1. INTRODUCTION

Remote sensing [1] is the acquisition of information about an object, without making any physical contact with that object. The various stages in Remote sensing technique are given as follows [2]:

- Emission of electromagnetic radiation (EMR) that can be sun/self- emission.
- Transmission of energy from the source to the earth, including absorption and scattering.
- Interaction of EMR with the surface of the earth: reflection and emission.
- Transmission of energy to the remote sensor from the earth.
- Sensor data as output.
- Data transmission, processing and analysis.

Now a day's sensor technology is also being used to detect and classify the objects by the means of signals propagated by satellites. A human analyst attempting to classify features in an image uses the elements of visual interpretation to identify homogeneous groups of pixels which represent various features or land cover classes of interest. ***Digital image classification*** uses the spectral information represented by the digital numbers in one or more spectral bands, and attempts to classify each individual pixel based on this spectral information. This type of classification is termed as spectral pattern recognition**.** In either case, the objective is to assign all pixels in the image to particular class or theme (e.g. water, coniferous forest, deciduous forest, corn, wheat, etc.).

Image classification plays an important role in the area of remote sensing. The intent of the classification process is to categorize all the pixels of a Multi spectral image into one of the several land cover classes such as water, vegetation, urban, rocky and barren; the resultant data can then be used to create thematic maps which are used for digital image analysis.

At present there exist several traditional classifiers such as Minimum Distance Classifier (MDC), Maximum Likelihood Classifier (MLC) etc which are being applied in the field of image classification but they are not able to provide efficient solutions. To solve this problem Natural Computation technique was being introduced.

Natural computation is the study of computational systems that makes use of the ideas and takes inspiration from natural systems, including biological, ecological and physical systems. It is an emerging interdisciplinary area in which a range of techniques and methods are studied in order to deal with large, complex, and dynamic problems. Though various Natural Computation techniques such as BBO [3], ACO [4], PSO, Artificial Neural Network, Membrane Computing etc are already been introduced as a classifier. Cuckoo Search which was recently being introduced under Natural Computation was being applied by us for image classification and for the resolution of mixed pixels.

1.1. Motivation

Satellite images capture the land cover features some of which cover significantly large area, while some (e.g. bridge and roads) occupy relatively much smaller regions. Image does not contain each feature independently; some features are intertwined with other features. Due to mixture of feature at some portion of images the concept of mixed pixels came into existence. Images contain various topographic traits and it is the specific task of the analysts to identify these features contained in an image. Image classification is one of the important approaches for recognizing these terrain features. In image classification domain we have huge database (which is provided by experts) so we need a technique which efficiently searches all possible cases.

Most of the existing natural computation technique like Biogeography Based Optimization (BBO), Ant Colony Optimization (ACO), and Particle Swarm Optimization (PSO) take an image in the form of clusters or group of pixels due to which the classification accuracy is moderate. To overcome this disadvantage Membrane Computing was introduced which is pixel by pixel approach but this algorithm was not able to completely classify all the terrain features.

This encouraged us to develop an image classification algorithm based on one of the new techniques of natural computation or nature inspired approach known as Cuckoo search. It is a meta-heuristic algorithm that was introduced by Xin-She Yang and Suash Deb in the year 2009 in there research paper titled ―Cuckoo Search via Lévy Flights [5]. This algorithm is based on the obligate brood parasitic behaviour of cuckoo species in combination with the Lévy flight behaviour of some birds and fruit flies. Resolution of mixed pixels is one of the challenges faced by the image classification algorithms so we have also extended our algorithm to solve this problem.

Our proposed work is motivated from nature-inspired technique called Cuckoo Search technique and bases its fundamentals from the research paper titled “Cuckoo search via Levy flights” by Xin-She Yang, Suash Deb [5]. CS has two important features of meta-heuristic algorithms that are intensification and diversification [6].

Some of the advantages of Cuckoo Search based satellite image classification over some of the existing algorithms are:

- CS works with single unit of object rather than group of objects i.e. the approach is pixel by pixel.
- It has a huge search space as compared with other meta-heuristic algorithm.

Thus we have used the strength of cuckoo search for the extraction of landscape traits.

1.2. Related Work

In remote sensing the problem of satellite image classification has been solved by using the traditional classifiers such as Parallelopiped Classification [7], Minimum Distance to Mean Classification [7], Maximum Likelihood Classification [7] etc. However these techniques show limited accuracy in information retrieval for high resolution multi-spectral satellite image. Also these techniques are insensitive to different degrees of variance in the spectral response data.

To provide a solution to the above problems, soft computing techniques were introduced in remote sensing. Some of the soft computing techniques are as follows fuzzy logic [8], rough set theory [9], neural network theory, and Probabilistic reasoning and Swarm Intelligence techniques [10], Genetic algorithms, chaos theory and parts of learning theory.

However the soft computing techniques like the fuzzy classifier [8], and the rough set classifier [9] were not able to provide good result in case of ambiguity since the main goal of these techniques was to synthesize approximation of concepts from the acquired data. Hence these techniques did not provide very much accurate and efficient result even with low spatial resolution images. Also these techniques were not able to handle the crisp and continuous data separately.

The solution to the above drawbacks was provided by recently introduced concept of swarm intelligence [10] [11] which come under natural computation. A lot of techniques like ACO, BBO, PSO were introduced but they were not able to obtain high level of accuracy as they were being implemented as cluster based approach. Then hybridized approaches such as ACO2/PSO, BBO/ACO2/PSO were formulized [12]. These techniques improved the classification of multi-spectral satellite images. Though these classifiers identify the homogeneous features efficiently on an image but it does not give accurate result for the tagging of heterogeneous regions. Recently Membrane Computing was introduced in order to overcome the disadvantages of previously developed cluster based approaches but this approach was not able to classify all the terrain features.

For the resolution of mixed pixels various approach were being introduced such as ICA-aided mixed-pixel analysis of hyperspectral data in agricultural land [14], Swarm Intelligence for Mixed Pixel Resolution [15], hybridization Ant Colony Optimization and Bio Geography Based Optimization [16]. But the results obtained are not satisfying as for some of the above approaches the performance of the algorithm degrades with increase in the number of mixed pixels which are to be resolved.

To overcome the above mentioned difficulties, to provide efficient and accurate results we have proposed Cuckoo Search in image classification and also for tagging (or resolution) of mixed pixels. Cuckoo tries to find out the most suitable host nest to lay their eggs in the same way our algorithm always find the best old case that perfectly matches with recent new case (i.e. the pixel to be classified).

1.3. Problem Statement

The problem in remote sensing is to extract different terrain features such as water, vegetation, urban, rocky, barren etc for a given dataset. The images used for feature extraction are generally high resolution multi-spectral satellite images. The aim of this work is to propose a branch of natural computation i.e. Cuckoo Search for terrain feature (such as water, vegetation, urban, rocky etc.) extraction and for the resolution of mixed pixels. Cuckoo search is a new technique which is recently being introduced under natural computation. Cuckoo search is basically a model of computation that is motivated by the reproduction strategy of Cuckoos. Most of the natural computation technique proposed in land cover feature extraction method uses clusters for each feature but in Cuckoo Search we did not use clusters rather we considered each pixel as a stand alone unit to be classified and is not affected by the other set of pixels to be classified.

Classification of mixed pixel is one of the most important problems being faced by almost all the satellite image classifiers so there is a need to find an optimal solution for this. For this Swarm Intelligence based approach i.e. Cuckoo Search was being extended. Thus our problem statement can be proposed as follows-

"To develop Nature Inspired Cuckoo Search based technique that captures the terrain features for homogeneous, heterogeneous as well as mixed pixels"

1.4. Scope of the work

Our approach is to use ***Cuckoo Search in Image classification as well as in resolution of mixed pixels*** to be able to perform some very important targets successfully completed. The proposed classifiers used in image classification and for resolution of mixed pixels are applied to data which is provided by Geosciences scientists. Experts have provided us with the data containing the main attributes which are necessary to know about the image. The attributes are the several band information of an image. Depending on the training set provided by the experts the query pixel is classified into one of the land cover classes. Our classifiers performed the classification successfully thus our classifier can be used even for the real life applications. It is a nature-inspired image classifier that classifies the image efficiently and effectively.

The classification algorithm has been applied to seven band Indian Resourcesat Satellite image of dimension 472 X 546 of Alwar Region in Rajasthan and six band satellite image of dimension 641 X 641 of Saharanpur Region in Uttar Pradesh (since they contain a variety of land cover features) to classify the area into one of the land cover classes and then assigning color codes to them hence solving the problem of land cover feature extraction. The resolution of mixed pixels using Cuckoo Search is done for the Alwar dataset.

The efficiency of meta-heuristic algorithm is that they imitate the best features in nature, especially the selection of the fittest in biological systems have evolved by natural selection over millions of years [13]. The Kappa coefficient (K) is usually considered for evaluating and analyzing the accuracy of the proposed approach in remote sensing domain. The main advantage of calculating Kappa coefficient is the ability to use this value as a basis for determining the statistical significance of any matrix or the differences among matrices. The above discussion can be summarized as follows:

- To develop Cuckoo Search based algorithm for the land cover feature extraction.

- Test the developed algorithm for Alwar and Saharanpur regions.
- To evaluate the accuracy of the classification process by calculating kappa coefficient.
- Comparative analysis of CS based image classification algorithm for the homogeneous and heterogeneous regions with the existing algorithms.
- To extend the above algorithm for the resolution of mixed pixels.

1.5. Organization of thesis

The remainder part of this thesis is organized in the following sections:-

Chapter 2: Presents introduction of remote sensing concepts. Then it covers details about how a satellite image is formed and how it is transformed into digital number representation. It also gives a brief explanation about homogeneous regions, heterogeneous regions and mixed pixels in an image. Then it covers the basic introduction to image classification

Chapter 3: State of art where we present overview of all the techniques that had been used till date in image classification. In this traditional, soft computing, semantic web based approach and natural computation methods are described.

Chapter 4: This section gives a detailed description of cuckoo search. This is the technique that we have applied and hence is a new addition to the area image classification techniques. It‘s modeling and its efficiency in such environment is also shown in this section.

Chapter 5: This section provides the proposed algorithm of Cuckoo search in multi-spectral satellite image classification. It also presents the overall mathematical formulation of the proposed algorithm.

Chapter 6: The proposed framework of Cuckoo Search for resolution of mixed pixels is presented here. It describes the Cuckoo Search algorithm in respect to feature extraction for the mixed pixels.

Chapter 7: The first section presents the result of the case study which is based on the experiments performed at DTRL Lab-DRDO for the classification of satellite image of Alwar area in Rajasthan and Saharanpur area in Uttar Pradesh using our proposed algorithm. The second section provides the result for the experiment performed at DTRL Lab-DRDO for mixed pixel resolution for the satellite image of Alwar, Rajasthan, India. Our algorithm has shown correct results in almost all the cases as provided.

Chapter 8: Here the result of the image classification algorithm is analyzed and is compared with other soft computing technique; it has shown that Cuckoo Search is at par with other technique.

Chapter 9: The details of publication from the thesis and the conference are covered in this section.

Chapter 10: In this section the conclusion of the thesis work and the future scope of the work are presented.

Chapter 11: This section gives the reference details of the thesis.

Appendix A: Abbreviations used.

Appendix B: Introduction to MATLAB software.

Appendix C: Introduction to ERDAS IMAGINE.

2. Remote Sensing and Image Classification

In this chapter we shall explain the remote sensing procedure, characteristics of remote sensing images and the various stages in the process of image classification. It also provides a detailed description about homogeneous regions, heterogeneous regions and mixed pixels in an image.

2.1. A Glimpse of Remote Sensing?

"*Remote sensing is the science (and to some extent, art) of acquiring information about the Earth's surface without actually being in contact with it. This is done by sensing and recording reflected or emitted energy and processing, analyzing, and applying that information.*"

Remote sensing [18] **procedure** requires an interaction between the incident radiation and the target. This is diagrammatically shown in figure 1. Not only has this it also requires sensing of the emitted energy and the use of non-imaging sensors.

- **Energy Source (A) –** the first and foremost requirement for remote sensing is to have an energy source which provides electromagnetic energy to the target taken into consideration.
- **Radiation and the Atmosphere (B) –** as the energy traverses from the source to the target of interest, it will come in contact with the atmosphere. This interaction may take place second time as well when the energy traverses from the target to the sensor.

- **Interaction with the Target (C)** - once the energy reaches the target after passing through the atmosphere, it interacts with the target depending on the characteristics of both the target as well as the radiation.

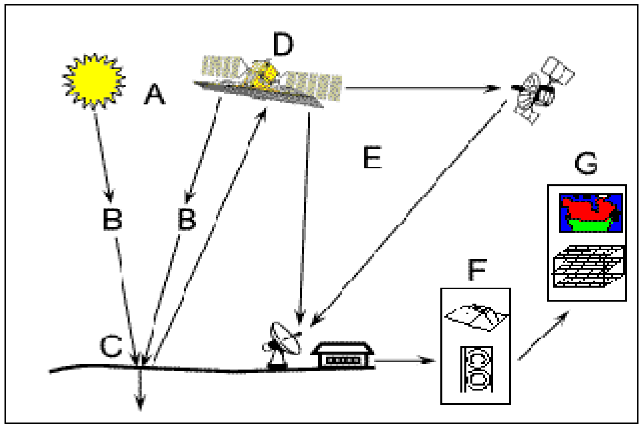


Fig 1: Capturing Image from Satellite

- **Recording of Energy by the Sensor (D)** - after the energy is scattered, or emitted from the target, a sensor (remote - not in physical contact with the target) is used in order to collect and record the electromagnetic radiation.
- **Transmission, Reception, and Processing (E)** - the energy recorded by the sensor is transmitted to a receiving and processing station in a electronic form where the data are processed into an image.
- **Interpretation and Analysis (F)** - the processed image is interpreted, visually or digitally or electronically, to extract information about the target which was taken into consideration.
- **Application (G)** - the final element of the remote sensing process is to apply the information that is being extracted from the imagery about the target in order to better understand it or to use it in solving a particular problem.

2.1.1. Electromagnetic Radiation

As explained above, the foremost requirement for remote sensing is to have an **energy source to illuminate the target** (unless the energy is being emitted by the target itself). This energy is in the form of electromagnetic radiation. All electromagnetic radiation has some fundamental characteristics and behaves in certain ways on the basics of the wave theory.

"*Electromagnetic radiation* *(**EM radiation** or **EMR**) is a form of energy emitted and absorbed by the charged particles, which exhibits wave-like behavior as it travels through space*". EMR has two components one is electric field and other is magnetic field. Both the components stand in a fixed ratio of intensity to each other, and oscillate in phase perpendicular to each other. This is shown in figure 2. In vacuum, EMR propagates with the speed of light. There are two properties of electromagnetic radiation which are crucial for understanding remote sensing. These are the **wavelength and frequency.**

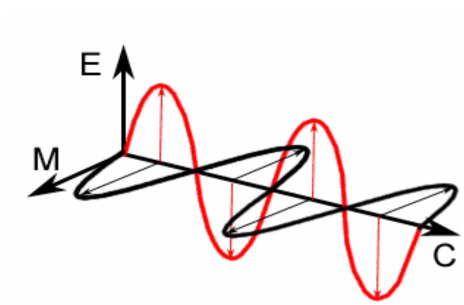


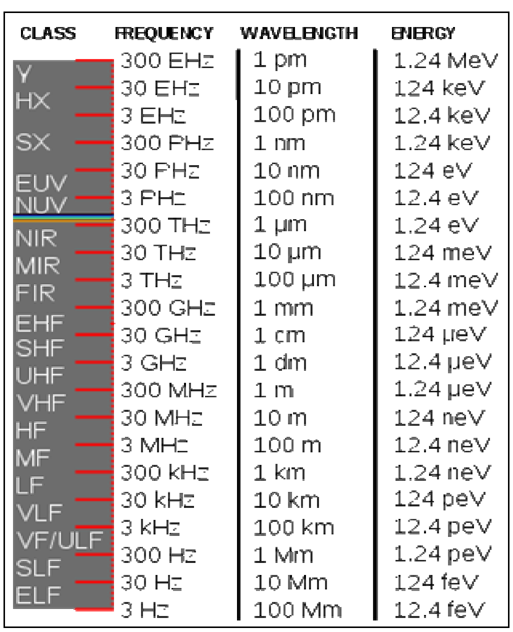
Fig 2: Electromagnetic Radiation

"*Wavelength is the length of one wave cycle, which can be measured as the distance between successive wave crests*". Its unit of measurement is **lambda (λ)**.

"*Frequency refers to the number of cycles of a wave passing a fixed point per unit of time*". Frequency is usually measured in **hertz (Hz)**. The two are inversely related to each other i.e. Longer the wavelength, lower the frequency and vice versa. Understanding the characteristics of EMR in terms of their wavelength and frequency is very crucial in understanding of the information to be extracted from the remote sensing data.

2.1.2. Electromagnetic Spectrum

The **electromagnetic spectrum** is the range of all the possible frequencies EM radiation. The EM spectrum of an object is the characteristic distribution of EM radiation emitted or absorbed by that object. The **electromagnetic spectrum** which is shown in figure 3 ranges from the shorter to the longer wavelengths. There are several portions of the EM spectrum which are useful for remote sensing. For most of the purposes, **ultraviolet or UV** region of the spectrum (which has the shortest wavelengths) are used for remote sensing. This radiation is just beyond the violet portion of the visible wavelengths, hence its name. Some Earth surface materials, especially rocks and minerals, emit visible light when illuminated by UV radiation.



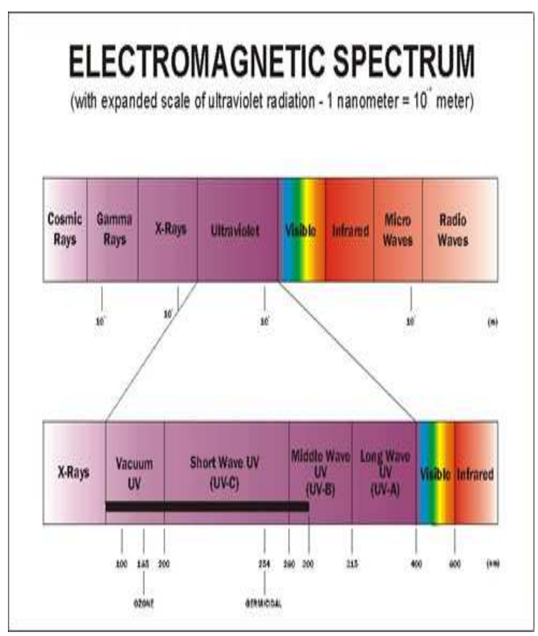


Fig 3: Electromagnetic Spectrum

Fig 4: Ultra Violet Portion of Spectrum

The light which our eyes can detect is part of the **visible portion** of the EM spectrum. But there are lots of radiations around us which are "not visible" to our eyes, but can be detected by some specialized remote sensing instruments and thus can be used to our advantage. The visible wavelength ranges from approximately 0.4 to 0.7 μm. It is only portion of the spectrum that can be associated with the concept of **colors**.

2.1.3. The Radiation

Radiations which can not be absorbed or scattered in the atmosphere can reach and interact with the surface of the Earth. There are three forms of interaction that can take place when energy is **incident (I)** upon a surface. These are given as follows:

- **Absorption (A)**;
- **Transmission (T)**; and
- **Reflection (R)**.

Absorption (A) occurs when radiation is absorbed by the target object while transmission (T) occurs when radiation passes through the target object. Reflection (R) occurs when radiation "bounces" or reflects back from the target and is redirected. In the field of remote sensing, we are most interested in measuring the radiation reflected from the target object. There are two types of reflection which are shown in figure 5 and 6. It represents the two extreme ways in which energy is reflected from a target: **specular reflection** and **diffuse reflection**.

When a surface is smooth we get **specular** or mirror-like reflection where almost all of the energy is bounce back from the surface in a single direction. When the surface is rough the energy is reflected almost uniformly in all directions and this phenomenon is referred to as **diffuse reflection**. Most earth surface features are neither perfectly specular nor perfectly diffuse reflectors. Whether a particular target reflects the energy specularly or diffusely, or somewhere in between, all depends on the roughness of the surface.

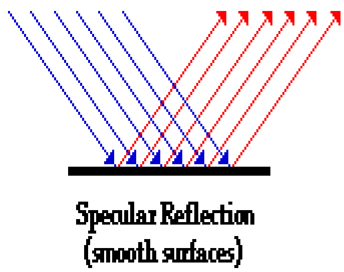


Fig 5: Specular reflection

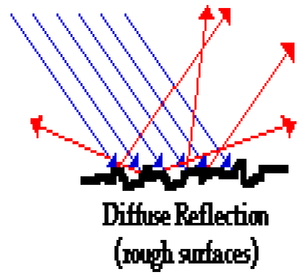


Fig 6: Diffuse Reflection

2.1.4. Passive vs. Active Sensing

So far, we have made various references to the sun as a source of energy. Sun acts as a very convenient source of energy for remote sensing. The sun's energy is either **reflected**, (as it is for visible wavelengths), or absorbed or **reemitted** (as it is for thermal infrared wavelengths). **Passive sensors** are remote sensing systems which measure the energy that is naturally available. For all the objects that reflect energy, this can only take place for the time when the sun is illuminating the Earth. While the energy that is naturally emitted (such as thermal infrared) can be detected during the day as well as during night, as long as the amount of energy is sufficient enough to be recorded.

Active sensors are the ones that provide their own energy source. The sensor emits radiation which is directed towards the target which is to be examined. The radiation reflected from that target object is detected and measured by the sensor. There are various advantages of active sensors such as the ability to obtain measurements anytime, regardless of the time of day. Active sensors can be used for investigating wavelengths that are not sufficiently provided by the sun, such as microwaves. Some of the examples of active sensors are a laser fluorosensor and synthetic aperture radar (SAR).

