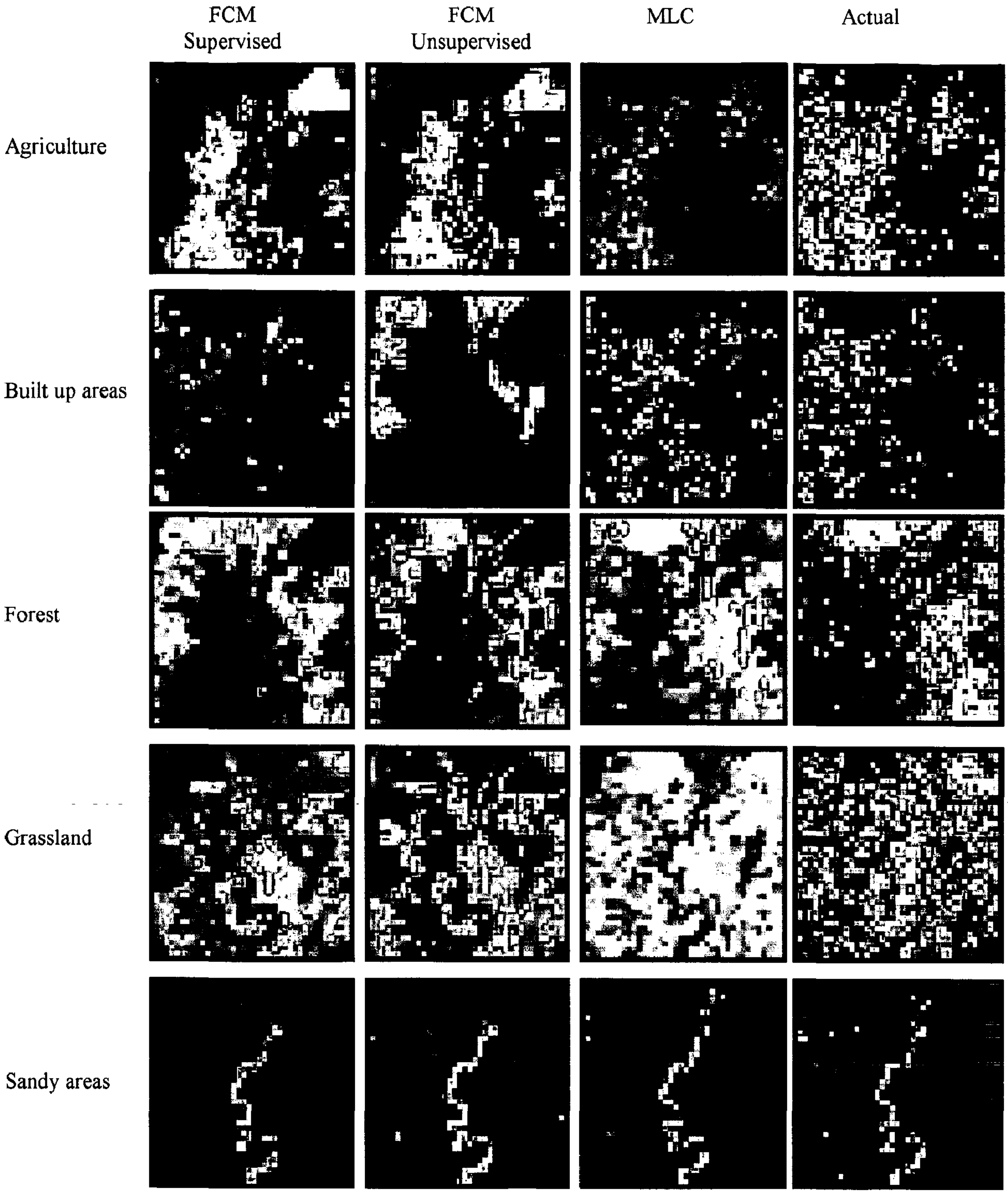

Plate 5.2 Fraction Images for each Classifier Compared with actual Proportions in Reference Data

5.3 A Comparison between various Accuracy Measures

5.3.1 Crisp accuracy measures

In order to assess the performance of a given measure, the accuracy measures for one particular classifier namely FCM (supervised) has been reported and discussed in this section. Thus, for this classifier, the various accuracy measures for the whole image are shown in Table 5.8. It may be emphasized that the comparison shall be meaningful if studied with the corresponding error matrix in Table 5.2.