2.2. Characteristics of Images

Electromagnetic energy can be detected in one of the following ways:

- Photographically
- Electronically

The photographic process makes use of chemical reactions that take place on the surface of light- sensitive film in order to detect and record energy variations. **Images** and **photographs** are closely related to each other in remote sensing, so we need to distinguish between them. Image refers to any pictorial representation, regardless of the wavelengths, the remote sensing device that was being used to detect and record the electromagnetic energy. Photographs are the images that have been detected as well as recorded on photographic film. Photos are normally recorded over the wavelength ranging from 0.3 µm to 0.9 µm. Based on the above discussion, we can say that all photographs are images, but not all images are photographs.

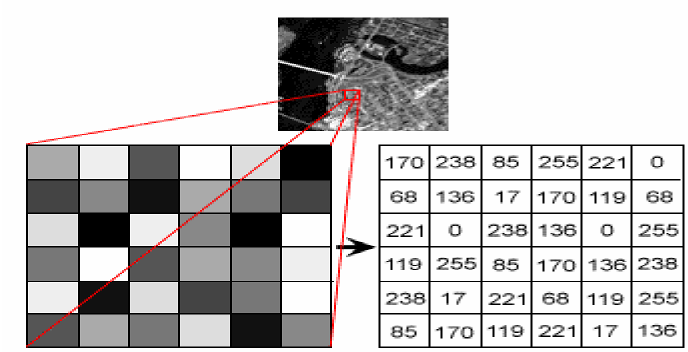


Fig 7: Digital Representation of an Image

A photograph could also be represented in a **digital** format by dividing the image into small equal-sized areas, known as the **pixels**, and representing the brightness of each area with a **digital number (**as shown in figure 7). The computer displays each of the digital value as different brightness levels.

2.2.1. Digital Numbers (DN value)

It is a positive integer value, which portrays the relative reflectance or emittance of an object in a digital image. For 8 bit images, the DN value or digital number lies in the range of 0-255. Digital image consists of discrete picture elements referred to as pixels. Associated with each pixel is a number represented as DN that reflects the average radiance of relatively small area within a scene. The range of DN values is normally from 0 to 255. The size of this area affects the reproduction details within the scene. As the pixel size is reduced more scene detail can be preserved in digital representation.

2.2.2. Types of regions in an image

There exist three types of regions in an image these are homogeneous regions, heterogeneous regions and mixed pixels. All these are explained in detail below.

Homogeneous Regions

A region in an image is said to be homogeneous if all the pixels in that region belong to the same land cover feature. For example in figure 8 the blue region represents homogeneous water region and green region represents homogeneous vegetation region.

Heterogeneous Regions

Heterogeneous region in an image are a common occurrence where a mixture of feature (more than two types of terrain features) type co-exist in a small region.

Mixed Pixels

The pixel is an explicit feature of Remote sensing, and a crucial concept of the raster GIS (Geographical Information System) which is the usual vehicle for integration [24]. Mixed pixels are ones having a signature representative of more than one land cover class (as with boundary pixels) or pixels saturated by reflectance or emittance of sub-pixel size features. The figure 8 illustrates an example of mixed pixel. In this figure the marked areas 1 and 2 shows the existence of mixed pixel in the image.

Fig 8: Illustration of mixed pixels

There are four different cases in which mixed pixel can exist they are as follows [24]:

- By the presence of small, sub-pixel targets within the area it represents.
- The pixel straddling the boundary of discrete thematic classes.
- Due the gradual transition observed between continuous thematic classes.
- Linear sub pixel.

All these are shown in figure 9.

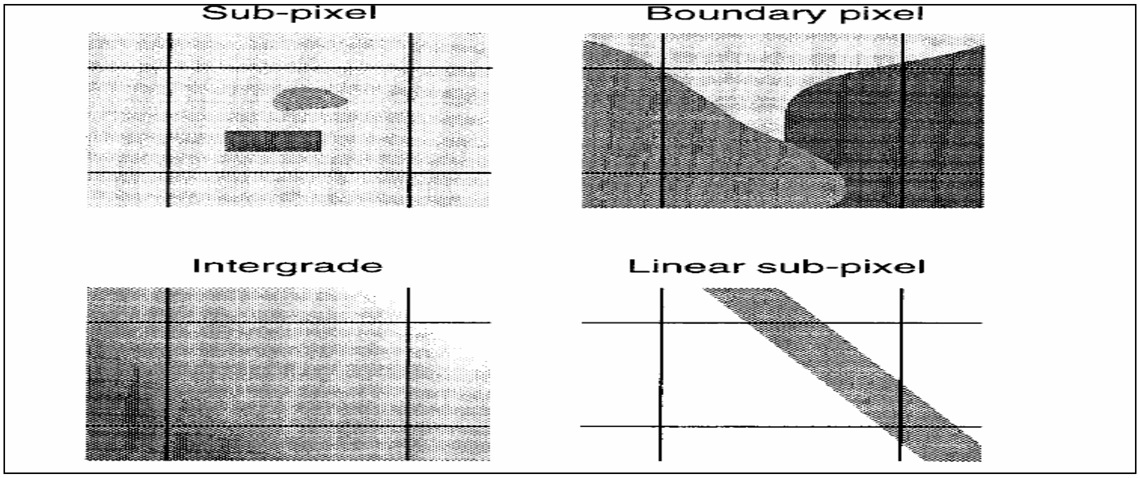


Fig 9: Different cases of mixed pixel.

2.3. Image classification procedure

The main objective of image classification procedure is to categorize all pixels of an image into one of the land cover features. The image that is obtained from remote sensors is usually a multi-spectral image and the spectral information present within it for each pixel is used as the numerical basis to categorize each of the pixels. Spectral pattern recognition belongs to the family of classification processes that utilizes spectral information of pixels as the basis for automated classification. Commonly used techniques for classification are Rough set, Rough-Fuzzy theory, and Biogeography Based Optimization etc... Some of the techniques are described below.

For example, in an agricultural map, each pixel could be assigned one of the following classes' wheat, rice, barley, or fallow. Fig 10(a) shows a picture captured by a geographical satellite. The picture obtained after applying classification process is shown in Figure 10(b).In this image water, highland regions, vegetation, and forest are classified and indicated with different colors. Since the number of classes is less than possible values for pixels we can consider classification as a simplification process. If we assume that the image was classified correctly, we can easily perform tasks such as area measurement and region extraction.

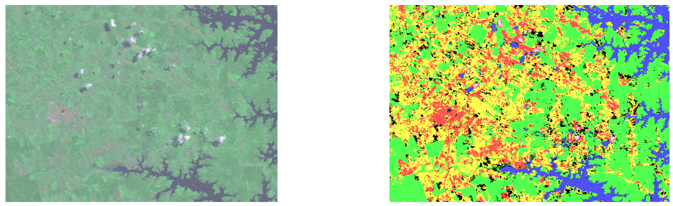


Fig 10: (a) Satellite Image (b) Classified Image

The main steps involved in the image classification procedure are shown in figure 11.

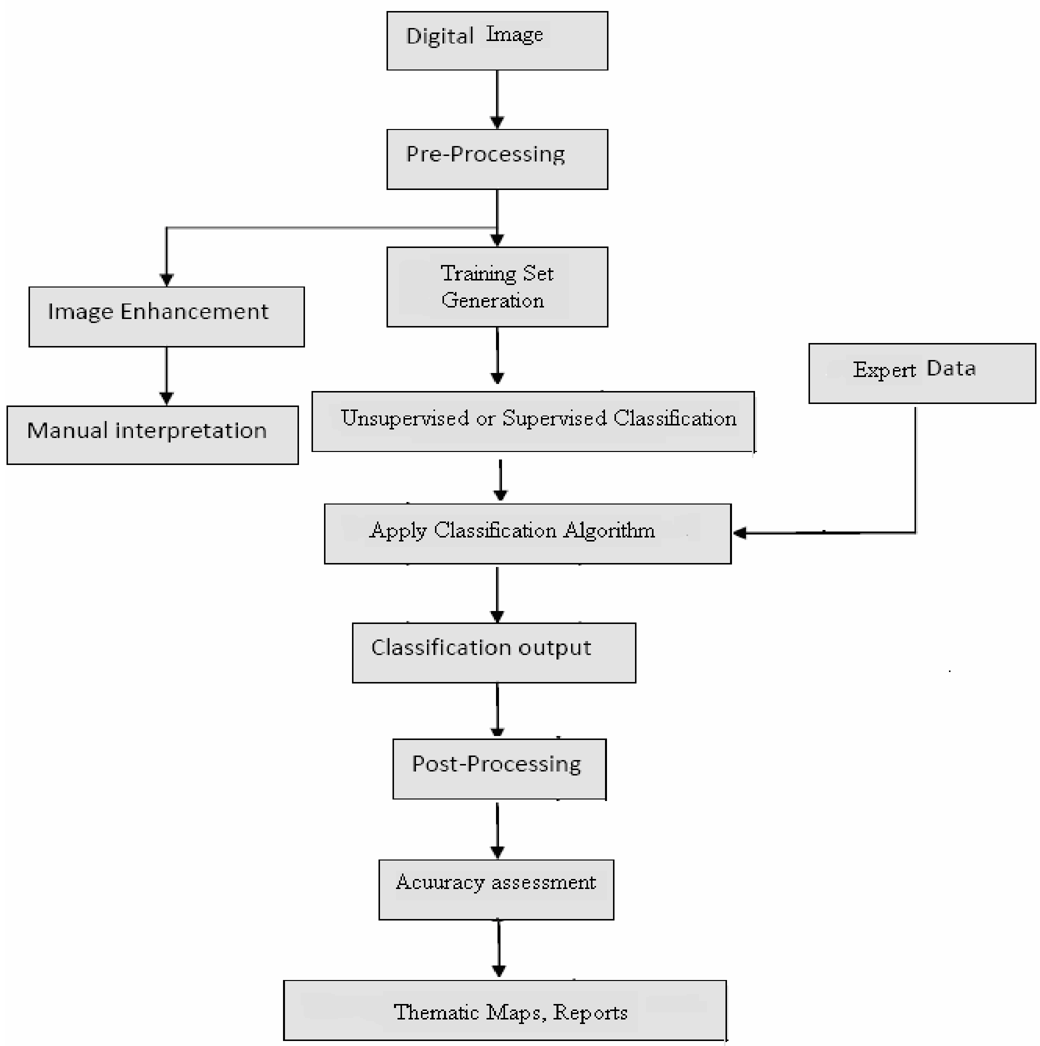


Fig 11: Image Classification Procedure

The detailed description about various steps in image classification procedure is as follows:

Step 1: Satellite image is given as input to the classification process. Usually multi spectral image is used.

Step 2: Here the pre processing of the given image is performed. In case of hyper spectral image dimensionality reduction is performed.

Step 3: From the given image training set is generated using ERDAS imagine software. Details about this software are given in appendix. Here image enhancement can also be performed which can further be used for manual interpretations about an image.

Step 4: In this step decision about classification algorithm to be applied is taken. Based on the application we can either apply supervised or unsupervised algorithm.

Step 5: In this step the algorithm selected in step 4 is applied to the input image. And if the classification is supervised expert data is used.

Step 6: The classification output obtained is stored for future analysis.

Step 7: Post processing is optional step and is performed if required by the application.

Step 8: Accuracy assessment is performed for the classification algorithm for this kappa coefficient is calculated.

Step 9: Based on the output image and accuracy assessment reports and thematic maps can be generated.

3. Image Classification Techniques

At present several Satellite Image Classification techniques exist. They can be categorized into four main types of classifiers such as:

- **Traditional Classifiers**
- **Soft Computing Techniques of Classification**
- **Semantic web based classification**
- **Natural Computation Techniques for Classification**

3.1. Traditional Classifiers

Some traditional classifying techniques are:

- **Parallelopiped Classifier**
- **Minimum distance to Mean Classifier**
- **Gaussian Maximum Likelihood**

3.1.1. Parallelopiped Classifier:

The Parallelopiped classifier comes under the category of supervised classification in which the class limits are stored within each class signature to determine if a given pixel falls within the class or not. One of the statements may be true for each of the pixel:

- If the pixel falls inside the Parallelopiped, it is assigned to that class.
- If the pixel falls within more than one class, it is put in the overlap class (with code 255).

- If the pixel does not fall inside any class, it is assigned to the null class (code 0).

This classifier is typically used when speed is the major requirement. The draw back is poor accuracy (in most of the cases) and a large number of pixels classified as ties.

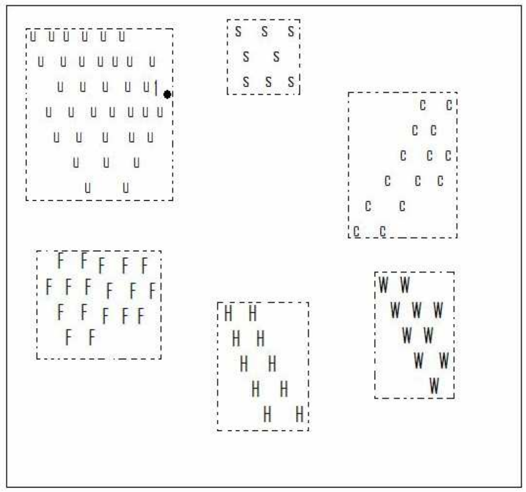


Fig 12: Parallelopiped Classifier

3.1.2. Minimum Distance to Mean Classification:

In Minimum distance to Mean classification [19] mean or average of all the spectral value in each band for each category is calculated. For each category mean vector is obtained after performing the above mentioned step. The pixels can be classified by computing the distance between the value of unknown pixel and each of category mean. The unknown pixel is put in the class to which it closely belongs. The main advantage of minimum-distance-to-means strategy is that it is mathematically simple and computationally efficient, but it has certain limitations. But the major disadvantage is that it is insensitive to different degrees of variance in the spectral data. This classifier is not used in applications where spectral classes are closely related to one another in the measurement space and have high variance.

3.1.3. Gaussian Maximum Likelihood Classification:

The maximum likelihood classifier [19] quantitatively computes both the variance and as well as the covariance of the category spectral information when classifying an unknown pixel. In this process an assumption is made that the distribution of the cloud of points forming the category training data is distributed. In Dimensional graph the vertical axis is associated with the probability value of a pixel that is a member of one of the classes. The bell-shaped surfaces thus formed are called probability density functions, and there is one such function for each spectral category. These functions are used to classify an unknown pixel by computing the probability value of the pixel belonging to each category.

The major disadvantage of this approach is that the decision rule is based on the probability value. The basic equation assumes that these probabilities are same for all classes, and that the input bands have normal distributions. It also assumes that the histograms of the bands of data have normal distributions, which is not always true and hence not applicable in all domains.

3.2. Soft Computing Techniques of Classification

Soft computing differs from conventional (hard or traditional) computing since unlike hard computing, it is tolerant of imprecision, uncertainty and partial truth. In effect, the role model or the basis for soft computing is the human mind. Principle of soft computing technique is as follows: Exploit the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness and low solution cost. The main constituents of soft computing are fuzzy logic [8], rough set theory [9], neural network theory and probabilistic reasoning, chaos theory and parts of learning theory [11].

Some Image classification techniques are:

- **Fuzzy Classifier.**
- **Rough set theory.**

3.2.1. Fuzzy Classifier:

Fuzzy classification attempts to handle the mixed pixel problem by employing the fuzzy set concept, in which a given a pixel may have partial membership in more than one classes. One approach to fuzzy classification is fuzzy clustering. This procedure is similar to “K-Means” unsupervised classification. The difference is that instead of having hard boundaries between classes fuzzy regions are established. So instead of each assigning each of measurement vectors solely to a single class, membership grade values are assigned that describes how close a pixel measurement is to the means of all the classes.

Another approach to fuzzy classification is fuzzy supervised classification. This approach is similar to maximum likelihood classification, the difference is that fuzzy mean vector and covariance matrices are developed from statically waited training dataset. For example, a vegetation classification might include a pixel with grades of 0.78 for a class “forest”, 0.18 for “street” and 0.04 for “grass”.

3.2.2. Rough Set theory:

The rough set philosophy is based on the assumption that with every object of the universe of discourse we associate certain information (i.e. data, knowledge) [9]. For example, if objects are patients suffering from a certain disease, symptoms of the disease will act as information about patients. Objects characterized by the same information are similar in view of the available information about them. The similarity relation generated in this way is the mathematical basis of rough set theory. A set of all similar objects is called an elementary set, and forms a basic granule of knowledge about the universe. Any union of some elementary sets is referred to as crisp set – otherwise the set is rough (imprecise, vague).

One of the main objectives of rough set data analysis is to reduce data size. Various notions such as indiscernibility, rough set, reduct are used to approximate inconsistent information and to exclude redundant data.

3.3. Semantic web based classification

The ontology design process is provided by Gupta 2008 [20] [21] has following steps:

- Training Set Generation
- Geo-Ontology Construction
- Matching Region Detection/ Classification

Framework for the system has been built in such a way so that if original image is available in Excel sheet format rather than TIFF format, then also we can construct ontology for it by the use of Excel to RDF/XML mapper. If the data for the image is available in tabular format of a database, then also we can build ontology for it by connecting database with protégé. The figure 13 describes the steps of this classifier.

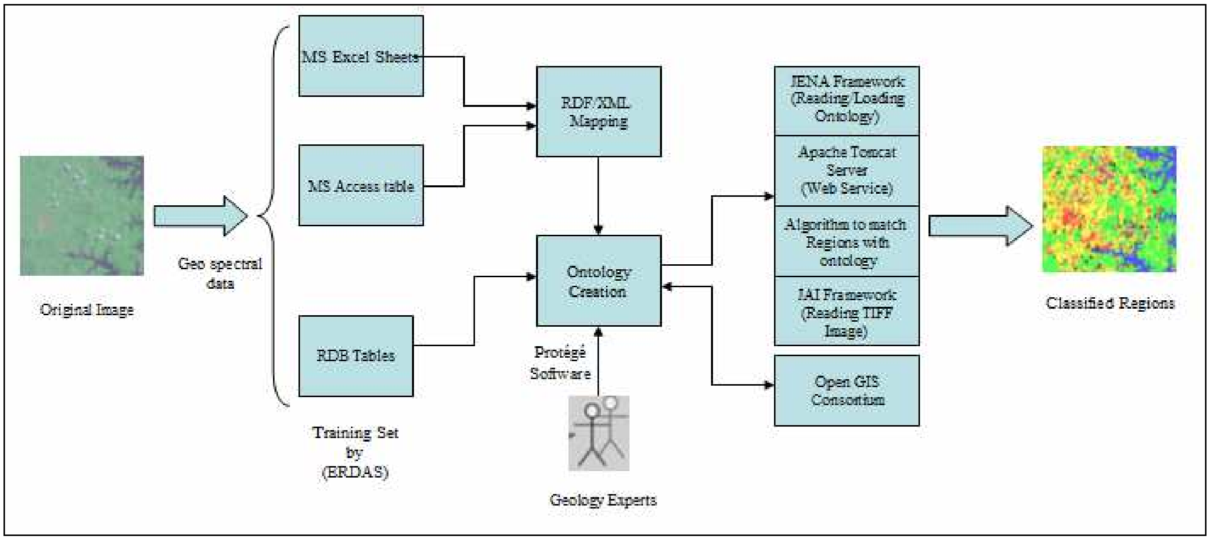


Fig 13: Framework for Semantic Classifier

3.4. Natural Computation Techniques

Some classifying techniques under this category are:

- **cAnt Miner**
- **Biogeography Based Optimization**
- **Hybridization of ACO/BBO**
- **Hybridization of ACO/PSO**
- **Hybridization of BBO/ACO2/PSO**
- **Membrane Computing**

3.4.1. cAnt Miner

Fernando, Freitas, and Johnson proposed an extension to Ant-Miner (based on ACO), named cAnt Miner [33], which was able to cope with the continuous values rather than only the discrete values. The goal of this algorithm is to extract the classification rules from the given dataset. Here we take the training set for the 7-Band Alwar image in .arff format as input to generate the classification rules from it using the Myra tool and then apply the extracted classification rules on each of the remainder clusters of the image. Each classification rule has the form IF <term1 AND term2 AND…> Then <CLASS>. Unlike Ant Miner cAnt Miner doesn't requires the discretization as a pre-processing method. So it is suitable for all the types of attributes.

3.4.2. Biogeography Based Optimization:

This is based on the concept of the Habitat Suitability Index (HSI). This HSI value is calculated for each of the pixel using 7 band values. Steps involved are

- First the dataset is divided into number of clusters.
- Then BBO is applied to each of this cluster.
- The pixels are classified into the class for which highest HSI is obtained.

3.4.3. Hybridization of ACO/BBO:

As its being mentioned above that the BBO approach is not able to classify all the land cover regions correctly. But it wrongly classifies some of the land cover features. Therefore, in order to classify the pixels efficiently, we first apply BBO algorithm and then apply ACO2 i.e. cAnt Miner Technique [29] on the remainder of the clusters of the image and thus obtain a more refined classification of the image with an improved classification accuracy [30]

3.4.4. Hybridization of ACO/PSO:

This swarm intelligence technique has been proposed by Gupta 2009 [22]. Basically work on Ant Colony Optimization for image classification has been proposed by Omkar [11].The technique uses Swarm Intelligence. It uses cAnt miner and modified hybrid PSO/ACO used for rule generation. There are training set which is fed as input to rule generation. The training set generated from original unclassified image can be in any tabular format as excel-sheet or access sheet or simple text file. Before this the training set is done unified by changing it to the Attribute Relation File Format. The cAnt miner algorithm is applied by using the MYRA tool and the Modified Hybrid PSO/ACO algorithm is applied by using the open source tool for Hybrid PSO/ACO. These algorithms provide rule set for the different regions. Then the regions are classified according to different rules.