Table 5.8: Crisp Measures for Whole Classification

Overall accuracy		0.532308
Average Accuracy	User's	0.572640
	Producer's	0.573264
Combined Accuracy	User's	0.552474
	Producer's	0.552786
Kappa Coefficient		0.360768
Weighted Kappa		0.173747
Tau Coefficient based on equal probability (T_e)		0.415385
Tau Coefficient based on unequal probability (T_p)		0.359811

The overall accuracy is an index of classification as a whole. However, it does not take into account the off-diagonal elements of the error matrix. Therefore, it will not give idea about the commission and omission errors

It can be seen that average accuracy is greater than overall accuracy because it has the tendency of bias towards the class having high percent of correctly classified samples (i.e., sandy areas), while overall accuracy is biased towards the class with less percent of correctly classified samples (i.e., grassland). The combined accuracy is balancing the overall accuracy and average accuracy. However, the values of overall, average and combined accuracy are more than the Kappa coefficient. Thus, these measures tend to overestimate the classification accuracy. Hence, Kappa may prove to be a desirable accuracy measure, because it has the ability to account for chance agreement as it uses all the elements of the error matrix. Kappa however gives equal weights to each class. Many a times, some classes may be more confused with each other than other classes. Therefore, some weights may be attached and a weight matrix generated to determine weighted Kappa. The weight matrix should be taken carefully, keeping in mind confusion between the various classes. The weights used in this study are shown in Table 5.9

Table 5.9 Weights Used to Obtain Weighted Kappa

	Agriculture	Built up	Sandy	Forest	Grassy
Agriculture	0	7	2	4	9
Built up areas	7	0	7	2	2
Sandy areas	2	7	0	1	1
Forest	4	2	1	0	10
Grassland	9	2	1	10	0

From the error matrix (Table 5.2), it may be observed that the most confusion exist between forest and grassland areas, therefore, the most weight has been given to these pair of classes. The least confusion exist between the forest and sandy areas and, hence, these classes have been given the least weight. After assigning these weights, the weighted Kappa obtained shall probably be the most realistic estimate of the accuracy. Though the weights are completely subjective.

Initial comparison between Kappa and Tau coefficients, reveals that Kappa coefficient constantly overestimates the chance agreement and underestimates the classification accuracy relative to Tau. Hence, Tau coefficient provides an intuitive and accurate measure of classification accuracy. Also, Tau may be viewed as the ratio between the number of pixels that were correctly grouped by random assignment. Thus, for the classification based on equal probability, the T_e value indicates that 41.5% or more pixels have been classified correctly than would be expected by random assignment. Similarly, the T_p value indicates that the classification, based on an unequal probability, makes 36% fewer errors than it would be expected by random assignment. In this respect, Tau values may be easier to understand and interpret than Kappa. Unlike Kappa, Tau also compensates for unequal probabilities of groups. The unequal probabilities used to calculate T_p in this study are shown in Table 5.10

Table 5.10 Unequal Probabilities for each Class for Tau Coefficient

Agriculture	Built up areas	Sandy areas	Forest	Grassland
0.23	0.09	0.04	0.29	0.35

Looking at the individual class accuracies obtained by various measures (Table 5.11), it may be concluded that conditional Kappa and conditional Tau may be used for the assessment of accuracy of individual classes after making some compensation for chance agreement. Hence, the values of conditional Kappa are less than user's and producer's accuracies.

Table 5.11 Crisp Accuracy Measures for individual class

	Accuracy per class		Conditional Kappa		Conditional Tau	
	User	Producer	User	Producer	User	Producer
Agriculture	0.447059	0.558824	0.300755	0.402574	0.252782	0.427044
Built up areas	0.608696	0.424242	0.564473	0.380393	0.579243	0.367299
Sandy areas	0.720000	0.818182	0.710191	0.810909	0.708333	0.810606
Forest	0.587444	0.651741	0.402758	0.469864	0.374915	0.509495
Grass land	0.500000	0.413333	0.235294	0.178161	0.295775	0.097436

5.3.2 Fuzzy accuracy measures

Fuzzy accuracy measures for the classification produced by FCM supervised have been reported in Table 5.12

Table 5.12 Accuracy Measures for Fuzzy Classification by FCM Supervised

	Average for Whole Image	The values for each class				
		Agriculture	Built up areas	Sandy areas	Forest	Grass land
Entropy	0.5647	0.1322	0.1183	0.0603	0.1151	0.1389
Cross-entropy	0.2865	0.0483	0.0272	0.0070	0.1087	0.0953
Distance	0.0600	0.0633	0.0334	0.0077	0.0999	0.0954
Information closeness	0.1596	0.0369	0.0335	0.0134	0.0386	0.0371
Correlation coefficients		0.4948	0.6257	0.8604	0.5825	0.3664

The high value of entropy indicates high fuzziness. In other words, a large number of pixels are mixed.

Other measures, such as, cross-entropy and generalized measure of information closeness may also be used. It may be observed that these two measures give a higher value for whole image than distance measure. This indicates that simple measure of distance may overestimate the classification accuracy as a whole. Also, it may be inferred that the cross-entropy, distance measure, and information closeness are have similar trend in expressing the accuracy of fuzzy classifications.

From the entropy for each class, it can be observed that sandy area has the smallest value than other classes, this indicates that the pixels containing sandy areas are having a higher probability of membership with this class. Therefore, this class is highly related to the actual reference data. In contrast, the grassland class has the highest value, which indicates the existance of precious proportions from other classes within the pixels in this class. From Table 5.12, it may also be observed that sandy areas have the least distance. Therefore, this class is a well representation of the actual class.

The correlation coefficients are used to indicate the accuracy of the classification on per-class basis. Higher the correlation coefficients, higher are the classification accuracy. It may be seen the sandy area has the highest correlation coefficient and thus is more close to the actual proportions as given in reference data.

CONCLUSIONS AND FUTURE SCOPE

6.1 Conclusions

The objective of this work is to focus the attention on the assessment of accuracy of classification of thematic maps derived from remote sensing images. Necessary software to perform the classifications in crisp and fuzzy form and their accuracy measures has been written in MATLAB environment. The program has been tested on a sample data set from remote sensing images obtained from IRS satellite. On the basis of the results obtained, following conclusions may be drawn:

1. RSICAA is an interactive user-friendly Window based software that provides various options to the user to perform a classification and use an appropriate accuracy measure.

2. Data format adopted in RSICAA is simple. Further, it allows portability between other commercial software packages such as the well known ERDAS Imagine.

3. As per the nature of classification output (i.e., crisp or fuzzy), proper accuracy measures may be used to derive qualitative information from thematic classifications.

4. For the data set considered, MLC and FCM supervised classifier have generally produced higher accuracies than FCM unsupervised classifier.

5. Depending upon whether the aim is to determine the accuracy of the whole classification or the accuracy of one class, a user may select a particular accuracy measure.

6. Though the Kappa coefficient has been used extensively, the current study shows that it may underestimate the classification accuracy. Therefore, Tau

coefficient is attractive as it provides an intuitive and accurate measure of classification accuracy. Also, this coefficient is easier to understand and interpret than Kappa. Unlike Kappa, Tau also compensates for unequal probabilities of classes or for difference in number of classes.