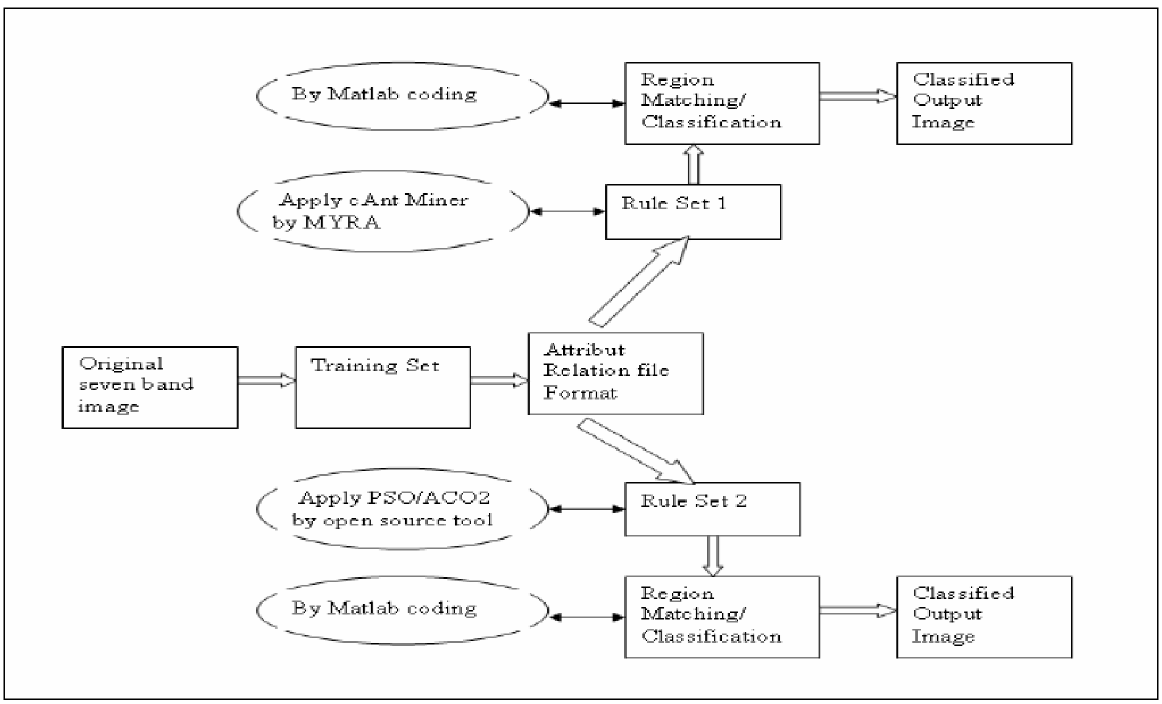


Figure 14: Diagrammatic representation of Swarm Intelligence Classifier [23]

3.4.5. Hybridization of BBO/ACO2/PSO

In the year 2010 Gupta [12] proposed a hybrid method of BBO and ACO/PSO. The previous swarm intelligence technique described ACO/PSO method in image classification. This section only describes how BBO is applied to image classification. Basically BBO has mainly 2 steps for classification: Selective cluster BBO Classification Applier layer and Feature extraction analysis layer. The main steps are described below:

(i) Selective Cluster BBO Classification Applier Layer: Hence, we apply BBO Technique on those clusters of the satellite image which show the maximum classification efficiency which is due to the fact that these are the clusters which predominantly show the presence of the feature that is most efficiently classified by the BBO Algorithm. And therefore, we apply BBO on the k^{th} cluster of the satellite image since this is the cluster which gives the maximum classification efficiency because it predominantly shows the presence of water body in the image.

(ii) BBO Feature Extraction Inefficiency Analysis Layer: BBO is not able to classify all the features with good classification efficiency and shows poor performance on some of the remaining features. In fact, in our illustration, it shows the poorest performance in classifying the urban pixels. Hence we need to proceed towards ACO2/PSO classification to improve the image classified by BBO.

3.4.6. Membrane Computing:

The main steps involved in the membrane computing based land cover feature extraction are formation of evolution rules, applying the rules, pixel absorption/rejection. The detailed description of these steps is given below:

(i) **Formation of Evolution rules:** Now the main layer of this classifier comes. All the rules that needed to be applied are formed here. Here all the rules are defined before starting of the main step i.e. applying all these evolution rules.

(ii) **Applying the rules:** This layer basically takes any combination of rules and apply rules promoter/inhibitor condition. Number of objects taken depends on the combination of rules. The rules are applied and depending on its symport the object is absorbed accordingly to that sub-membrane.

(iii) **Pixel absorption/rejection:** If the rule is satisfied then object get absorbed in a sub-membrane that is implied in the rule otherwise it remains in the membrane from where it was taken.

4. Overview of Cuckoo Search

Cuckoo Search proposed by Yang and Deb [5] is a nature inspired algorithm that is based on their aggressive reproduction strategy i.e. some species of cuckoo lay eggs in other bird’s nest in a parasitic manner.

There are three basic types of brood parasitism [25]:

- Intraspecific brood parasitism.
- Cooperative breeding
- Nest takeover.

Some species such as Ani and Guira cuckoos lay their eggs in the communal nests, though they may remove others’ eggs to improve the hatching probability of their own eggs [26]. CS satisfies two important characteristics of modern metaheuristic algorithm that are intensification and diversification [6]. Intensification refers to the fact that the problem initially searches for current best solution and then selects a global solution from the best solutions obtained while diversification means the algorithm explores the search space efficiently.

In cuckoo search three ***idealized rules*** [5] are formalized as follows:

- Each cuckoo lays one egg at a time, and dumps its egg in randomly chosen nest.
- The best nest with high quality of eggs will carry over the next generation.
- The number of available hosts is fixed and the egg laid by the cuckoo is discovered by the host nest with a probability $p_a \in [0, 1]$. In this case, the host bird can either throw the egg away or abandon the nest so as to build a completely new nest in a new location.

The fraction p_a of n nests is being replaced by new nests. For a maximization problem, the quality or fitness of a solution can simply be proportional to the objective function. Based on the above three rules an algorithm of cuckoo search is given by Yang and Deb [5].

4.1. Cuckoo Search Algorithm

Main steps involved in the process of Cuckoo search are given below [5]:

- Compare the cuckoo's egg with the set of available hosts (the number of host nests are fixed).
- Randomness is added to it by using Levy flight to choose host nests.
- The above steps produce a set of quality solutions and a set of discarded solutions.
- Based on the Ranking function a global solution is obtained from the set of quality solutions while discarding the worst nests.

When a new solutions is generated $x^{(t+1)}$ for, say, a cuckoo i, a Levy flight is performed [5] as :

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \alpha \oplus \text{Lévy}(\lambda),$$

Here $\alpha > 0$, is the step size which should be related to the scales of the problem taken under consideration. In most cases, we use $\alpha = 1$. The Levy flight essentially provides a random walk while the random step length is drawn using Levy distribution given by the equation given below [4]:

$$\text{Lévy} \sim u = t^{-\lambda}, \qquad (1 < \lambda \leq 3),$$

It has an infinite variance with an infinite mean. The steps form a random walk process with a power-law step-length distribution with a heavy tail. Some of the new solutions are being generated by Levy walk nearby the best solution obtained so far, this helps to speed up the local search. However, some fraction of the new solutions is generated by taking into consideration the randomization concept. But, whose locations should conceptually be far enough from the current best solution, this ensures that the system does not get trapped in a local optimum.

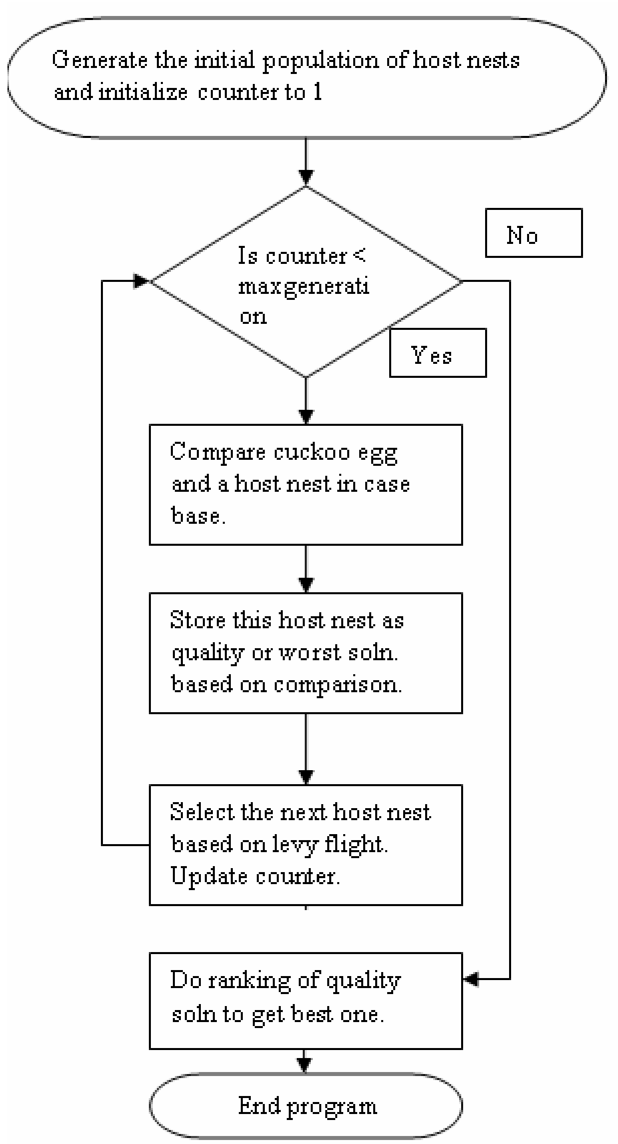


Fig 15: Cuckoo Search algorithm

Cuckoo Search [5] is a nature inspired metaheuristic algorithm which is based on their parasitic reproduction strategy, but is ***different from other existing metaheuristic*** algorithms (such as PSO, GA) in the following ways [10]:

- CS is a population based algorithm like GA and PSO but the selection mechanism is more similar to harmony search.
- The randomization is more efficient as the step length is heavy tailed and thus any large step is possible.
- Number of parameters in CS is less than GA and PSO and thus it is more general and it can be adapted to wider class of optimization problems. CS can even be extended to meta-population algorithm.

4.2. Performance Analysis on Michaelwicz function

Yang and Deb validated their algorithm in Michaelwicz function. The function looks like:

$$F(x, y) = -sin(x)\ sin^{2m}\left(\frac{x2}{\Pi}\right) - sin(y)\ sin^{2m}\left(\frac{2y2}{\Pi}\right)$$

Where m = 10 and (x, y) Є [0, 5] X [0, 5]. This function has a global minimum $f^* \sim -1.8013$ at (2.20319, 1.57049). The landscape of this function is shown in figure 16. The global optimum has been found by cuckoo search and the results are shown in figure 17 where the final locations of the nests are also marked ◊ within the figure. Here n = 15 nests, $\alpha = 1$ and $p_a = 0.25$. In the experiment n = 15 to 50.

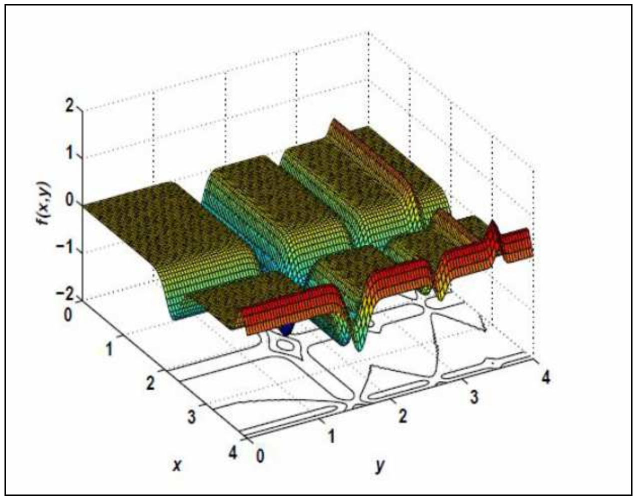


Figure 16: The landscape of Michaelwicz's function

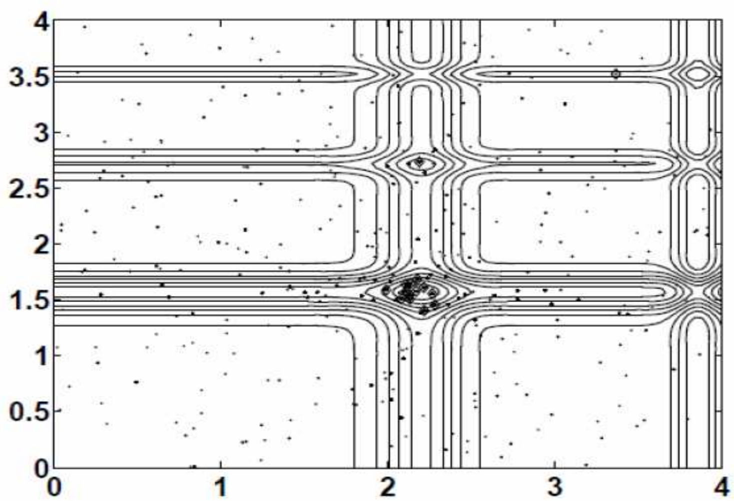


Figure 17: Search paths of nests using Cuckoo search. The final locations of the nests are marked with ◊ in the figure

From the figure 17, the optimum is approaching, most nests aggregate towards the global optimum. We also notice that the nests are also distributed at different (local) optima in the case of multimodal functions. This means that CS can find all the optima simultaneously if the number of nests is much higher than the number of local optima. This advantage may become more significant when dealing with multimodal and multiobjective optimization problems.

4.3. Comparison of CS with PSO and GA

To carry out the comparison some benchmark functions have been used. The benchmark functions are used to test the performance of optimization algorithms. The following test functions have been taken to carry out the comparative study:

- De Jong‘s first function is a sphere function

$$F(x) = \sum_{i=1}^{d} x_i^{2} \qquad x_i \in [-5.12, 5.12]$$

- Easom‘s test function is unimodal

$$F(x, y) = -\cos(x)\cos(y)\exp[-(x-\Pi)^{2} - (y-\Pi)^{2}],$$

Where, (x,y) Є [-100, 100] X [-100,100].

It has a global minimum of f*=-1 at (Π, Π) in a very small region.

- Shubert‘s bivariate function

$$F(x, y) = -\sum_{i=1}^{5} i \cos[(i+1)x+1] \sum_{i=1}^{5} \cos[(i+1)y+1],$$

Has 18 global minima in the region (x, y) Є [-10, 10] X [-10, 10]. The value of global minima is f* = -186.7309

- Griewangk‘s test function has many local minima

$$F(x) = \frac{1}{4000} \sum_{i=1}^{d} x_i^{2} - \prod_{i=1}^{d} \cos\left(\frac{x}{\sqrt{i}}\right) + 1,$$

But a single global minimum f* = 0 at (0,0 …0) for all -600≤ xi ≤600 where i = 1,2...d.

- The generalized Rosenbrock‘s function is given as ,

$$F(x) = \sum_{i=1}^{d-1}[(1 - x_i)^2 + 100\,(x_{i+1} - x_i^{\,2})^2]$$

Which has a minimum f(x*) = 0 at x* = (1, 1,…1).

- Schwefel's test function is multimodal

$$F(x) = \sum_{i=1}^{d}[-x_i \sin(\sqrt{|x_i|})], \qquad -500 \le x_i \le 500,$$

With a global minimum of f* = -418.9829d at xi* = 420.9687 (i = 1,2,….d).

- Rastrigin's test function

$$F(x) = 10d + \sum_{i=1}^{d}[x_i^{\,2} - 10\cos(2\Pi x_i)]$$

A global minimum f*=0 at (0, 0 ….0) in a hypercube -5.12≤ *xi*≤5.12 where i=1, 2... d.

Table 1: Comparison of CS, PSO and GA

FUNCTION/ ALGORITHM	GENETIC ALGORITHM	PARTICLE SWARM OPTIMIZATION	CUCKOO SEARCH
De Jong (d = 256)	25412±1237 (100%)	17040±1123(100%)	4971±754(100%)
Eastom's	19239±3307(92%)	17273±2929(100%)	6751±1902(100%)
Shubert's (18 minima)	54077±4997(89%)	23992±3755(92%)	9770±3592(100%)
Rosenbrock's (d = 16)	55723±8901(90%)	32756±5325(98%)	5923±1937(100%)
Schwefel's (d = 128)	227329±7572(95%)	14522±1275(97%)	8829±625(100%)
Rastrigin's	110523±5199(77%)	79491±3715(90%)	10354±3755(100%)
Griewank's	70925±7652(90%)	55970±4223(92%)	10912±4050(100%)

We can see that the CS is much more efficient in finding the global optima with higher success rates. Each function evaluation is virtually instantaneous on modern personal computer. For all the test functions, CS has outperformed both GA and PSO. The primary reasons are:

- A fine balance of randomization and intensification.
- Less number of control parameters.

As for any metaheuristic algorithm, a good balance of intensive local search strategy and an efficient exploration of the whole search space will usually lead to a more efficient algorithm. On the other hand, there are only two parameters in this algorithm, the population size n, and p_a. Once n is fixed, p_a controls the elitism and the balance of the randomization and local search. Few parameters make an algorithm less complex and thus potentially more generic.

5. Image Classification by applying Cuckoo Search

The main aim of this section is to describe our problem domain along with the algorithm proposed to obtain a solution. Our problem is to classify each and every pixel of an image into one of the land cover features such as water, vegetation, rocky, barren, urban etc.

5.1. Structural Design

The architectural view of the Cuckoo Search based image classifier shown in figure 18. It is a layered architecture which consists of four major components which are explained in detail as follows:

Input Layer: Multi-spectral satellite image to be classified is given as input to the algorithm. They contain spectral response value for different frequency bands.

Training the classifier: In this layer training set is generated for the image given in the input layer. This is performed using ERDAS IMAGINE software. Then this dataset is used by the classifier for learning as these are pixels for which the class to which they belong is already known.

Cuckoo Classifier: It has three sub-components which are explained as follows:

- **Generate quality solution**: First we read several multi-spectral images and then attempt to classify each and every pixel of an image. This pixel is compared with the training set based on the spectral response values obtained from the multi-spectral images of a region. Training set is the set of pixels for which we have the information about the land cover class to which it belongs. The training set pixels which show similarity with the pixel to be classified form the quality solutions. These quality solutions are stored in a database for further evaluation.

- **Perform Ranking**: Next step is to perform ranking of these quality solutions to obtain the best solution. For this correlation coefficient is calculated between the pixel to be classified and the quality solutions obtained in previous step.
- **Class Selector**: Last step is to determine the class to which the pixel will belong. For this we first examine the expert data to determine the class to which the best solution belongs, the pixel will also belong to the same class.

These steps are repeated for every pixel of the image.

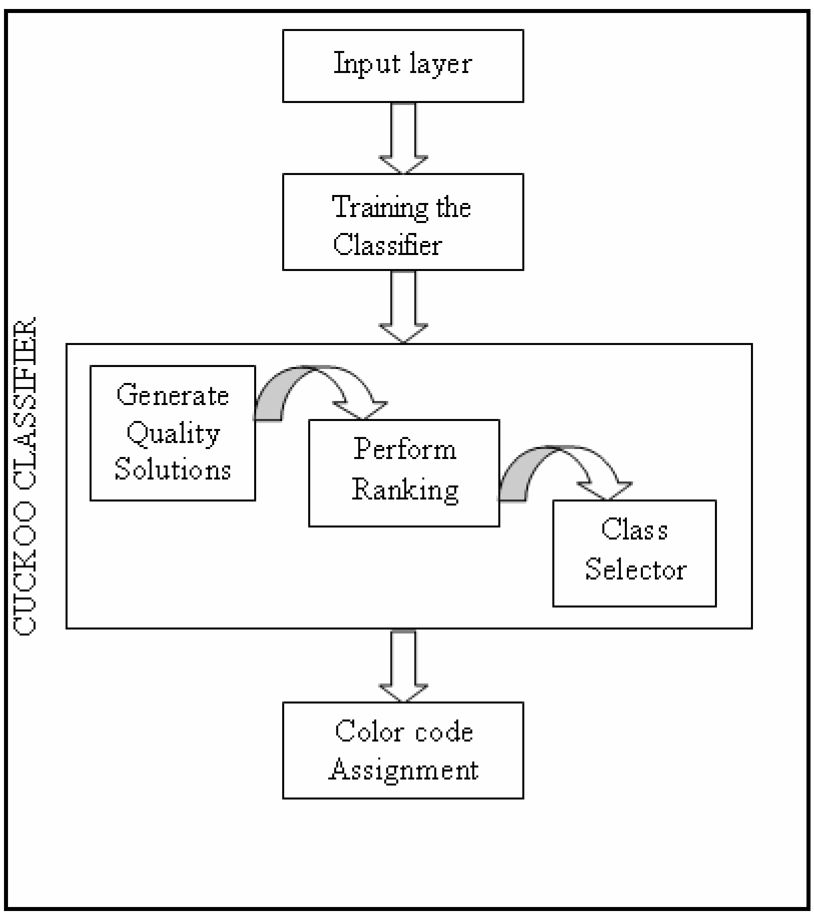


Fig 18: Structural design for image classification using CS algorithm

Colour Code assignment: The pixel of the image is assigned the colour based on the land cover feature into which it is classified. Each land cover feature is represented by a different colour.

5.2. Proposed Algorithm

The basic terminologies used in this algorithm are:

- Cuckoo egg: This is the pixel to be classified.
- Case: Each pixel of an image is represented in the form case. It is a tuple which stores the spectral responses of that pixel for different frequency bands.
- Case base: It stores the training set provided by the experts. This is used for training the algorithm.
- Host nests: These are the cases stored in the case base.
- Quality solutions: The cases in the case base that match the current query.
- Best solution: The case in the case base for which we obtain highest ranking.

Experts have provided us with the database of cases. For Alwar dataset each of the case is a tuple of the form C={R, G, NIR, MIR, RD1, RD2, DEM} and for Saharanpur dataset each of the case is a tuple of the form C= {B1, B2, B3, B4, B5, B6} where C represents the case, R represents the value of the pixel for Red band, G represents the value the pixel for Green band, NIR represents value of the pixel for Near Infra-Red band, MIR represents the value of the pixel for Middle Infra-Red band, RD1 represents the value of the pixel for Radarsat-1, RD2 represents value of the pixel for Radarsat-2, DEM represents the value of the pixel for Digital Elevation Model band and B1 to B6 represents different bands. Analyzing this case base we have to obtain the class to which the query pixel will belong using the different band values of the query pixel. The output classes (to which the pixel will belong) in case of Alwar dataset are: water, vegetation, urban, rocky, and barren and in case of Saharanpur dataset are: barren, dense, medium vegetation, sparse vegetation, urban and water. The main steps involved in our proposed algorithm are explained as follows:

Input: case base of pixels (host nests) of different land cover classes and multi-spectral images of region to be classified.

Output: classified image representing different regions by different colours.

1- Initialization

i. Generate initial population of the case base. Calculate the objective function value for all the host nests using the 7 band values.

ii. Read the set of query pixels (i.e. cuckoo eggs) which are to be classified. Calculate objective value for them.

2- Find Best Solutions

Loop until all the pixels are being classified

Find quality solutions

a. Loop to take into consideration all the host nests in the case base. Point the counter to the first host nest in the case base.

- Find similarity b/w host nest & cuckoo egg (query).(for this calculate distance between two pixels)
- Store similarity value in a database.
- Increment counter.

End Inner Loop

b. Sort the above evaluated database to find the top quality solutions.

c. The top k solutions are stored in a database and the worst nests are discarded.

d. Loop for the above obtained quality solution (k is number of quality solutions)

Do the ranking of the quality solutions by calculating Pearson correlation b/w cuckoo egg & a quality solution. (Value of the coefficient varies between -1 to +1. If the value is positive it means both the quantities are positively correlated. If it's negative then the quantities are negatively correlated. But if the value is between -0.09 to +0.09 that means there is no relation between them)

End this Loop

e. Sort the quality solutions on the basis of correlation value. The one with the highest correlation value is the best solution or the required solution. The best solution is the one that matches our cuckoo egg to great extend.

f. Find the class to which best solution belongs based on the expert data. The query will also belong to the same class to which the best solution belongs. Hence the query pixel is classified.

End Outer Loop

The representation of the above algorithm in the form of flowchart is given below:

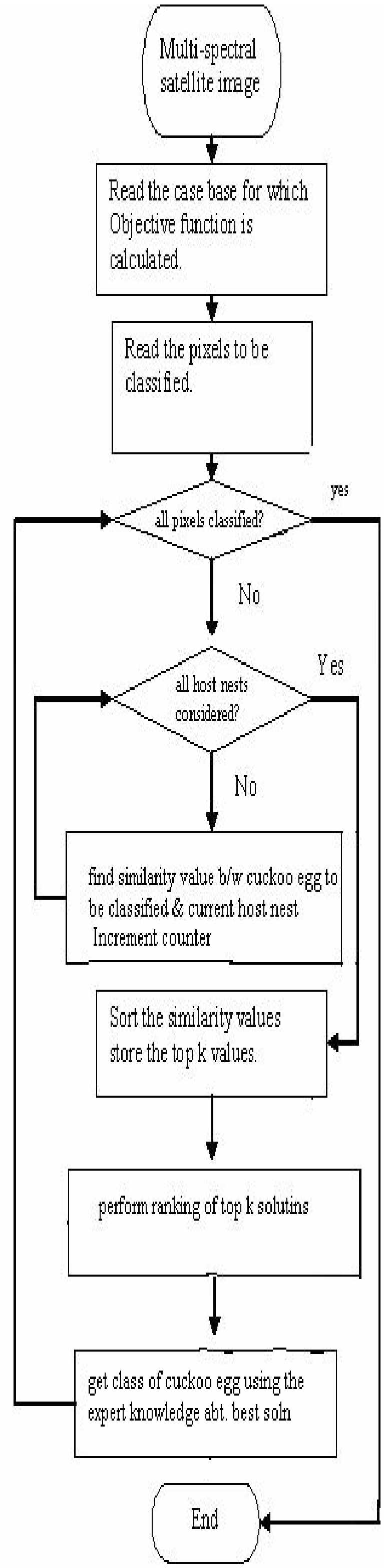


Fig 19: Cuckoo Search for image classification

5.3. Detailed description of algorithm

Step 1: In this step basically we have to find out number of cases need to be considered. Though objective function had to be found out for each case in the case base but the real evaluation is done in later steps. Thus we have considered a total of 956 cases in our case base. The left out cases (i.e. 475 pixels) are needed for validation process. Read the pixels to be classified from an excel sheet.

Step 2: This is loop for all the pixels to be classified. Initialize counter to 1, this will keep track of the current case in the case base.

- Loop to find the quality solution from a case base. For this we calculate the distance of the query pixel from every case in case base. To find the quality solutions KNN was used. In this approach we have used Euclidean distance concept. It is given by the equation:

  $$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2}.$$

  Where, i vary from 1 to 7.

  d(p,q)= distance b/w p & q.

  p_i= In our algorithm it represents the value of ith band for pixel 1.

  Q_i= In our algorithm it represents the value of ith band for pixel 2.

Where, pixel 1 and pixel 2 are the two pixels b/w which distance is to be calculated. More less the value of d(p,q), more similar the two pixels are.

- The result obtained from the above step is sorted on the basis of value of the distance. Those with the lowest value are of importance to us as they are the cases which match the most with our current query. Thus they become the quality solutions for the pixel to be classified. The k quality solutions are stored and the rest of the solutions or the worst nests are discarded. Selection of the value of k is a very critical task. If small value of k is taken then noise will have higher influence on the result. A large value of k makes it computationally expensive and will defeat the basic philosophy behind KNN. Simple approach is to select ‘k’ as the square root of number of cases in case base.

- The best solution is found from the above obtained quality solutions by using ranking. The concept used for ranking in this algorithm is Pearson product-moment correlation coefficient (sometimes referred to as the PPMCC or PCCs) [27]. This is denoted by ‘r’, given by formula:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

Where, i vary from 1 to 7.

X_i = i th band value of sample X.

Y_i= i th band value of sample Y.

X bar= mean value of all the band values of X.

Y bar= mean value of all the band values of Y.

The interpretation of the value of r is given in table1.

Table 2: Correlation value understanding

Correlation	Negative	Positive
None	-0.09 to 0.0	0.0 to 0.09
Small	-0.3 to -0.1	0.1 to 0.3
Medium	-0.5 to -0.3	0.3 to 0.5
Strong	-1.0 to -0.5	0.5 to 1.0

6. Resolution of mixed pixels by applying Cuckoo Search

Image classification is one of the most important areas of digital image analysis. There is a fundamental assumption that each pixel in an image represents properties of a single land cover feature (or class). But, this is not always the case, because of the existence of mixed pixels that show the properties of more than one land cover features [3]. At present several soft computing techniques are being used for the classification of an image, but their performance is not satisfying for the classification of mixed pixel. The aim of this section is propose an algorithm for the classification of mixed pixels.

6.1. Block Diagram

The architectural view of the Cuckoo Search based mixed pixel resolution is given below:

- The first step involved in the resolution of mixed pixels is to take satellite image as input. This is the image of the region containing the mixed pixels.
- In the next step the Geoscience's experts examine the satellite image minutely and study the different land cover features in the image.
- Analysis of image performed in previous step helps the expert to identify the heterogeneous regions in an image. Excel sheets containing the spectral response for different frequency bands are created.
- Then next step is to apply the Cuckoo Search algorithm in which we first obtain the quality solution and then the best solution.

- The tagging of the mixed pixels is achieved as the mixed pixel will belong to the same class to which the best solution belongs.

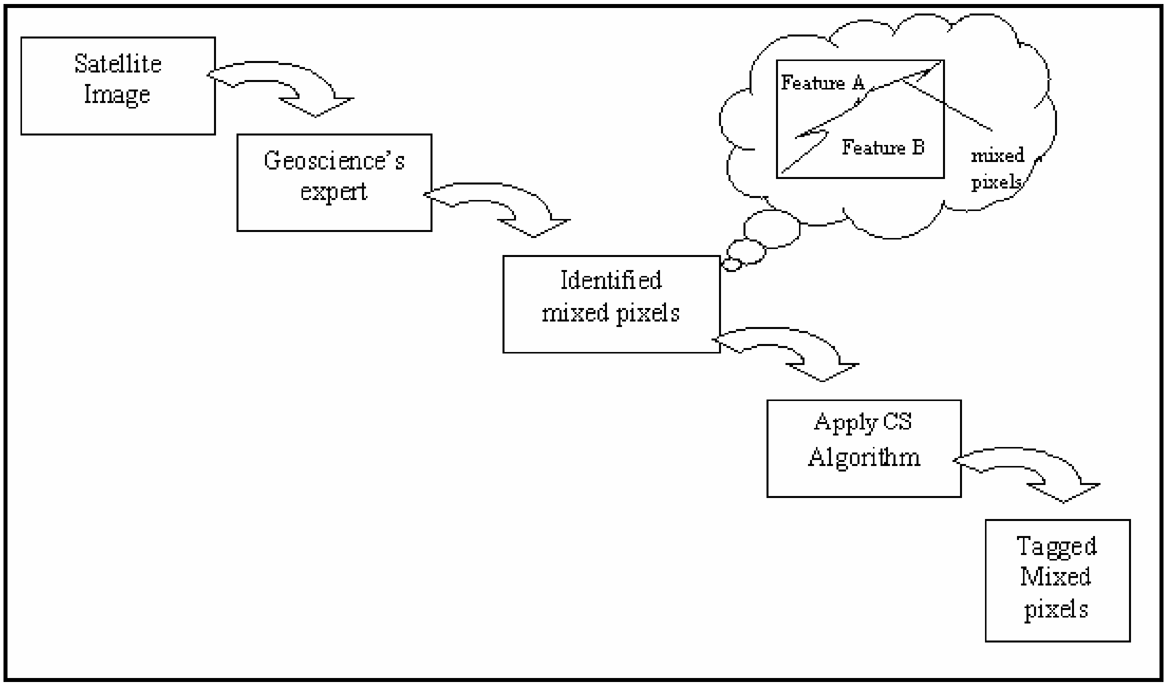


Fig 20: Block diagram for resolution of mixed pixels using CS

6.2. Proposed Algorithm

The basic terminologies used in this algorithm are:

- Cuckoo egg: This is the pixel to be classified.
- Case: Each pixel of an image is represented in the form case. It is a tuple which stores the spectral responses of that pixel for different frequency bands.
- Case base: It stores the training set provided by the experts. This is used for training the algorithm.
- Host nests: These are the cases stored in the case base.
- Quality solutions: The cases in the case base that match the current query.
- Best solution: The case in the case base for which we obtain highest ranking.

Input: Dataset of pure and mixed pixels of land features.

/*A case base of pure pixels is formed which acts as available host nests used for the classification of the mixed pixel*/

Output: Classified mixed pixels.

/*each of the mixed pixel is classified into one of the pure classes.*/

1. Read the set of pure pixels. /* These are the pixels that form the case base*/
2. Take the input query of mixed pixel to be classified.
3. Initialize k=sqrt (n). /* the number of nearest neighbour to be selected and its equal to sqrt(no. of entries in case base i.e. n) */
4. for i= 1 to n /*loop for all the cases in the case base*/
    a. Apply KNN to find similarity b/w the mixed pixel (i.e. the cuckoo egg or the query) and the current case in the case base (i.e. the available host nest in the case base).
    b. Store the similarity value in a temporary database and i++.

   End
5. Sort the similarity values in the database and store the top k results i.e. the quality solutions and discarding the worst nests.
6. for j= 1 to k
    a. Find the correlation coefficient b/w the mixed pixel and the current case in the database of top k quality solutions./*done for the purpose of ranking */
    b. Store the correlation value into a database and j++.
7. Sort the database of correlation value to get the best solution.

8. Based on the expert knowledge determine the class to which the best solution belong. Based on the similarity the mixed pixel will also belong to the same class.

The flowchart for the above algorithm is given below:

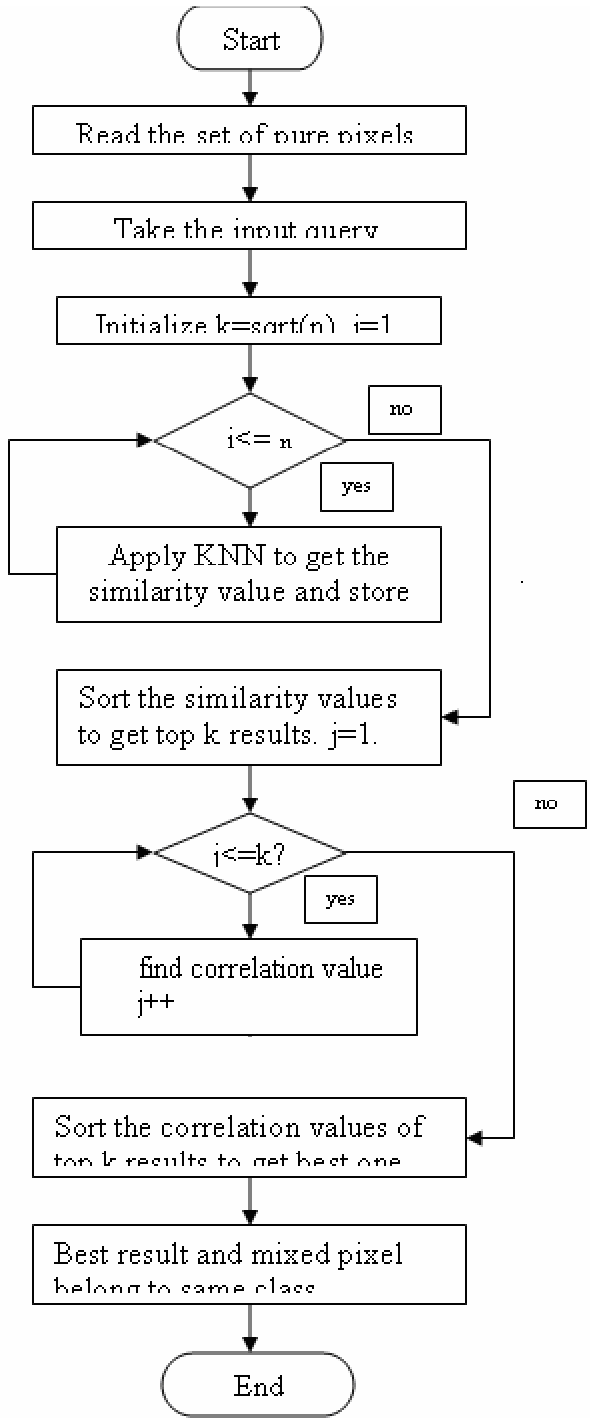


Fig 21: Flowchart for resolution of mixed pixels

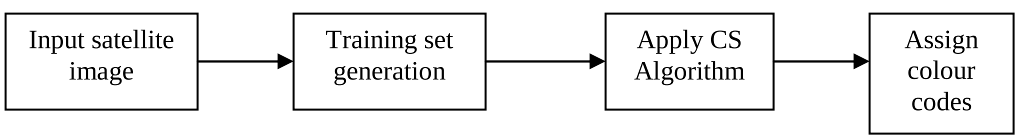
7. Experiment and Results

We have to differentiate between the regions/classes in ALWAR AREA and SAHARANPUR AREA images. In the final image these classes are depicted in different colors. The accuracy of the system has been measured on the basis of the Kappa coefficient.

We have 7 bands Satellite Image of The ALWAR REGION as shown in figure 22 (a) and 6 bands image of SAHARANPUR REGION as shown in figure 22 (b). The various steps of our system are applied on these images.

7.1. Implementation Details

The functional flow of the algorithm is shown below:



- The image is read using imread function provided in MATLAB.
- Training set is generated using the ERDAS IMAGINE software.
- The coding for the algorithm is done in MATLAB using the image processing toolbox.
- The resultant image is created by assigning colour codes to each pixel of the image and then displaying the image using imshow function of MATLAB.

7.2. Case Study 1: Image Classification for ALWAR and SAHARANPUR REGION

An image is nothing but collection of pixels. The image of Alwar taken by us for validating our algorithm has 2, 55,136 pixels and the image of Saharanpur has 4, 10, 881 pixels. Alwar region has mainly the following features: Water, Vegetation, Urban, Rocky, and Barren. Whereas Saharanpur region has following features: barren, dense vegetation, medium vegetation, sparse vegetation, urban, and water.

Our geosciences experts recognized all these features in Alwar and Saharanpur region. Thus our aim is to extract all these features from a satellite image. The data provided to us by our experts are collection of band images. Each band image represents spectral response for each pixel in that frequency band.

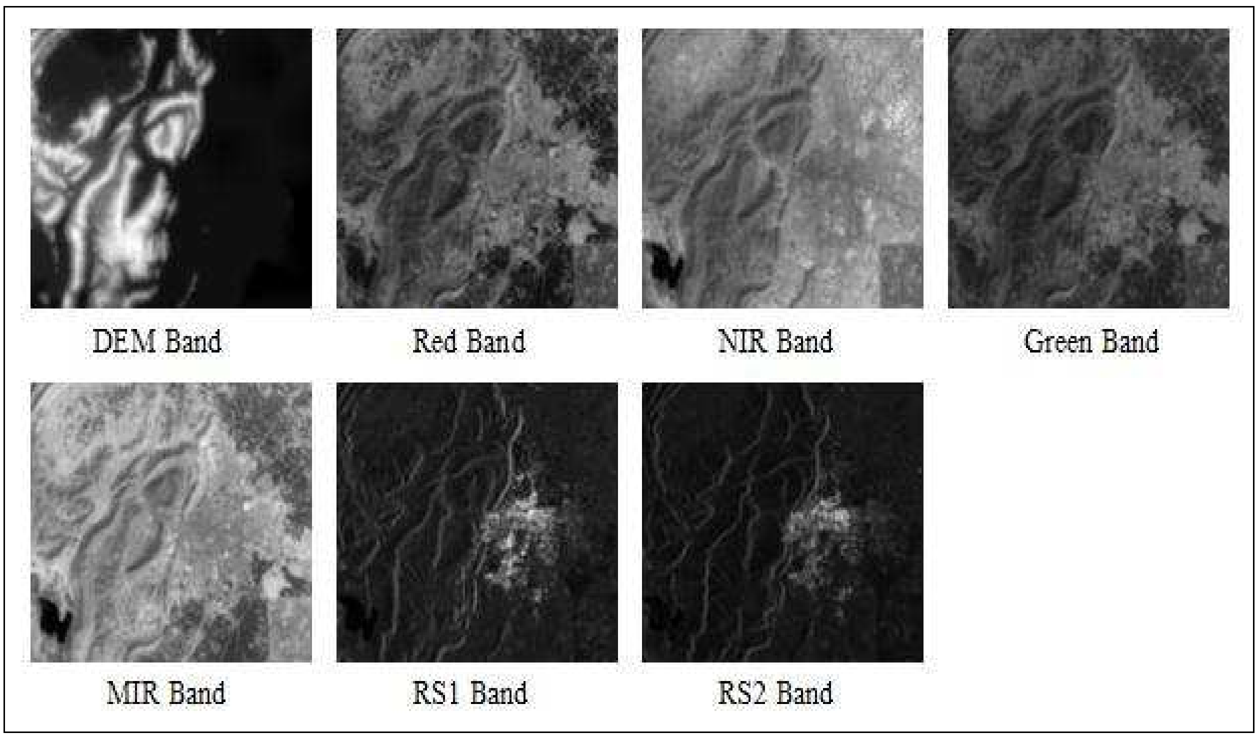


Fig 22(a): Seven band Satellite Image of ALWAR REGION

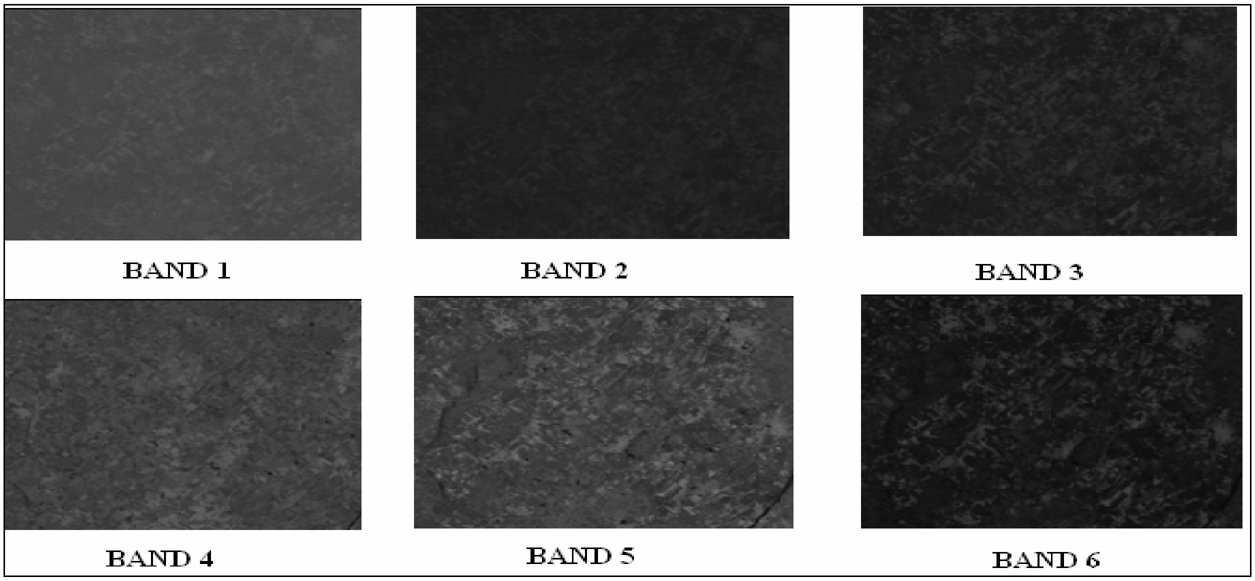


Fig 22(b): Six band Satellite Image of SAHARANPUR REGION

7.2.1. Training Set

Initially we have different bands satellite images of any region. The image comprises of several land covers like rocky, water, vegetation, open land, barren etc. In a satellite image the regions are not displayed appropriately. So, we need some system expert, to differentiate between these land covers. ERDAS software is used by our geoscientist expert. In order to classify the image by different techniques and check their accuracy assessment each classified image has been compared with the image generated by expert system. Screenshot of how the training set is generated by this software manually is shown in figure 23. The individual classes' pixel has been picked from an image and their proper class is then been assigned.