7. For fuzzy classification, the entropy value gives no indication of whether the proportions obtained are close to the reference data or not. Therefore, distance measures seem more appropriate. However, simple distance measure may overestimate the classification accuracies. Therefore, cross-entropy and measure of information closeness may be more appropriate.

6.2 Future scope

Though all care has been taken to develop a versatile software, no package is ultimate. The modifications are always necessary and therefore newer versions shall keep coming. There are some points that crept in while working on this software.

1. Other classifiers such as Linear Mixture Modeling (LMM) and Artificial Neural Network (ANN), which yield fuzzy classified images, can be incorporated in this package.

2. The effect of different factors affecting the classification accuracy such as training and testing data characteristics and number of wavebands may be studied with the help of this package.

Appendix A

This appendix shows a sample of the output file for crisp classification accuracy measures generated by RSICAA. A typical error matrix has been used in the computation of various accuracy measures.

Error Matrix : (Source: Arora and Ghosh, 1998)

Ground Data (Reference)

		Forest	Built up	Range land	Water	row total
Classified Image	Forest	310	20	0	0	330
	Built up	60	120	0	0	180
	Range land	2	4	60	0	66
	Water	30	20	0	10	60
	column total	402	164	60	10	636

Number of correctly classified: 500

Overall Accuracy = 0.786164

User's Accuracy :

0.939394
0.666667
0.909091
0.166667

Producer's Accuracy :

0.771144
0.731707
1.000000
1.000000

Average Accuracy(User's) = 0.670455

Average Accuracy(Producer's)=0.875713

Combined Accuracy(User's)= 0.728309

Combined Accuracy(Producer's)=0.830938

Kappa Coefficient = 0.636198

Weights used to calculate Weighted Kappa :

$$\begin{bmatrix} 0 & 2 & 1 & 10 \\ 2 & 0 & 1 & 7 \\ 1 & 1 & 0 & 1 \\ 10 & 7 & 1 & 0 \end{bmatrix}$$

Weighted Kappa = 0.364344

Conditional Kappa for User's approach :

0.835276
0.550847
0.899621
0.153355

Conditional Kappa for Producer's approach :

0.524339
0.625802
1.000000
1.000000

Tau Coefficient based on equal probability = 0.714885

The unequal probability for each class for Tau Coefficient :
class 1 :0.250000 class 2 :0.250000 class 3 :0.250000 class 4 :0.250000
Tau Coefficient based on unequal probability = 0.714885

Priori probabilities of class membership for Classified Data :
class 1 :0.250000 class 2 :0.250000 class 3 :0.250000 class 4 :0.250000
Conditional Tau for User's approach :

0.919192
0.555556
0.878788
-0.111111

Priori probabilities of class membership for Refrence Data :
class 1 :0.250000 class 2 :0.250000 class 3 :0.250000 class 4 :0.250000
Conditional Tau for Producer's approach :

0.694859
0.642276
1.000000
1.000000

Appendix B

This appendix shows a sample of the output file for fuzzy classification accuracy measures generated by RSICAA. This file has been generated using the data set is used in this dissertation.

The average Entropy value for Whole Image is:
0.5264

The Entropy values for each class are:
Agriculture: 0.1006
Built up : 0.1040
Sandy : 0.0404
Forest : 0.1336
grassy : 0.1478

The average Cross-entropy (d) value for Whole image is:
0.2622

The Cross-entropy (d) values for each class are:
Agriculture: 0.1010
Built up : 0.0206
Sandy : 0.0084
Forest : 0.0560
grassy : 0.0761

The average Distance (S) value for Whole image is:
0.0569

The Distance (S) values for each class are:
Agriculture: 0.0666
Built up : 0.0447
Sandy : 0.0070
Forest : 0.0773
grassy : 0.0890

The average Measure Of information closeness (D) value for Whole image is:
0.1454