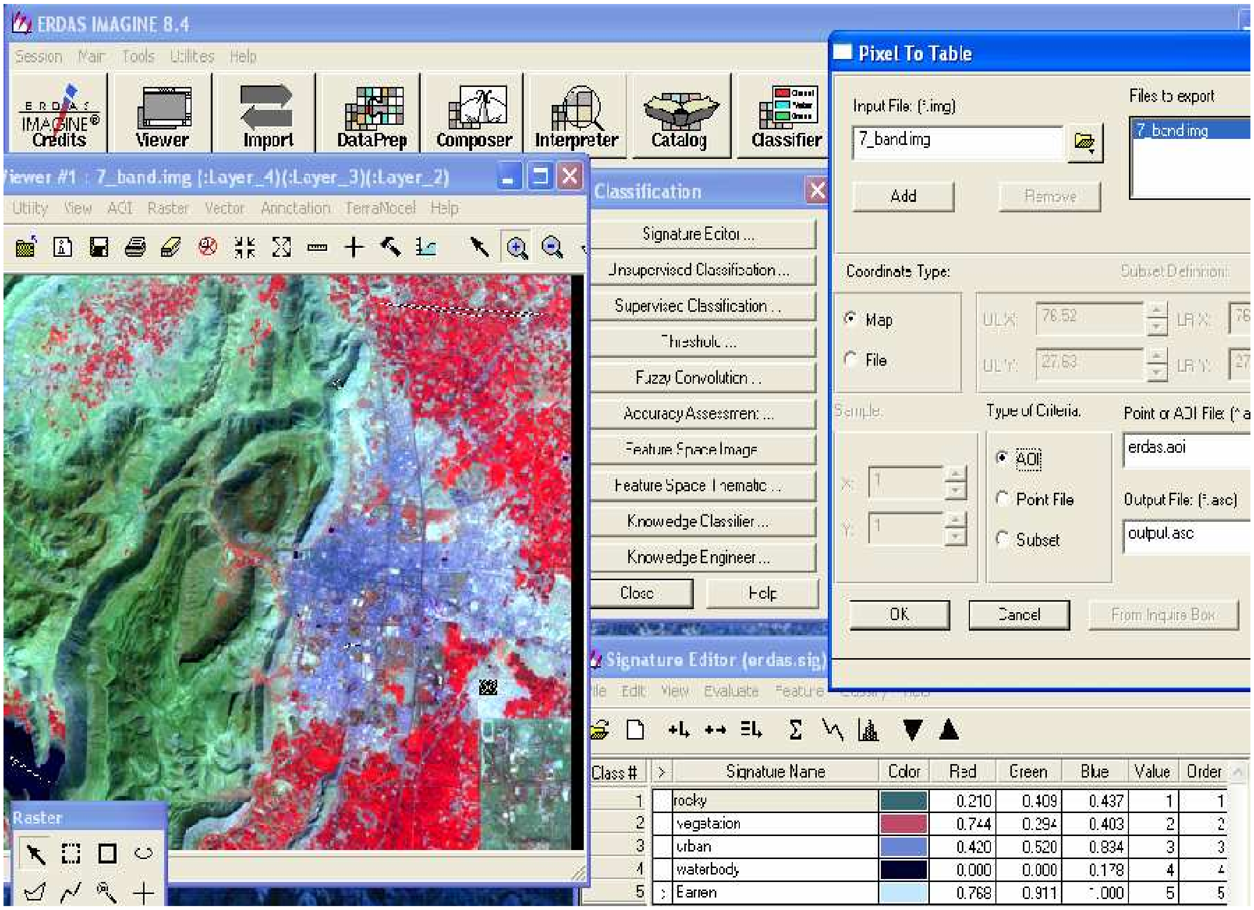


Fig 23: Snap shot for training set generation by ERDAS IMAGINE

In ERDAS first we have to open the 7 band image in View. Then we will select type of Classification i.e. Supervised Classification from classifier menu. We will open the tool box for Raster Image; it is shown in left bottom corner. Then we will select appropriate tool for the selection of the pixel/ pixels and merge all similar pixels into one class. Class creation has been done in Signature Editor where we merge the selected pixel of one class into one value and correspondingly one ASCII file has been created for it. In order to see the selected pixels we have to save them as an AOI layer as shown in figure 24.

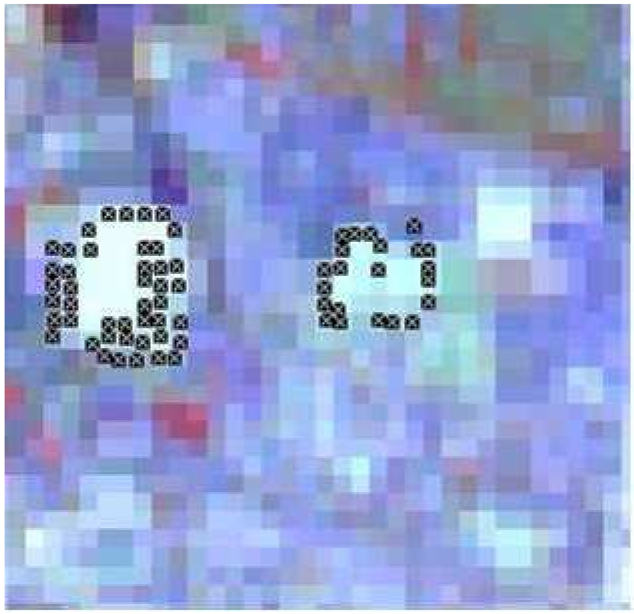


Fig 24: Showing the pixels selected for URBAN class

An example of training set generated for water, vegetation, urban, barren and rocky classes of Alwar area is shown as under:

Table 3: Training Set for ALWAR REGION IMAGE

RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
115	91	182	126	20	15	30	BARREN
111	90	173	131	17	34	15	BARREN
121	91	182	118	26	17	40	BARREN
125	98	188	128	25	21	27	BARREN
62	49	135	91	44	40	94	ROCKY
84	64	160	102	20	25	165	ROCKY
52	45	129	85	15	29	107	ROCKY
91	69	171	106	10	46	123	ROCKY
128	106	184	142	22	35	15	URBAN
128	120	155	118	18	36	15	URBAN
117	109	157	122	19	21	15	URBAN
123	109	166	137	10	13	15	URBAN
13	30	64	228	28	24	12	VEGETATION
11	28	60	245	14	23	13	VEGETATION
15	33	64	245	13	30	13	VEGETATION
15	33	82	255	15	20	11	VEGETATION
23	30	16	14	3	4	30	WATER
21	24	14	10	1	2	30	WATER
23	25	14	12	1	1	30	WATER
23	27	12	10	2	1	30	WATER

Here as we have taken the 7 band image i.e. Red band , Green Band, NIR (Near Infrared), MIR (Middle Infrared), DEM(Digital Elevation Model), RS1(Radar Sat 1) and RS2

(Radar sat 2). In the above the last column showed the decision of the expert system for the different band values of the pixel.

The training set for each class has been constructed by selecting some pixels of each class from the image. The sample of training set generated for Alwar region are depicted in Table 4, 5, 6, 7, 8.

Table 4: Training set for Water Class of Alwar

RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
17	35	1	3	15	2	54	WATER
21	35	7	7	1	2	30	WATER
19	32	14	7	4	2	30	WATER
17	35	7	1	7	2	30	WATER
19	35	3	3	3	3	30	WATER
21	38	1	1	5	0	30	WATER
19	33	1	1	6	1	54	WATER
21	32	10	5	5	0	30	WATER
17	35	1	1	1	3	30	WATER
19	37	1	5	3	3	30	WATER
21	32	3	1	8	4	48	WATER
17	33	3	3	10	2	44	WATER
17	37	5	3	8	3	30	WATER

Table 5: Training set for Vegetation Class of Alwar

RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
21	35	65	192	24	33	9	VEGETATION
23	37	73	192	18	33	14	VEGETATION
27	37	82	170	30	30	13	VEGETATION
21	35	69	252	20	22	11	VEGETATION
19	37	85	234	29	37	10	VEGETATION
17	32	69	246	25	37	10	VEGETATION
29	40	76	186	23	21	14	VEGETATION
15	25	60	237	13	26	12	VEGETATION
15	28	64	219	20	25	12	VEGETATION
13	30	64	228	28	24	12	VEGETATION
11	28	60	245	14	23	13	VEGETATION
15	33	64	245	13	30	13	VEGETATION
15	33	82	255	15	20	11	VEGETATION

Table 6: Training set for Urban Class of Alwar

RED	**GREEN**	**NIR**	**MIR**	**RS1**	**RS2**	**DEM**	**DECISION**
128	117	148	117	68	84	15	URBAN
123	104	155	118	48	124	15	URBAN
160	135	219	159	14	27	15	URBAN
146	127	190	149	22	11	15	URBAN

127	117	149	118	52	40	15	URBAN
158	130	201	160	11	15	15	URBAN
136	115	186	144	15	30	9	URBAN
134	119	164	135	61	60	15	URBAN
138	125	199	164	8	16	15	URBAN
134	117	155	118	85	118	15	URBAN
168	146	213	171	20	24	12	URBAN
93	83	122	115	27	31	15	URBAN
123	111	164	131	29	29	15	URBAN

Table 7: Training set for Rocky Class of Alwar

RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
54	38	118	69	17	24	90	ROCKY
62	43	140	84	36	35	214	ROCKY
72	53	159	93	20	25	237	ROCKY
54	37	128	80	14	18	165	ROCKY
44	32	106	60	74	32	86	ROCKY
58	41	122	76	16	22	115	ROCKY
56	40	138	78	29	41	240	ROCKY
37	27	98	53	67	19	89	ROCKY
64	46	138	85	13	33	101	ROCKY
58	41	140	85	21	37	234	ROCKY
66	49	146	87	21	41	238	ROCKY
41	33	115	65	43	25	101	ROCKY
54	43	133	84	46	20	102	ROCKY

Table 8: Training set for Barren Class of Alwar

RED	GREEN	NIR	MIR	RS1	RS2	DEM	DECISION
119	91	186	129	21	23	29	BARREN
134	111	195	151	28	25	30	BARREN
134	111	184	133	17	44	42	BARREN
132	99	206	128	31	39	30	BARREN
127	98	197	137	22	23	29	BARREN
140	112	190	155	16	17	30	BARREN
152	125	221	157	14	22	30	BARREN
140	109	190	149	14	24	30	BARREN
150	130	212	173	20	53	30	BARREN
154	130	210	166	22	36	30	BARREN
144	115	193	149	31	48	28	BARREN
134	112	186	157	34	58	28	BARREN
170	143	208	151	22	24	32	BARREN

Similarly, the training set for different classes of Saharanpur region can also be created. Table 9, 10, 11, 12, 13, 14 portrays the training set for Saharanpur region. The first two columns show the pixel coordinates followed by the 6 band reflectance values.

Table 9: Training set for Barren Class of Saharanpur

176505	3319844	117	66	85	79	133	85
176535	3319844	120	67	84	80	137	86
176835	3319784	111	65	83	76	130	83
176865	3319784	111	63	81	75	138	85
176835	3319694	107	61	76	75	128	79
176805	3319664	109	61	77	73	125	79
176295	3319514	112	63	79	76	129	84
176925	3316544	81	42	52	76	102	49
176865	3315584	84	41	45	75	70	28
176655	3315374	83	44	48	82	89	39

Table 10: Training set for Dense vegetation Class of Saharanpur

168315	3327044	69	29	28	56	49	21
167835	3326654	69	29	27	54	59	21
167865	3326654	68	28	27	54	58	21
167835	3326624	68	28	28	57	58	20
167865	3326624	67	29	27	56	56	19
169065	3326594	66	28	25	59	45	16
169065	3326564	67	27	25	61	51	17
166575	3325064	67	27	25	55	47	17
166605	3325064	69	27	25	52	47	18
166575	3325034	66	27	24	55	45	16

Table 11: Training set for Medium vegetation Class of Saharanpur

174465	3321314	69	30	26	94	74	24
174435	3321284	69	29	25	93	66	20
168525	3321104	70	29	27	75	59	21
164325	3321074	71	30	28	80	66	24
168495	3321044	70	30	27	75	54	18
168465	3321014	70	29	27	65	55	22
168465	3320984	69	29	27	73	54	19
168465	3320954	69	29	27	73	56	20
168075	3320594	69	28	25	80	62	21
168015	3320534	68	28	25	77	60	19

Table 12: Training set for Sparse vegetation Class of Saharanpur

165585	3330014	75	38	36	92	70	26
165405	3329984	74	37	35	91	69	25
165135	3329954	76	38	37	83	71	26
165585	3329954	76	38	37	90	69	25
165525	3329894	76	40	36	98	69	25
165555	3329894	76	40	36	98	69	25
166275	3329894	74	37	33	93	65	23
166305	3329894	73	36	33	88	63	23
165165	3329864	74	36	33	91	70	26
165195	3329864	75	37	34	93	70	25

Table 13: Training set for Urban Class of Saharanpur

167685	3324824	76	34	39	56	83	43
167715	3324824	76	34	38	56	82	42
167745	3324824	74	32	34	57	72	36
167655	3324794	75	34	36	59	74	33
167685	3324794	74	35	37	56	75	36
167715	3324794	76	34	38	55	72	36
167745	3324794	76	35	36	60	78	38
168375	3324614	77	33	38	40	65	41
168405	3324614	75	32	36	48	65	36
168435	3324614	73	32	33	58	62	28

Table 14: Training set for Water Class of Saharanpur

167955	3329684	74	33	29	26	22	10
167925	3329564	74	32	29	34	35	16
167895	3329534	74	33	29	28	38	10
167835	3329414	76	34	31	36	32	18
167805	3329384	74	34	20	29	24	11
167655	3329204	74	34	30	28	24	12
167085	3328604	72	33	30	35	35	12
166965	3328514	75	34	30	32	28	11
166995	3328514	75	33	29	23	19	9
166815	3328394	75	34	29	24	19	8

7.2.2. Result and Discussion

We have applied our classification technique on two different dataset and it has shown good results with both dataset, confirming its applicability for different features. First we have shown our result on Alwar region and then on Saharanpur region. Our objective is to use the proposed natural computation algorithm as an efficient land over classifier for satellite image.

Figure 25(a) shows the original Alwar image and figure 25(b) shows the false color image of Saharanpur region. Figure 26(a) and (b) shows the classified image obtained by applying our algorithm to Alwar and Saharanpur region respectively.

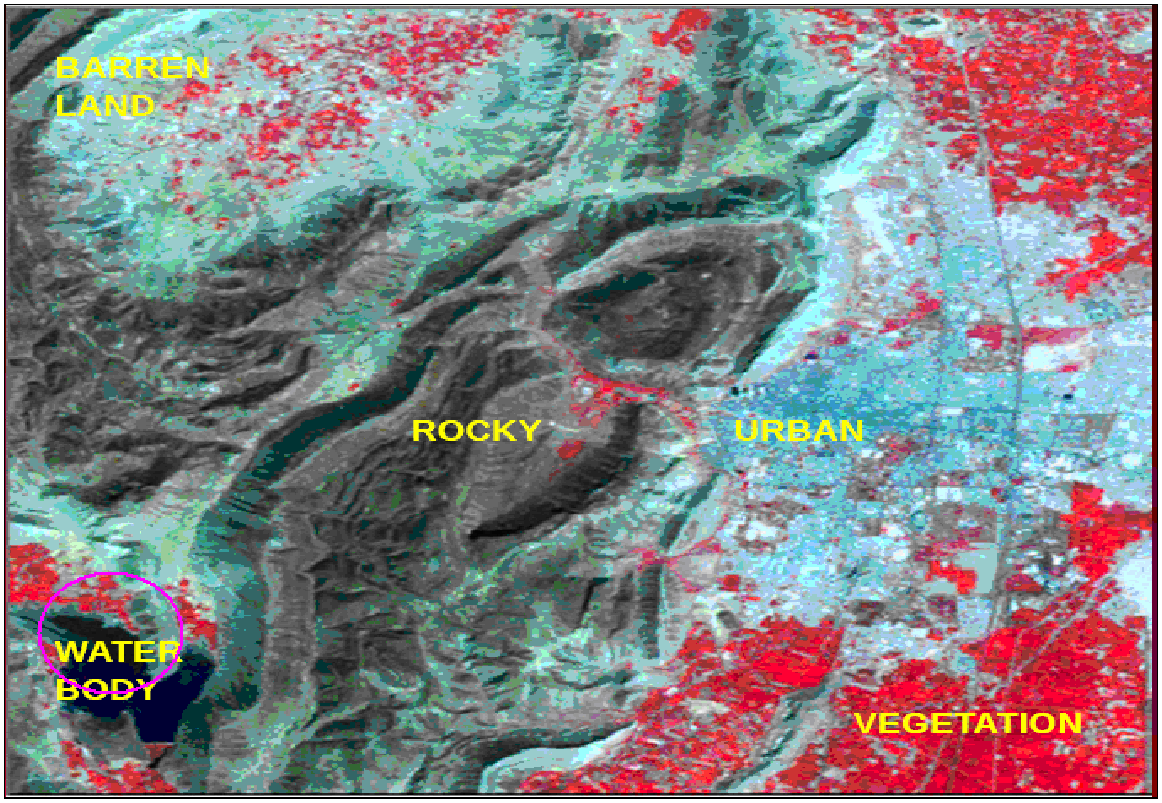


Fig 25 (a): Original image of Alwar Region

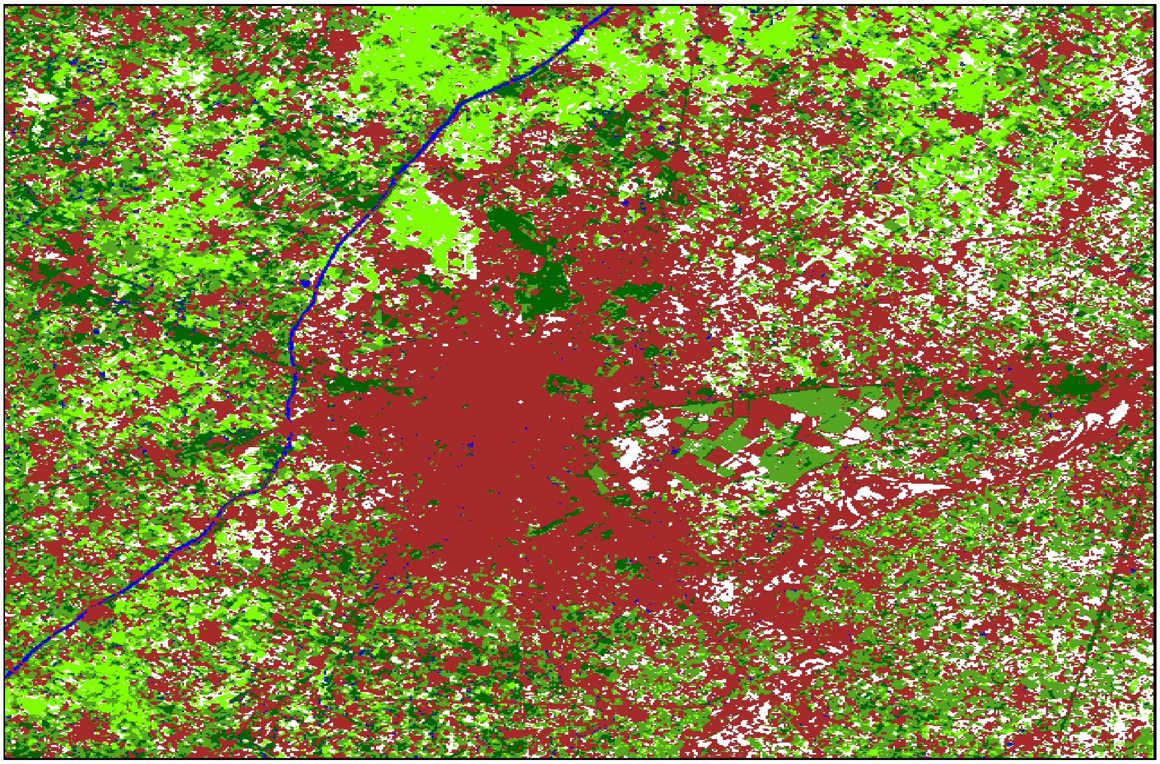


Fig 25 (b): Original image of Saharanpur Region

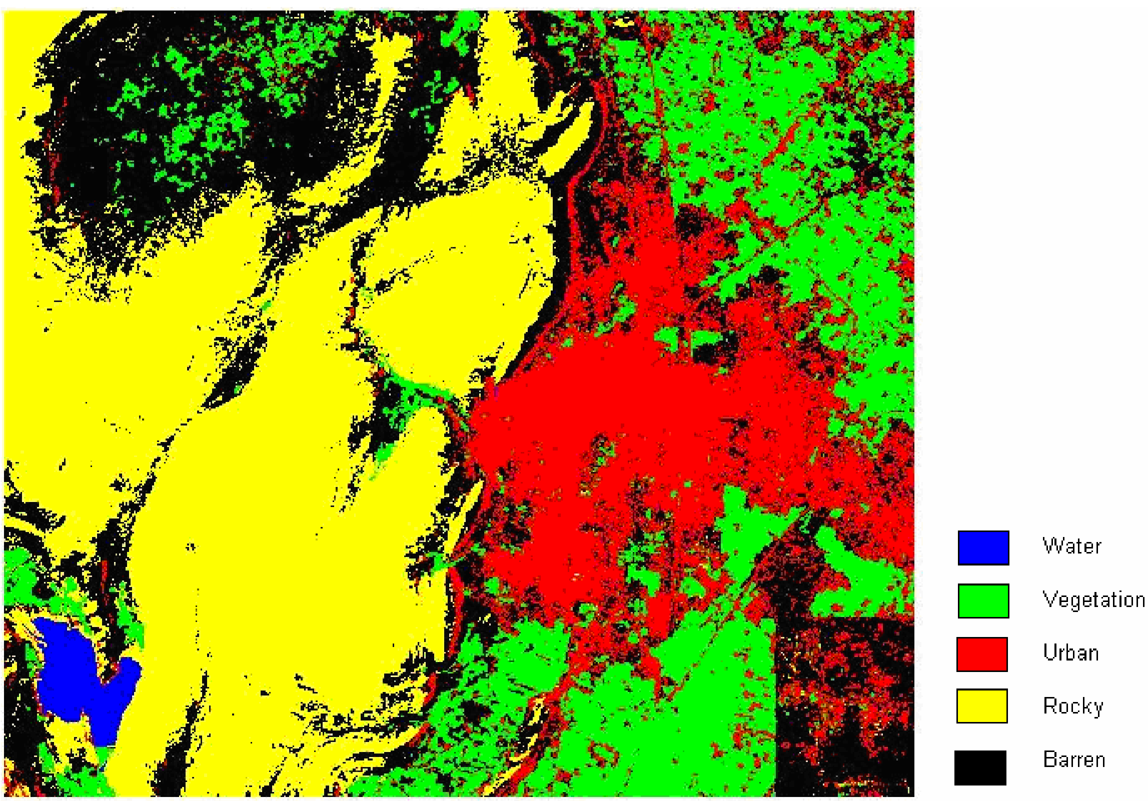


Fig 26(a): Classified image of Alwar Region

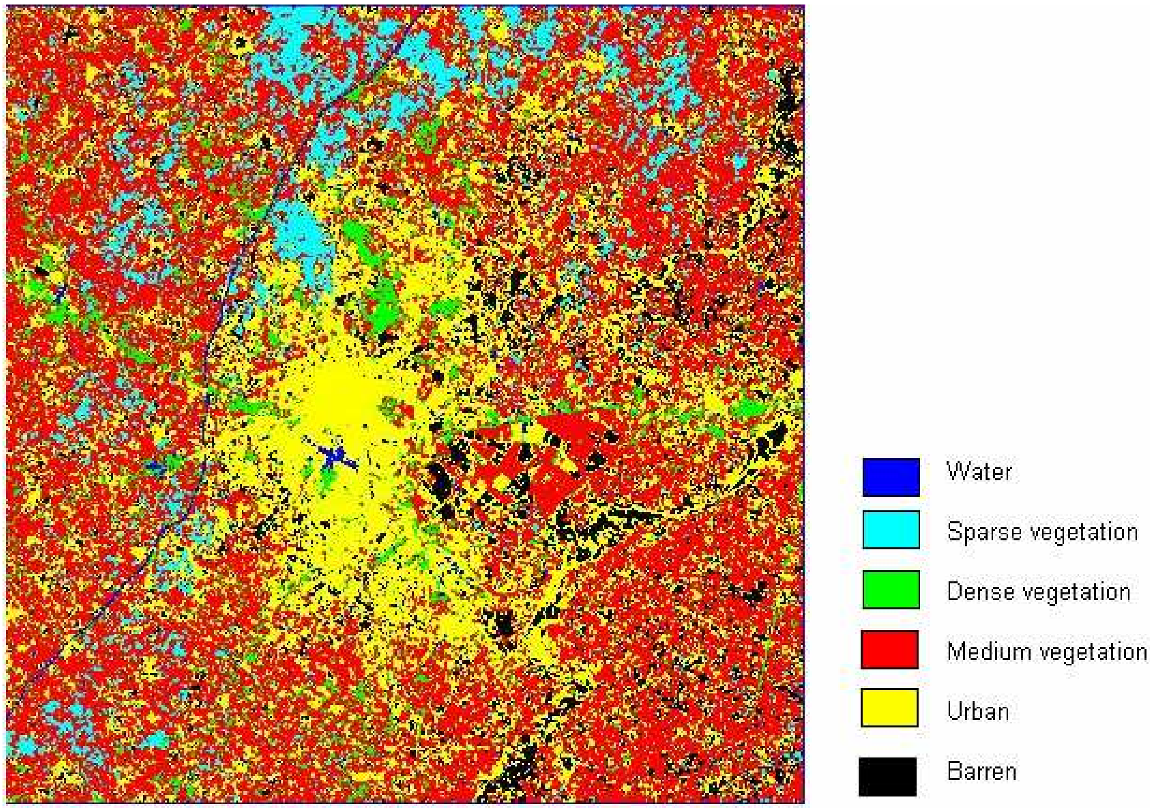


Fig 26(b): Classified image of Saharanpur Region

Original image depicted the image which is obtained after classification by our geoscientist and classified image is obtained after applying our proposed Classifier. The image obtained after classification has different colour. In figure 26(a) yellow, black, blue, green, red colour represents rocky, barren, water, vegetation and urban region respectively. Comparing the original image of Alwar and our classified output we can clearly see the good correlation between the two. All the features namely water, vegetation, urban, rocky and barren are well classified.

In figure 26(b) black, green, red, light blue, yellow, and royal blue represents barren, dense vegetation, medium vegetation, sparse vegetation, urban and water respectively. From the above two figures we observe high correlation and notice that all the 6 features have been very well extracted.

7.3. Case Study 2: Resolution of mixed pixels for ALWAR REGION

Problem Statement: Image classification is one of the most important areas of digital image analysis. There is a fundamental assumption that each pixel in an image represents properties of a single land cover feature (or class). But, this is not always the case, because of the existence of mixed pixels that show the properties of more than one land cover features [24]. At present several soft computing techniques are being used for the classification of an image, but their performance is not satisfying for the classification of mixed pixel. The water-vegetation mixed pixels of Alwar region is shown in figure 27.

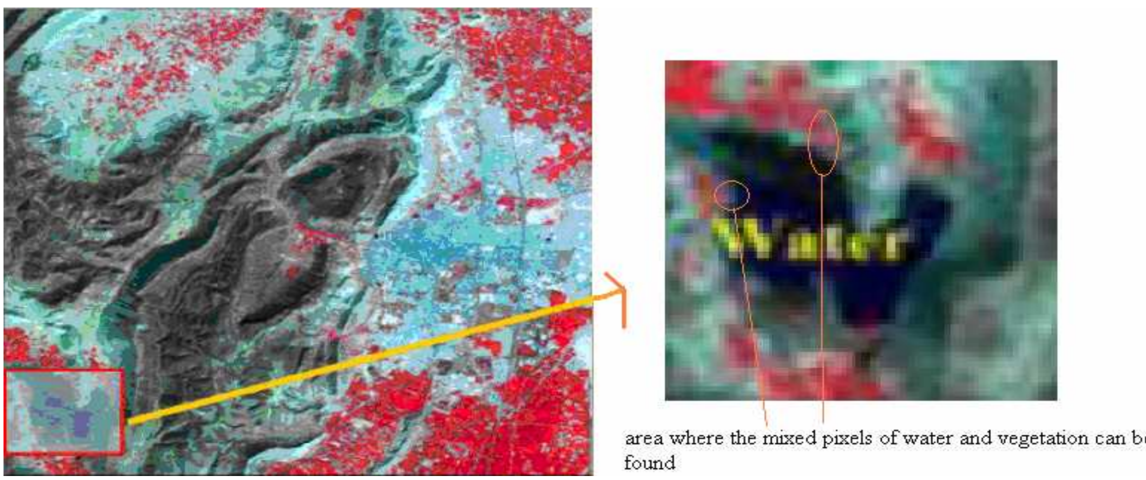


Fig 27: Image showing mixed pixels

This image has a portion near the water body region where at the edge of water body both probability of water and vegetation is mixed and experts have conflicts to decide about the assignment of this portion. Basically this conflict arises when spectral signatures become similar for different features as shown in figure 28.

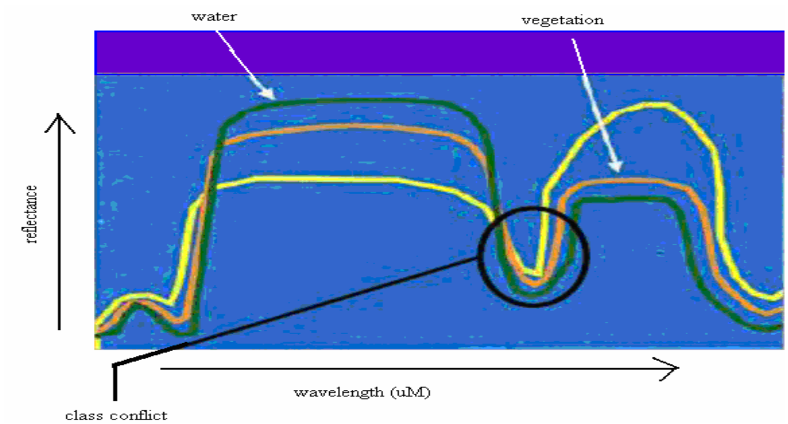


Fig 28: Mixture of spectral signature

7.3.1. Training Set used for resolution

Here, we are working on the water-vegetation mixed pixels. Table 15 and 16 shows the sample of the dataset that was being used for tagging of the mixed pixel by an appropriate land cover class.

Table 15: Water pixels for mixed pixel resolution

RED	GREEN	NIR	MIR	RS1	RS2	DEM
21	27	12	14	1	4	30
23	25	14	12	1	1	30
23	28	14	10	2	1	30
21	27	12	10	3	2	30
21	28	14	10	3	3	30
23	27	14	12	1	3	30
25	25	12	10	7	4	30
21	25	10	10	4	2	30
23	24	16	12	2	1	30
21	28	12	10	2	1	30
21	27	12	12	2	1	30
25	24	10	12	2	1	30
25	28	10	10	5	1	30
21	25	12	12	3	0	30

Table 16: Vegetation pixels for mixed pixel resolution

RED	GREEN	NIR	MIR	RS1	RS2	DEM
37	51	98	182	33	33	15
35	49	95	179	20	30	15
31	40	91	188	24	22	15
37	51	98	182	30	35	15
35	49	95	179	19	33	15
31	43	87	199	23	26	15
33	43	89	184	24	23	15
29	43	89	186	26	26	15
33	45	91	195	16	24	15
29	41	93	182	23	19	15
42	48	102	181	19	28	15
33	43	91	181	24	54	15
33	45	96	181	26	29	15
52	51	115	170	26	39	15

7.3.2. Result and Discussion

We have worked on the multi-spectral image of Alwar, Rajasthan, India. The following mixed pixel dataset were being made available by our experts: urban-barren, barren-rocky, urban-rocky, vegetation-barren, vegetation-rocky, water-rocky, and water-vegetation. Here, we have worked on the water-vegetation dataset.

The case base consists of a collection of 7 band values of pure pixels of water and vegetation. The 7 band values of the mixed pixel to be classified are taken as an input. Then using KNN we find those pixels from the case base that matches the given input query. From the obtained quality solutions we determine the best solution by performing ranking. For this purpose we have used Pearson product-moment correlation coefficient. Then we find the class to which our best solution belongs. The mixed pixel will also belong to that same class. This same approach can be applied to other mixed pixel datasets as well. Judging from the experimental results, the mixed pixel is resolved and hence the whole image is being classified. The sample input in shown in table 17 and the output for this data is shown in table 18.

Table 17: Set of mixed pixels as input

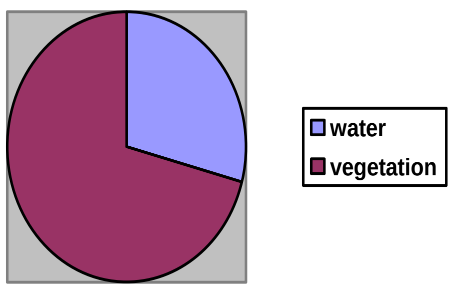
X	Y	RED	GREEN	NIR	MIR	RS1	RS2	DEM
76.54119	27.52842	62	59	104	89	2	10	30
76.54144	27.52842	35	41	42	25	1	4	30
76.54169	27.52842	23	37	18	7	1	2	30
76.52974	27.52817	41	45	93	138	21	0	30
76.52999	27.52817	46	49	69	56	17	3	30
76.53024	27.52817	25	35	20	10	16	2	30
76.54094	27.52817	80	64	131	91	1	58	30
76.54119	27.52817	70	62	115	100	2	6	30
76.54144	27.52817	41	46	54	43	1	2	30
76.54169	27.52817	25	38	18	10	1	2	30
76.52974	27.52792	54	54	102	138	16	0	30
76.52999	27.52792	42	45	51	60	20	3	30
76.54119	27.52792	72	66	115	115	0	5	30
76.54144	27.52792	37	41	42	34	0	3	30
76.54119	27.52767	62	59	96	98	0	2	30
76.54144	27.52767	31	38	25	16	0	2	30

A total of 203 mixed pixels were being classified by this strategy. The figure 29 shows that the result obtained by our approach are satisfying taking into consideration hybridization of ACO and BBO algorithm for resolution of mixed pixels [16] approach as standard for comparison.

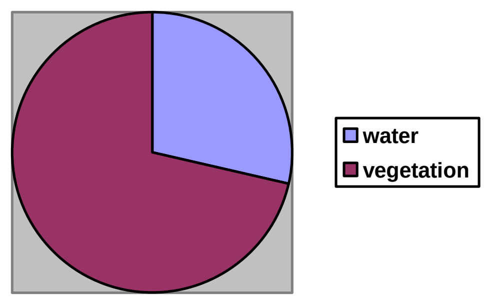
Using our approach we found 59 pixels of water and 144 pixels of vegetation. Using the hybridization of ACO and BBO 58 pixels of water and 145 pixels of vegetation were obtained. This is being shown in figure 29.

Table 18: Output obtained

X	Y	Decision
76.54119	27.52842	vegetation
76.54144	27.52842	water
76.54169	27.52842	water
76.52974	27.52817	vegetation
76.52999	27.52817	water
76.53024	27.52817	water
76.54094	27.52817	vegetation
76.54119	27.52817	vegetation
76.54144	27.52817	water
76.54169	27.52817	water
76.52974	27.52792	vegetation
76.52999	27.52792	water
76.54119	27.52792	vegetation
76.54144	27.52792	water
76.54119	27.52767	vegetation
76.54144	27.52767	water



Cuckoo Search



ACO & BBO

Fig 29: Result Comparison

8. Accuracy Assessment of Image Classification using CS

Another area that is continuing to receive increased attention by remote sensing specialists is that of classification accuracy assessment. Historically, the ability to produce digital land cover classifications far exceeded the ability to meaningfully quantify their accuracy. In fact, this problem sometimes precluded the application of automated land cover classification techniques even when their cost compared favorably with more traditional means of data collection. The lesson to be learned here is embodied in the expression “A classification is not complete until its accuracy is assessed." Accuracy assessment is used for comparing the classification to geographical data that are assumed to be true (ground truth), in order to determine the accuracy of the classification process. Practically it is not possible to test each and every pixel of a classified image. So, instead a set of reference pixels is used. Reference pixels are points on the classified image for which actual features are (or will be) known. The selection of these reference pixels is done randomly. Here the main aim is to quantitatively determine how effectively pixels were grouped into the correct feature classes in the area under investigation.

8.1. Error Matrix

One of the most common means of expressing classification accuracy preparation of a classification *error matrix* (sometimes referred to as a confusion matrix or a contingency table). Error matrix [1] compares, on a category by category basis, the relationship between known reference data (ground truth) and the corresponding results of an automated classification [12]. Such matrices are square, with the number of rows and columns equal to the number of categories whose classification accuracy is being assessed.

This matrix stems from classifying the sampled training set pixels and listing the known cover types used for training (columns) versus the pixels actually classified into each land cover category by the classifier (rows). Several characteristics about classification performance are expressed by an error matrix. For example, one can study the various classification errors of omission (exclusion) and commission (inclusion).

Once accuracy data are collected (either in the form of pixels, cluster of pixels, or polygons) and summarized in an error matrix, they are normally subject to detailed interpretation and further statistical analysis.

For the validation process of Alwar region we have taken into consideration following number of pixels:

- 68 water pixels.
- 109 vegetation pixels.
- 139 urban pixels.
- 96 rocky pixels.
- 63 barren pixels.

Following number of pixels are taken for validation of Saharanpur region classification:

- 15 barren pixels.
- 24 dense vegetation pixels.
- 35 medium vegetation pixels.

- 64 sparse vegetation pixels.
- 70 urban pixels.
- 13 water pixels.

Using the training set the classification is performed and based on the result error matrix is calculated which is shown in table 19(a) and (b). The diagonal elements show number of properly classified pixels while non diagonal elements represent number of wrongly classified pixels.

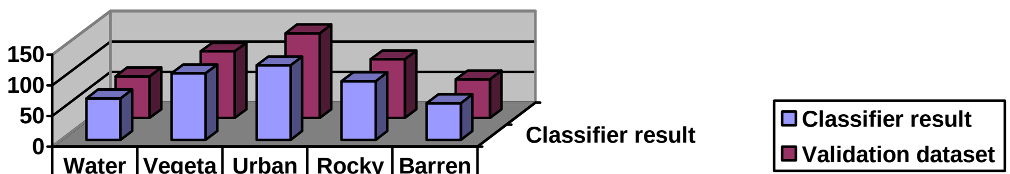
Table 19(a): Accuracy Table for Cuckoo Search based Classifier for ALWAR Region

Type	Water	Vegetation	Urban	Rocky	Barren	Total
Water	68	0	0	0	0	68
Vegetation	0	109	1	0	0	110
Urban	0	0	122	0	3	125
Rocky	0	0	0	96	0	96
Barren	0	0	16	0	60	76
Total	68	109	139	96	63	475

Table 19(b): Accuracy Table for Cuckoo Search based Classifier for SAHARANPUR Region

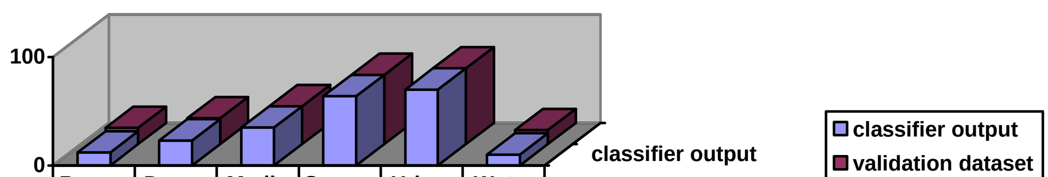
Type	Barren	Dense vegetation	Medium vegetation	Sparse vegetation	Urban	Water	Total
Barren	12	0	0	0	0	0	12
Dense vegetation	0	23	0	0	0	0	23
Medium vegetation	0	0	35	0	0	3	38
Sparse vegetation	0	1	0	64	0	0	64
Urban	0	0	0	0	70	0	70
Water	3	0	0	0	0	10	13
Total	15	24	35	64	70	13	221

Table 19(a) and (b) shows how our classifier classifies each feature perfectly whereas figure 30(a) and (b) shows the graphical view for Alwar and Saharanpur dataset respectively. In this we have shown all features along x-axis and number of pixels that are being classified along y-axis. The blue bar represents number of pixels classified correctly by our classifier whereas the purple bar depicts the sample validation data set that had been given to us by our geology expert.



	Water	Vegeta	Urban	Rocky	Barren
Classifier result	68	109	122	96	60
Validation dataset	68	109	139	96	63

Fig 30(a): Comparison study for Alwar region



	Barren	Dense	Mediu	Sparse	Urban	Water
classifier output	12	23	35	64	70	10
validation dataset	15	24	35	64	70	13

Fig 30(b): Comparison study for Saharanpur region

8.2. Producer Accuracy

Several other descriptive measures can be obtained from error matrix. The accuracy of individual category can be calculated by dividing the number of correctly classified pixels in each category by either the total number of pixels in corresponding row or column. Producer‘s accuracies(as shown in table 20 (a) and (b)) result from dividing the number of correctly classified pixels in each category (on the major diagonal) by the number of training set pixels used for that category(the column total). This figure indicates how well the training pixels of a given cover type are classified.

Table 20(a): Producer's Accuracy for Alwar area

Feature	Accuracy Calculation	Producer's Accuracy
Water	68/68	100%
Vegetation	109/109	100%
Urban	122/139	87.76%
Rocky	96/96	100%
Barren	60/63	95.23%

Table 20(b): Producer's Accuracy for Saharanpur area

Feature	Accuracy Calculation	Producer's Accuracy
Barren	12/15	80%
Dense vegetation	23/24	95.83%
Medium vegetation	35/35	100%
Sparse vegetation	64/64	100%
Urban	70/70	100%
Water	10/13	76.92%

Plot of terrain features of Alwar region extracted by our algorithm with respect to actual value of each feature is shown in figure 31(a) and for Saharanpur region is shown in figure 31(b).

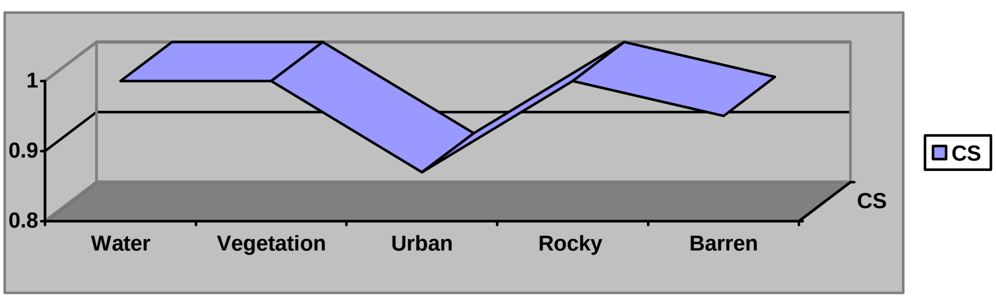


Fig 31(a): Terrain Features Extracted by CS for Alwar Region

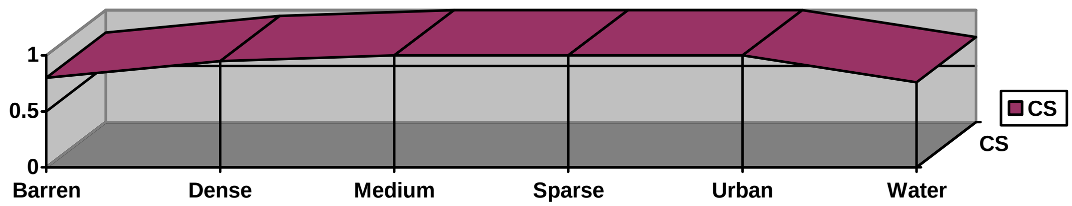


Fig 31(b): Terrain Features Extracted by CS for Saharanpur Region

8.3. User's Accuracy

User‘s Accuracies (as shown in following table) is computed by dividing the number of correctly classified pixels in each category by the total number of pixels that were classified in that category (the row total).This figure is a measure of commission error and indicates the probability that a pixel classified into a given category actually represents that category on the ground.