The Measure Of information closeness (D) values for each class are:
Agriculture: 0.0283
Built up : 0.0336
Sandy : 0.0084
Forest : 0.0373
grassy : 0.0379

the correlation coefficients (r) values for each class are:

Agriculture	: 0.5898
Built up	: 0.5071
Sandy	: 0.8540
Forest	: 0.7075
grassy	: 0.4015

References

Arora, M.K. and Ghosh S.K., 1998. Classification Accuracy Indices: Definitions, Comparisons And A Brief Review. Asian-Pacific Remote Sensing and GIS Journal 10 (2): 1-9.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data, Remote Sensing of Environment, 37: 35-46.

Congalton, R.G., R.G. Oderwald, and R.A. Mead. 1983. Assessing Landsat Classification accuracy using discrete multivariate analysis statistical techniques, Photogrammetric Engineering & Remote Sensing, 49: 1671-1678.

Foody, G. M., 1992. On the compensation for chance agreement in image classification accuracy assessment, Photogrammetric Engineering & Remote Sensing, 58: 1459-1460

Foody, G. M., 1995a. Land cover classification by an artificial neural network with ancillary information, International Journal of Remote Sensing, 16, 527-542.

Foody, G. M., 1995b. Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data. ISPRS Journal of Photogrammetry and Remote Sensing, 50, 2-12.

Foody, G. M., 1996b, Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data, International Journal of Remote Sensing, 17: 1317-1340.

Foody, G. M., 1996c, Relating the land-cover composition of mixed pixels to artificial neural network classification output, Photogrammetric Engineering & Remote Sensing, 62: 491-499.

Foody, G. M., and Arora, M. K., 1996, Incorporating mixed pixel in the training, allocation and testing stages of supervised classification, Pattern Recognition Letters, 17, 1389-1398.

Foody, G. M., and Cox, D. P., 1994, Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions, International Journal of Remote Sensing, 15, 619-631.

Foody, G. M., Campbell, N. A., Trodd, N. M. and Wood, T. F., 1992, Derivation and applications of probabilistic measures of class membership from maximum likelihood classification, Photogrammetric Engineering & Remote Sensing, 58: 1335-1341.

Fung, T. and E. LeDrew, 1988. The determination of optimal threshold levels for change detection using various accuracy indices, Photogrammetric Engineering & Remote Sensing, 54: 1449-1454.

Gopal, S., and Woodcock, C., 1994. Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets, Photogrammetric Engineering & Remote Sensing, 60(2): 181-188.

Jensen, J.R., 1986. Introductory Digital Image Processing, (New York, Prentice Hall)

Ma, Z. and R.L. Redmond, 1995, Tau coefficients for accuracy assessment of classification of remote sensing data, Photogrammetric Engineering & Remote Sensing, 61: 435-439.

Maselli, F., Conese, C., and Petkov, L., 1994, Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications, ISPRS Journal of Photogrammetry and Remote Sensing, 49(2): 13-20

Miguael-Ayanz, J.S. and G.S. Bigging, 1996, An iterative classification approach for mapping natural resources from satellite imagery, International Journal of Remote Sensing, 17, 957-981.

Naesset, E., 1995, Tests for conditional Kappa and marginal homogeneity to indicate differences between user's and producer's accuracy, International Journal of Remote Sensing, 16: 3147-3159.

Robinove, C. J. 1981. The Logic of Multispectral Classification and Mapping of Land, Remote Sensing of Environment, 11: 231-244.

Rosenfield, G. H., and Fitzpatrick-Lins, 1986, A coefficient of agreement as a measure of thematic classification accuracy, Photogrammetric Engineering & Remote Sensing, 52: 223-227.

Story, M., and Congalton, R. G., 1986, Accuracy assessment: A user's perspective, Photogrammetric Engineering & Remote Sensing, 52: 397-399.

Wang, F., 1990b. Fuzzy supervised classification of remote sensing images, IEEE Transactions on Geoscience and Remote Sensing, 28: 194-201.