Table 21(a): User's Accuracy for Alwar area

Feature	Accuracy Calculation	User's Accuracy
Water	68/68	100%
Vegetation	109/110	99%
Urban	122/125	97.6%
Rocky	96/96	100%
Barren	60/76	78.95%

Table 21(b): User's Accuracy for Saharanpur area

Feature	Accuracy Calculation	User's Accuracy
Barren	12/12	100%
Dense vegetation	23/23	100%
Medium vegetation	35/38	92.10%
Sparse vegetation	64/64	100%
Urban	70/70	100%
Water	10/13	76.92%

It shows that commission error is high in case of barren region. Thus it can be further considered for our future work.

8.4. KHAT statistics

A further point to be made about interpreting classification accuracies is the fact that even a completely random assignment of pixels to classes will produce percentage correct values in the error matrix. In fact, such a random assignment could result in a surprisingly good apparent classification result. The k ("KHAT") statistic is a measure of the difference between the actual agreement between reference data and an automated classifier and the chance agreement between the reference data and a random classifier. Conceptually, k can be defined as

$$k = \frac{\text{observed accuracy - chance agreement}}{\text{1 - Chance agreement}}$$

This statistic serves as an indicator of the extent to which the percentage correct values of an error matrix are due to "true" agreement versus "chance" agreement. As true agreement (observed) approaches 1 and chance agreement approaches 0, k approaches 1. This is the ideal case. In reality, k usually ranges between 0 and 1. For example, a k value of 0.87 can be thought of as an indication that an observed classification is 87 percent better than one resulting from chance. A k with the value of 0 suggests that a given classification is no better then a random assignment of pixels. In cases where chance agreement is large enough, k can take on negative values-an indication of very classification performance. (Because the possible range of negative values depends on specific matrix, the magnitude of negative values should not be interpreted as an indication of relative classification performance). The principle advantage of computing KHAT coefficient is the ability to use this value as a basis for determining the statistical significance of any matrix or the differences among matrices. The KHAT coefficient incorporates the non-diagonal elements of error matrix (and hence error of omission and commission) as a product of the row and column marginal.

The Ꝃ -coefficients of the classification by using Cuckoo Search is given below. Ꝃ-coefficient of the classification results from MLC, MDC, Fuzzy, Rough Set, Membrane Computing and Semantic Classifier is given in table 22. The KHAT statistic is computed as

$$k = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} \cdot x_{+i})}$$

Where:

r = number of rows in the error matrix

x_{ii} = the number of observations in row i and column i (on the major diagonal)

x_{i+} = total of observations in row i (shown as marginal total to right of the matrix)

x_{+i} = total of observations in column i (shown as marginal total at bottom of the matrix)

N = total number of observations included in matrix

For Alwar region

To illustrate the computation of KHAT for the error matrix in table 19(a),

N= 475

$\sum x_{ii}$ =68+109+122+96+60= 455

$N(\sum x_{ii})$ = 475*455= 216125

$\sum(x_{i+} * x_{+i})$ =(68*68)+(109*110)+(125*139)+(96*96)+(76*63) = 47993

N^2= 475*475 = 225625

ˆk(KAPPA) =(216125 – 47993) / (225625-47993) = 0.9174

Similarly we can calculate for table 19(b) i.e. for Saharanpur region

N= 221

$\sum x_{ii}$ =12+23+35+64+70+10=214

$N(\sum x_{ii})$ = 221*214=47294

$\sum(x_{i+} * x_{+i})$ =(12*15)+(23*24)+(38*35)+(64*64)+(70*70)+(13*13) = 11227

N^2= 221*221 = 48841

ˆk(KAPPA) =(47294-11227) / (48841-11227) = 0.9588

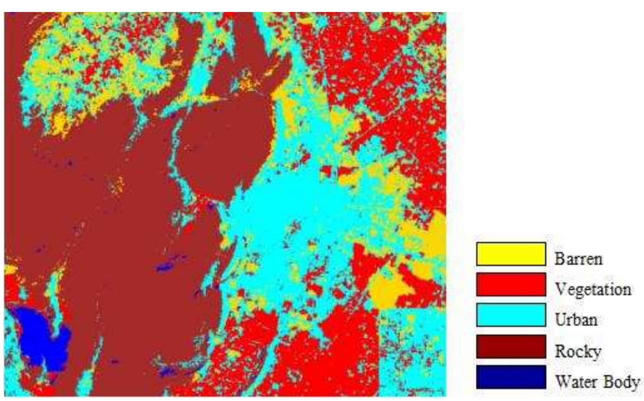
These values are a substantial agreement between the ideal classifier and the proposed algorithm. The Kappa (K) coefficient of the Alwar image is 0.9465 which indicates that an observed classification is 94.65% better than one resulting from chance. The Kappa (K) coefficient of the Saharanpur image is 0.9588 which indicates that an observed classification is 95.88% better than one resulting from chance.

8.5. Comparison with existing classifiers

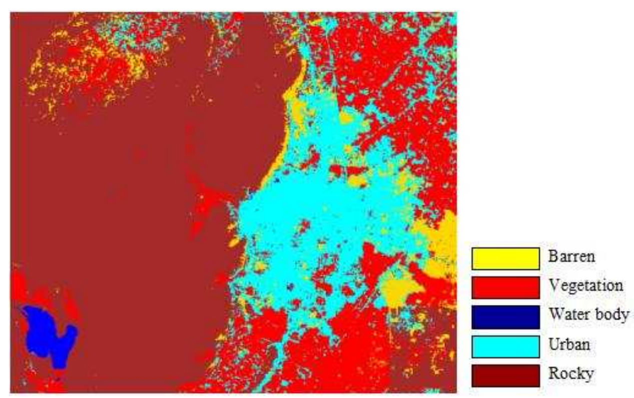
In this section we are going to perform the comparative analysis of the result obtained by our algorithm and the result for some of the existing image classification algorithms. The analysis is divided into two parts: first part performs the comparison for homogeneous regions and second part for heterogeneous regions.

8.5.1. Homogeneous region Comparison for Alwar image

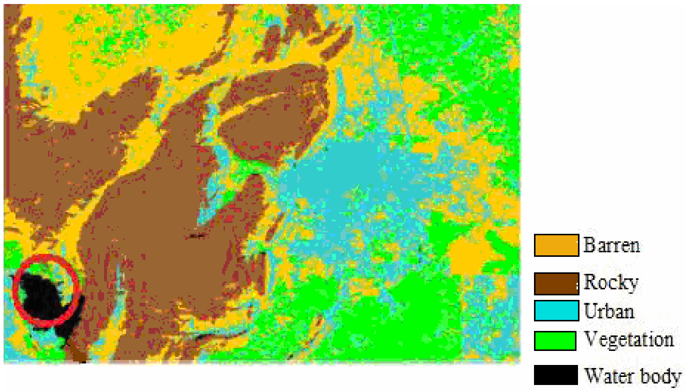
One of the important characteristic features of Cuckoo Search is that it works on single unit of object rather than group of objects. Consequently the clarity observed using Cuckoo Search is much clearer than other classifiers because most of them use certain clustering mechanism. Like BBO uses fuzzy-c-means cluster to group all the similar pixels together. The Kappa coefficient of the Alwar image when Cuckoo search was applied is 0.9465. Comparing this value with the Kappa coefficient of MDC, MLC, BBO and Membrane Computing (MC), we deduced that Cuckoo search has shown best results for image classification since it has obtained the highest Kappa coefficient as shown in table 22. The comparison of figure 32 with figure 25(a) shows that our algorithm has captured almost all the terrain features and showed high degree of efficiency for almost all the regions (water, vegetation, urban, rocky, and barren) with a Kappa coefficient of 0.9465.



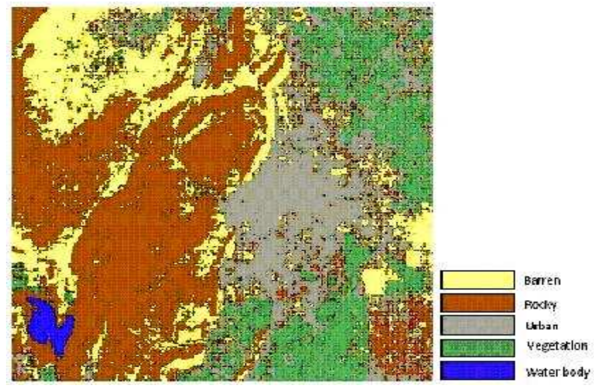
Minimum Distance Classifier



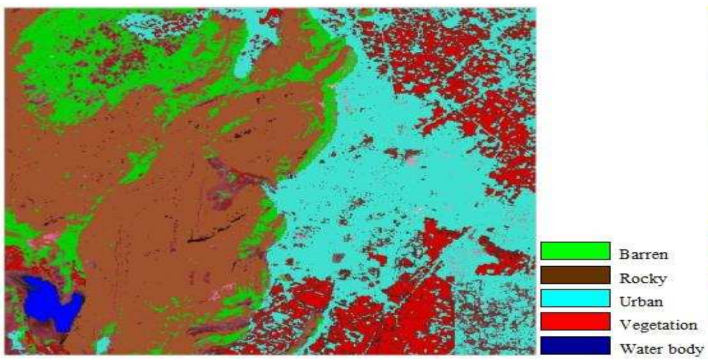
Maximum likelihood Classifier



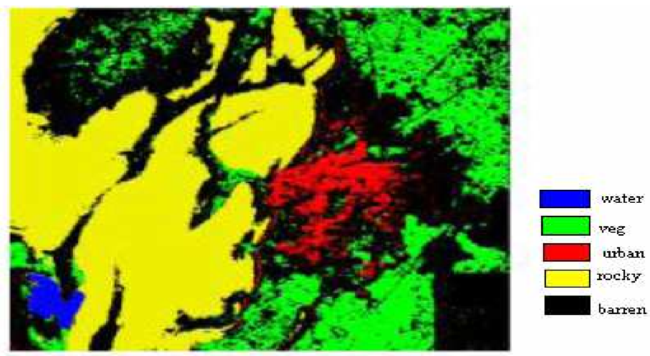
Fuzzy classifier



Rough Fuzzy tie up



Semantic Classifier



Membrane computing

Fig 32: Existing Classifiers output

Table 22: Kappa Coefficient comparison

Minimum Distance Classifier(MDC)	Maximum Likelihood Classifier(MLC)	Fuzzy set	Rough set classifier	Semantic Web Based Classifier	Biogeography Based Classifier	Membrane Computing	**Cuckoo Search**
0.7364	0.7525	0.9134	0.9525	0.9881	0.6715	0.68812	**0.9465**

Though kappa coefficient is more for some of the other classifiers such as semantic web, rough set classifier etc but the classification of the heterogeneous region is done more efficiently and effectively by cuckoo Search. This is discussed in detail in the section 8.5.2.

8.5.2. Heterogeneous region Comparison for Alwar image

Till date all the classifiers only classifies a given image and achieves efficiency but they do not focuses on region where mixture of many features are possible. Though they classify homogeneous region perfectly but they do not have any impact on heterogeneous portion of the image. Heterogeneous region in an image are a common occurrence where a mixture of feature type coexist in a small region. Figure 25(a) shows the portion of the Alwar image where probability of water, barren, and vegetation area is high.

We have gone through all other classifiers and had seen that some classifiers assigned this portion as rocky and some as vegetation. But the image shows that this classification is not fully correct.

Our classifier observed this region minutely and concluded that it is neither pure water nor pure vegetation area. Rather it is the conflict dataset match has shown a strong correlation with water, vegetation, and barren dataset. Thus our Classifier has identified this heterogeneous portion and classified it to a great extend.

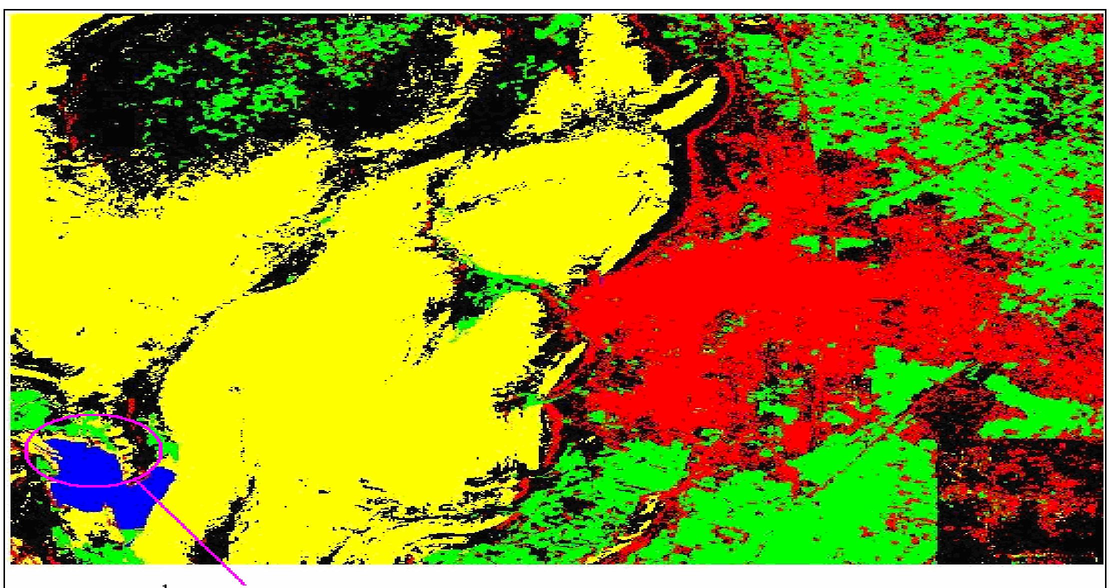
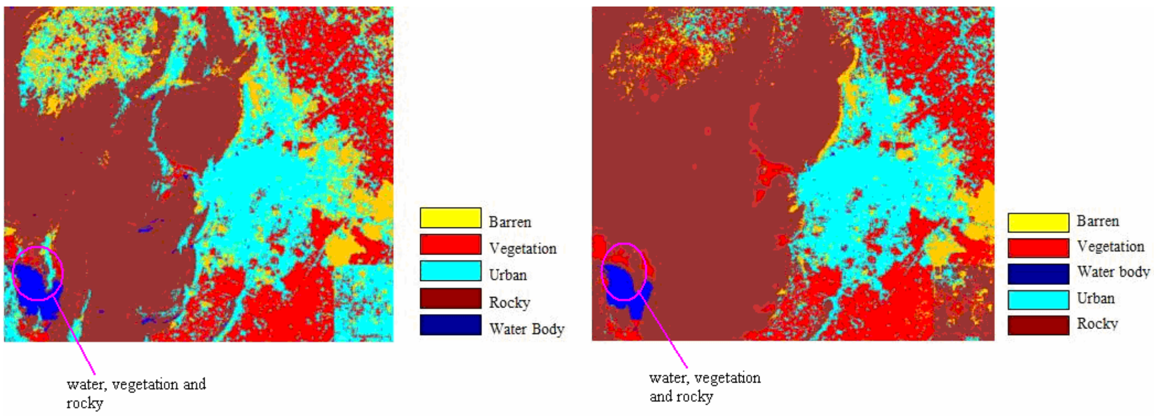


Fig 33: Identified heterogeneity and assigning proper feature by CS

In the above figure the yellow colour shows the rocky area, blue colour shows water, green colour shows vegetation and the black colour shows the barren area. Now consider our traditional classifier such as Minimum distance classifier and Maximum Likelihood classifier.

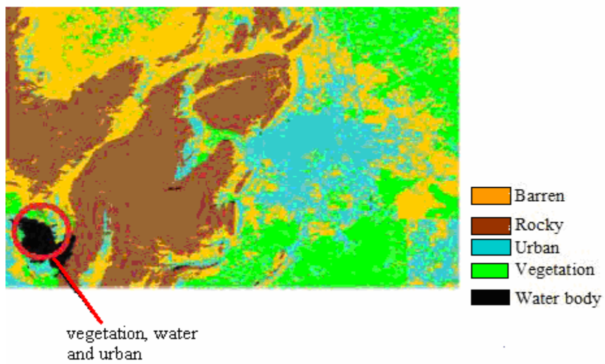


(a) Minimum distance classifier **(b) Maximum Likelihood classifier**

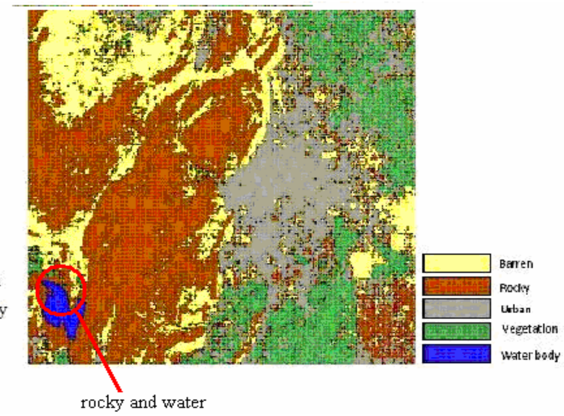
Fig 34: Traditional Classifier's output

Here we are presenting the report of the analysis performed on the results obtained by the different algorithm for a heterogeneous region of Alwar region shown in figure 25(a). When we observe the original image of Alwar region minutely we identify that the marked portion contains the pixels of water, vegetation, barren and rocky. But the result obtained for most of the algorithm discussed above is not promising for this heterogeneous region. The traditional classifiers have wrongly classified the region as water, vegetation and rocky only this is shown in figure 34. The Fuzzy classification has wrongly marked the barren region as urban and the Rough Fuzzy tie-up has wrongly marked the vegetation and barren portion as rocky. The result for cAnt Miner states that the outer portion of the marked region has only vegetation pixel which is incorrect as there exist barren pixels as well. So we can say that cAnt Miner was only able to classify water pixels correctly for the marked region. For hybrid of PSO and ACO2 the water and vegetation pixels in the marked portion are being classified correctly to some extent but the barren area is being wrongly classified as rocky.

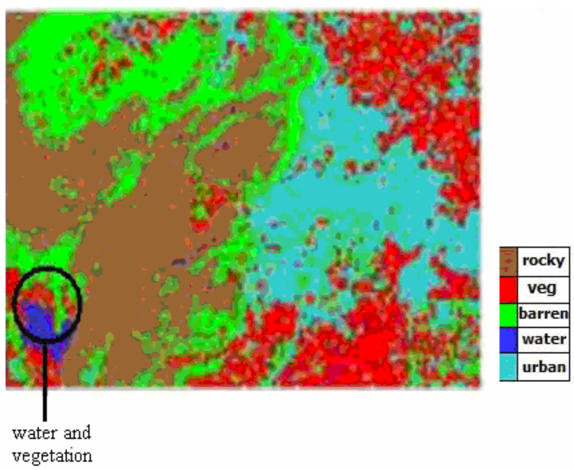
In the BBO based implementation the barren and some of the vegetation pixels of encircled area is wrongly classified as rocky. Not only this, there is an urban portion in the right side of the image which is being wrongly classified as barren. This observation shows that BBO is not able to classify some the homogeneous part of the image as well. From the results of ACO/BBO we observe that it is able to classify water and vegetation portion but this approach has wrongly classified the barren region as rocky for the portion of the image taken into consideration. But combination of ACO with BBO has provided one advantage that it is now able to correctly classify the urban and barren region which is marked on the right side of the image. Even the combination of ACO2, BBO, and PSO is not able to classify the encircled heterogeneous area correctly. Here, it is only able to classify the water pixels and has wrongly classified the vegetation and barren pixels as rocky. Semantic web based approach is only able to classify the water pixels correctly and rest of the pixels are wrongly classified as rocky. Membrane computing is the only approach which has been able to correctly classify the water and the barren pixels. But this approach also has some disadvantages firstly it has wrongly classified some of the vegetation pixels in encircled area as barren and secondly the urban area on the right side of the image is wrongly classified as barren.



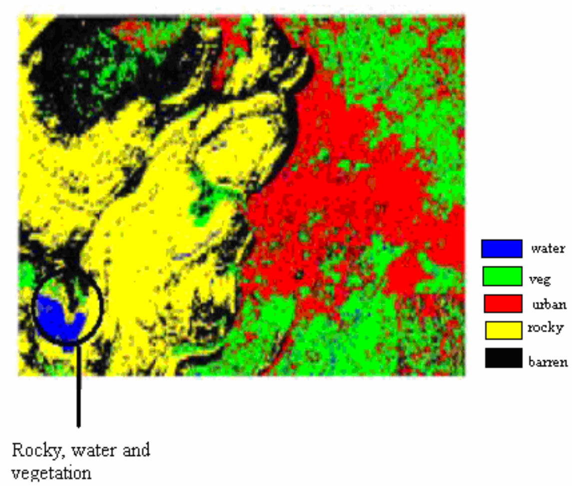
(a) Fuzzy Classification



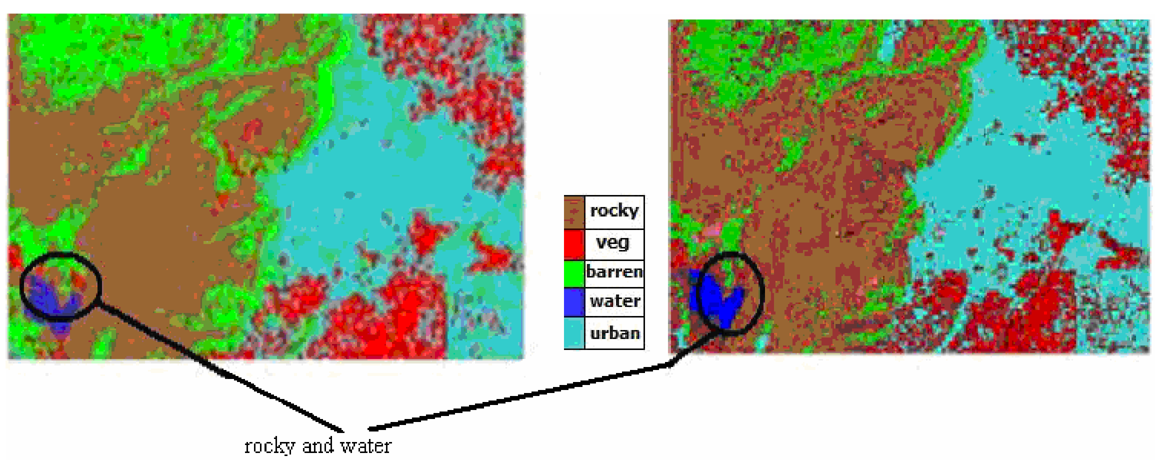
(b) Rough-Fuzzy tie up



(c) cAntMiner Algorithm

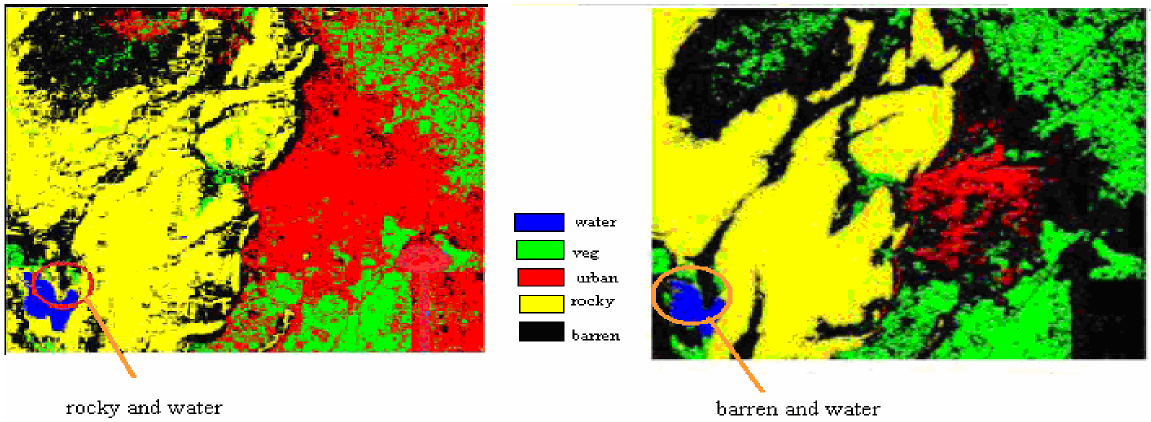


(d) Hybrid ACO-BBO Algorithm

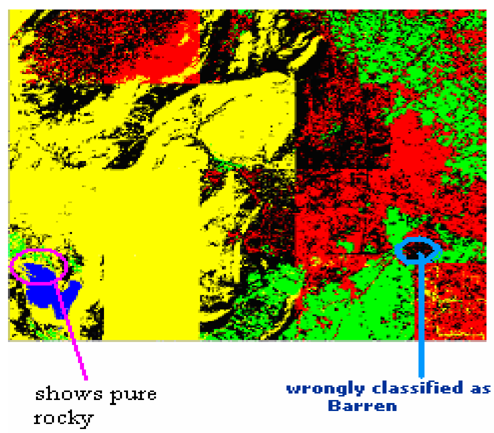


(e) Hybrid ACO2/PSO

(f) Semantic Web based classifier



(g) Hybrid ACO2/PSO/BBO **(h) Membrane computing classifier**



(i) BBO based classifier

Fig 35: Results for various algorithms

We have seen that a lot of discrepancy and uncertainty exist in separating each feature from the region. From the result analysis showed in figure 34 and 35 we can observe that the existing algorithms are not able to correctly classify the marked heterogeneous region. It shows that algorithms such as cAnt Miner, ACO2/PSO, ACO/BBO, ACO/BBO/PSO despite of having high 'k' value have wrongly classified the heterogeneous region. On the other hand MC even with low 'k' value is able to classify the heterogeneous region effectively to a certain extent. Thus we can say that a high kappa coefficient does not correspond to the correct classification of these conflict regions.

9. Publication from Thesis

During the period of working over this project we interacted with International community working on Nature Inspired algorithms. We discussed our approach with them and collected the reviews and worked over the suggestions send to us.

9.1. The details of Conference publications

1. Conference Name*:* The 2012 World Congress in Computer Science, Computer Engineering, and Applied Computing **(Accepted),** The 2012 International Conference on Artificial Intelligence (ICAI'12).

URL: http://www.worldacademyofscience.org/worldcomp12/ws

Paper Title: —"Applying Nature Inspired Metaheuristic Technique to capture the Terrain Features"

Authors: Akanksha Bharadwaj, Daya Gupta, V. K. Panchal.

Location: Las Vegas, Nevada, USA.

Conference date: July 16 - 19, 2012

Publisher/Proceedings: The paper is included in the conference proceedings, which has an ISBN number. The proceedings will also be submitted for several database indexes.

2. Conference Name: International Conference on Computer Science and Engineering 2012, Nainital, India **(Published)**.

URL: http://www.interscience.in/Conference/Nainital/iccse

Paper Title: "Comparative Analysis of Nature Inspired Techniques for Heterogeneous region in Remote Sensing"

Authors: Akanksha Bharadwaj, Srideepa Banerjee, Daya Gupta, V. K. Panchal.

Location: Nainital, Uttrakhand, India.

Conference Date: May 19, 2012

Publishers/ proceedings: The accepted papers will be included in the conference proceedings, which has an ISBN number. The proceedings were made available during the conference.

10. Conclusion and Future Scope

In this section the conclusion and the future scope of this work is presented.

10.1. Conclusion

A large numbers of soft computing techniques have been used for the classification of multi-spectral satellite image. All these techniques classified the terrain features but suffered from some uncertainties. Cuckoo Search is a new nature inspired metaheuristic algorithm which hardly has any footprint in any of the application. Cuckoo Search is an emerging area which is not been applied to image classification. Thus terrain features classification is taken as a case study. Our proposed implementation is pixel by pixel and hence overcomes the disadvantages of the previous techniques like BBO, ACO which were implemented as cluster based approaches. The experiments have shown that our algorithm is an efficient land cover classifier for the satellite image. Many of the land cover features are identified much more clearly when proposed algorithm is used. By using this approach we are able to classify the satellite image according to different areas like water, urban, vegetation, barren region etc with different colors assigned to each feature's pixels. It is perceived that the Kappa coefficient can be considered as a well-founded metric for assessing accuracy of classification in remote sensing. A novel approach for feature extraction from high resolution multi - spectral satellite image is presented in this thesis. The result of the experiment for Alwar region shows that the water, vegetation and rocky regions are classified with 100% efficiency while urban region with almost 88% and barren region with 95% of efficiency. The Kappa coefficient obtained for Alwar region is 0.9463 which is considered to be highly accurate as compared the Kappa coefficient of MDC, MLC, BBO, Fuzzy set and Membrane Computing which were 0.7364, 0.7525, 0.6715, 0.913 and 0.68812 [12] respectively. When this algorithm was applied to Saharanpur region the kappa coefficient obtained was 0.9588. In this experiment 100% efficiency was obtained for medium vegetation, sparse vegetation and urban regions. Thus we can say that our proposed method for the classification of pixels is extremely helpful in providing high level of accuracy.

Mixed pixel is one of the most crucial problems being faced by image classification algorithms, so there was a need to present a technique that can resolve such problem. Our proposed algorithm given in chapter 6 can solve this problem with great ease. The results obtained from our experiment depicts that the labeling of mixed pixels with an appropriate land cover class has been done efficiently and effectively. Moreover, the main advantage of this approach is that its performance doesn't degrade even with the increase in the number of mixed pixels which was not the case in some of the previous algorithms proposed [16]. Cuckoo Search not only identified homogeneous region but also successfully tagged the heterogeneous regions and the mixed pixels. Thus our proposed methods were successfully able to extract the land cover features from the given dataset and also maintained high levels of classification accuracy

10.2. Future Scope

The future scope of the research includes proposing certain modification to the algorithm so that the Kappa coefficient can be improved further. The current system is implemented using KNN i.e. K- Nearest Neighbor and a simple heuristic technique; the system performance can be increased by using other heuristic functions. The system performance can be further increased by using better unsupervised classifications and better training sets. Moreover it had identified most of the land cover features but some regions need to be identified properly. For this it can be combined with other bio-inspired algorithm in order to classify all the homogeneous and heterogeneous region of the image more efficiently and effectively.

11. REFERENCES

[1] Ralph W. Kiefer, Thomas M. Lillesand, “Remote Sensing and Image Interpretation”, John Wiley, New York, 2000.

[2] Shefali Aggarwal “Principle of remote sensing” Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, pp. 23-38.

[3] V.K Panchal, Samiksha Goel, Mitul Bhatnagar, “Biogeography Based Land Cover Feature Extraction”, VIII International Conference on Computer Information Systems and Industrial Management (CISIM), Coimbatore, 2009.

[4] Marco Dorigo and Thomas Stutzle, “Ant Colony Optimization” IEEE Transaction on Evolutionary Computation, 2002, Vol 6, Issue 4, pp-358-365.

[5] Xin-She Yang, Suash Deb, “Cuckoo search via Levy flights”, In Proceedings of World Congress on Nature & Biologically Inspired Computing, (NaBIC), IEEE Publications USA, India, 2009, pp 210-214.

[6] Blum C. and Roli A., “Metaheuristics in combinatorial optimization : Overview and conceptual comparison”, ACM Comput. Surv. 2003, pp-268-308.

[7] T.M. Lillesand, R.W. Kiefer & J.W. Chipman,”Remote Sensing & Image Interpretations” 5th edition, Wiley & Sons Pvt. Ltd., England, 2005, pp 586-592.

[8] L.A. Zadeh, “Fuzzy sets”, Information & Control, 1965, vol.8, no.3, pp 338-353.

[9] Z. Pawlak, “Rough Set Theory & its Application to Data Analysis”, Cybernetics & Systems, 1998, vol. 29, no. 7, pp 661-688.

[10] E. Bonabeau, M. Dorigo, G. Theraulaz, “Swarm Intelligence from Nature to Artificial System”, First Edition, Oxford University Press, USA, 1999, pp 1-24.

[11] S.N.Omkar, K.M. Manoj, D. Mudigree & D. Muley, “Urban Satellite Image Classification using Biologically Inspired techniques", In Proceedings of IEEE International Symposium on Industrial Electronics, Vigo, Spain, 2007, pp- 1767-1772.

[12] Daya Gupta, Lavika Goel, V.K.Panchal, “Embedding Expert Knowledge to Hybrid Bio-Inspired techniques- An Adaptive Strategy towards Focussed Land Cover feature Extraction”, International Journal of Computer Science & Information Security, (IJCSI), 2010, vol.8, no.2.

[13] Xin-She Yang, Suash Deb, "Engineering Optimization by Cuckoo search", International Journal Mathematical Modelling & Numerical Optimization, 2010, vol. 1, no.4, pp 330-343.

[14] Pi-Fuci Fsieh Lee, L.C. Nai-Yu Chen Nat. Space Program, Hsinchu, "ICA-aided mixed-pixel analysis of hyperspectral data in agricultural land" IEEE Geoscience and Remote Sensing Letters, 2005, vol. 2, pp 220-224.

[15] V.K. Panchal and Nitish Gupta. "Swarm Intelligence for Mixed Pixel Resolution", IEEE Geo Sciences and Remote Sensing Symposium (IGRASS), 2011, pp- 2801-2804.

[16] Suruchi Sinha, V.K. Panchal, Abhishek Bhola, Siddhant Singhal, Ajith Abraham "Resolving Mixed Pixel Through Hybridization Ant Colony Optimization and Bio Geography Based Optimization ", accepted in IEEE World Congress Computational Intelligence, Brisbane Australia 2012.

[17] ERDAS Software -- http://www.erdas.com/.

[18] Canada Center for Remote Sensing, "Fundamentals of Remote sensing", pg no. 1 - 64.

[19] T.M. Lillesand and R.W. kiefer "Remote Sensing & Image Interpretation " , 1994, 3rd edition.

[20] Sonal Kumar, Daya Gupta, V.K.Panchal, Shashi Kumar, "Enabling web services for classification of satellite images", In International Conference on Semantic Web & Web Services, SWWS, Orlando, 2009.

[21] Daya Gupta, J. Kartar, "Knowledge Representation with Ontology & Relational Database to RDF Converter", In Proceedings of Enterprise Information Systems & Web Technologies, 2008, pp 141-146.

[22] Shelly Bansal, Daya Gupta, V.K.Panchal, Shashi Kumar, "Swarm Intelligence Inspired Classifiersin Comparison with Fuzzy & Rough Classifiers – A Remote Sensing Perspective", International Conference on Comtemporary Computing, IC3, Noida, India, 2009.

[23] Lila Kari, Grzegorz Rozenberg, "The Many Facets of Natural Computing", Communication of the ACM, 2008, vol.51, no.10.

[24] P. Fisher, "The pixel: a Snare or a Delusion", International Journal of Remote Sensing, 1997, Vol. 18, pp. 679-685.

[25] Ereck R. Speed, "Evolving a Mario agent using cuckoo search and softmax heuristics," 2010 2nd International IEEE Consumer Electronics society's games innovations conference, ICE-GIC, Hong Kong, pp 1-7.

[26] Payne R. B., Sorenson M. D., and Klitz K., "The Cuckoos", Oxford University Press, 2005.

[27] Albert Barabasi et al., "The human disease network", Plos.org.

[28] Daya Gupta, Bidisha Das, V.K. Panchal, “A methodological study for the Extraction of Landscape Traits using Membrane Computing Technique”,Proceedings of International conference on genetic & evolutionary methods”,GEM ’11.

[29] Fernando E.B. Otero, Alex A. Freitas, and Colin G. Johnson, *c*Ant-Miner: An Ant Colony Classification Algorithm to Cope with Continuous Attributes in Springer-Verlag Berlin, Heidelberg 2008.

[30] Lavika Goel, V.K. Panchal, Daya Gupta, Rajiv Bhola “Hybrid ACO-BBO Approach for predicting the Deployment Strategies of enemy troops in a military Terrain Application”, 4th International MultiConference on Intelligent Systems & Nanotechnology (IISN), 2010.

[31] Zhou, Wei, “Survey and analysis of land satellite remote sensing applied in highway Transportation, Infrastructure and System Engineering”, IEEE International Geoscience and Remote Sensing Symposium, 2008.

[32] “An Introduction to Remote Sensing & Geographic Information System”, Yale University, Genocide Studies Project, Remote Sensing & Geographic Information System Reasearch.

[33] Fernando E.B. Otero, Alex A. Freitas, and Colin G. Johnson, *c*Ant-Miner: An Ant Colony Classification Algorithm to Cope with Continuous Attributes in Springer-Verlag Berlin, Heidelberg 2008.

Appendix - A

Abbreviations

EM – Electromagnetic

DN – Digital Number

BBO – Biogeography Based Optimization

MDC – Minimum Distance Classifier

MLC – Maximum Likelihood Classifier

SC – Soft Computing

RS – Rough set theory

SVM – Support Vector Machine

RS1 – Radarsat-1

RS2 – Radarsat2

LS-SVM – Least square support vector machine

MC – Membrane Computing

ACO – Ant Colony Optimization

PSO – Particle Swarm Optimization

CS – Cuckoo Search

SI – Swarm Intelligence

FL –Fuzzy Logic

NIR – Near Infra-Red

MIR – Middle Infra-Red

DEM – Digital Elevation Model

Appendix - B

Introduction to MATLAB Software

URL: http://mathworks.com/

MATLAB is a high performance language for technical computing. It integrates computation, visualization and programming in an easy-to-use environment where problems and solutions are expresses in familiar mathematical notation. Typical uses include:

- Math and computation.
- Algorithm, simulation and prototyping.
- Modeling, simulation and prototyping.
- Data analysis, exploration and visualization.
- Scientific and engineering and visualization.
- Application development, including graphical user interface building.

MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations in a fraction of the time it would take to write a program in a scalar non interactive language such as C or FORTRAN.

The name MATLAB stands for Matrix Laboratory. MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects. Today MATLAB uses software developed by the LAPACK and ARPACK projects, which together represent the state-of-the-art in software for matrix computation.

MATLAB has evolved over a period of years with input from many users. In university environments, it is the standard instructional tool for introductory and advanced courses in mathematics, engineering and science. In industry, MATLAB is the tool of choice for high-productivity research, development and analysis.

MATLAB features a family of application-specific solutions called toolboxes. Very important to most users of MATLAB toolboxes allow you to learn and apply specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control system neural networks, fuzzy logic, wavelets, simulation and many others.

FILE TYPES

MATLAB can read and write several types of files. There are mainly five different types of files used in MATLAB which is used for storing data or programs.

- **M-FILES**- They are the standard ASCII files, with a .m extension to the file name. There are basically two types of files and they are SCRIPT and FUNCTION file. In general, mostly MATLAB files are saved as M-FILES.
- **MAT-FILES**- They are the binary data-files, with a .mat extension to the filename. These files are created when you save the MATLAB data with the save command. The data which you save in MATLAB can only be read by mat lab as it saves in a special format.
- **FIG-FILES-** They are the binary figure-file, with a .fig extension to the filename. Such files are created by saving a figure in this format by using the save and save

as option in it. These files basically create all kind of information which is used for again recreating a figure and can be opened by filename.fig.

- **P-FILES**- These are the compiled M-File , with a .p extension to the filename. These file can be executed directly without using any compiler and parsed in it. These files are created with the P-CODE command.
- **MEX-FILES-** These are MATLAB-callable FORTRAN and C programme, with the .mex extension to the filename. Use of these file require some experience in MATLAB and lot of patience in it.

Appendix - C

Introduction to ERDAS IMAGINE

URL: http://www.ERDAS.com/

Overview

ERDAS is pleased to provide ERDAS IMAGINE® version 8.4 [17]. Many private and commercial users who need to extract and interpret information from imagery recognize ERDAS IMAGINE as a must have. With ERDAS IMAGINE 8.4, ERDAS'latest, most advanced release of ERDAS IMAGINE; production workflows are enhanced and simplified like never before. As an example, the Batch Wizard streamlines repetitive procedures such as importing; reprojecting, and exporting large numbers of files at once, using a wizard approach to record and "re-play" commonly used procedures. Also featured is the IMAGINE Expert Classifier™ – a tool for graphically building and executing geographically aware, rules-based expert systems. This tool can be used to build decision support systems, classifiers for high-resolution imagery, GIS analysis techniques, etc. These can then be distributed to other users for use with their own data.

Also featured is the IMAGINE Expert Classifier™ – a tool for graphically building and executing geographically aware, rules-based expert systems. This tool can be used to build decision support systems, classifiers for high-resolution imagery, GIS analysis techniques, etc. These can then be distributed to other users for use with their own data.

The software consists of two parts — the Knowledge Engineer and the Knowledge Classifier.

Knowledge Engineer:

This program provides a graphical user interface for the "expert" to build a knowledge base. The knowledge base is represented as a tree diagram consisting of final and intermediate class definitions (hypotheses), rules (conditional statements concerning variables), and variables (raster, vector, or scalar). Hypotheses are evaluated by the use of rules – if one or more rules are true, then that hypothesis may be true at that particular location. A rule is evaluated based on input variables to determine if it is true. For instance, a rule could be that slopes must be gentle (less than 5 degrees) – to evaluate this, a variable is required determining the slope at every location. This could be in the form of an existing image specifying slope angles, it could come from a spatial model calculating slope on-the-fly from an input DEM, or it could even be an external program. Variables can also be defined from vectors and scalars. If the variables' value indicates that the rule is correct, this (combined with other correct rules) indicates that the hypothesis (class allocation) is true.

Key Features

- Graphical drag-and-drop tool for building the knowledge tree.
- Confidence value definition and propagation, or the ability to handle uncertainty, is of vital importance to the knowledge base. The expert places confidence in each rule, and as multiple rules are triggered within a tree, the Knowledge Classifier combines the confidences.

- Several rules could be true at a particular location – the one with the highest confidence is most likely to be the class for that pixel.
- Variables can be from various sources – images, vectors, scalars, graphical models, and even user-defined programs.
- The ability to include prompts for particular data files and variables enables the creation of portable knowledge bases
- Use spatial operators (as opposed to traditional per-pixel classifiers) via Model Maker.
- Enables the multiple AND'ing or OR'ing of rules through the construction of the tree branches horizontally or vertically.
- Pathway cursor enables quick feedback on the results of a classification to aid in developing and fine-tuning a knowledge base.
- Access to existing ERDAS IMAGINE tools, such as Model Maker for defining spectral/spatial operators, shortens the learning curve.
- Direct read and edit of ESRI‘s Shapefiles.
- Direct read of ESRI‘s SDE data.
- Enhanced and expanded native raster file handling.
- Re-projection of raster data on-the-fly.
- Batch processing wizard.
- Enhanced Viewer functionality
- Improved print versatility on Windows NT
- International 2-byte font support in Annotation layers
- Support for ERDAS IMAGINE .img files larger than 2 GB

Knowledge Classifier:

With a previously created expert knowledge base, a less experienced user may use the Knowledge Classifier is to apply the knowledge base to data and perform a classification.

Key features:

- Wizard interface allows non-experts to apply the knowledge base to their own data.
- Evaluate all possible classification classes, or only consider a subset of rules.
- Identify missing files and prompt user to find them automatically.
- Options to output fuzzy sets and confidence layers, as well as a classification.
- Operator only requires an IMAGINE Advantage™ license.

The Knowledge Engineer is a standard part of IMAGINE Professional 8.4. The Knowledge Classifier is a standard part of IMAGINE Advantage 8.4. Consequently expert users can design their knowledge bases using IMAGINE Professional, but then these knowledge bases can be distributed to the thousands of IMAGINE Advantage users around the world to apply the knowledge-based classification process to their own data. This portability of knowledge bases is one of the keys to the strength of the expert systems approach.