The Islamic University of Gaza
Deanery of Post Graduate Studies
Faculty of Information Technology

Optimum Feature Selection for Recognizing Objects from Satellite Imagery Using Genetic Algorithm

By:

Eyad A. Alashqar (120110378)

Supervised by:

Prof. Nabil M. Hewahi

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master in Information Technology

1436 H – December 2014

المنارة للاستشارات

www.manaraa.com

إقـــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

Optimum Feature Selection for Recognizing Objects from Satellite Imagery Using Genetic Algorithm

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وإن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

DECLARATION

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification

Student's name:	Eyad Ahmed Alashqar	اسم الطالب: إياد أحمد الأشقر
Signature:		التوقيع:
Date:		التاريخ: 19/01/2015

بسم الله الرحمن الرحيم

الجامعة الإسلامية – غزة
The Islamic University - Gaza

مكتب نائب الرئيس للبحث العلمي والدراسات العليا   هاتف داخلي: 1150

الرقم ج س غ/35/ Ref

التاريخ 2015/01/11م Date

نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحث/ **اياد احمد عبداللطيف الأشقر** لنيل درجة الماجستير في **كلية *تكنولوجيا المعلومات*** **برنامج تكنولوجيا المعلومات** وموضوعها:

اختيار الخواص الأمثل للأجسام الملتقطة من الأقمار الصناعية للتعرف عليها باستخدام الخوارزميات الجينية

Optimum Feature Selection for Recognizing Objects from Satellite Imagery Using Genetic Algorithm

وبعد المناقشة التي تمت اليوم الأحد 20 ربيع أول 1436هـ، الموافق 2015/01/11م الساعة الحادية عشرة صباحاً، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

أ.د. نبيل محمود حويحي(ينوب عنه د. توفيق برهوم)	مشرفاً ورئيساً	....
د. أشرف محمد العطار	مناقشاً داخلياً	....
د. إيهاب صلاح الدين زقوت	مناقشاً خارجيًا	....

وبعد المداولة أوصت اللجنة بمنح الباحث درجة الماجستير في **كلية *تكنولوجيا المعلومات*/ برنامج تكنولوجيا المعلومات.**

واللجنة إذ تمنحه هذه الدرجة فإنها توصيه بتقوى الله ولزوم طاعته وأن يسخر علمه في خدمة دينه ووطنه.

والله ولي التوفيق ،،،

مساعد نائب الرئيس للبحث العلمي والدراسات العليا

أ.د. فؤاد علي العاجز

ABSTRACT

Object recognition is a research area that aims to associate objects to categories or classes. Usually recognition of object specific geospatial features, as building, tree, mountains, roads, and rivers from high-resolution satellite imagery is a time consuming and expensive problem in the maintenance cycle of a Geographic Information System (GIS).

Feature selection is the task of selecting a small subset from original features that can achieve maximum classification accuracy and reduce data dimensionality. This subset of features has some very important benefits like, it reduces computational complexity of learning algorithms, saves time, improve accuracy and the selected features can be insightful for the people involved in problem domain. This makes feature selection as an indispensable task in classification task.

In our work, we propose wrapper approach based on Genetic Algorithm (GA) as an optimization algorithm to search the space of all possible subsets related to object geospatial features set for the purpose of recognition. GA is wrapped with three different classifier algorithms namely neural network, k-nearest neighbor and decision tree J48 as subset evaluating mechanism. The GA-ANN, GA-KNN and GA-J48 methods are implemented using the WEKA software on dataset that contains 38 extracted features from satellite images using ENVI software. The proposed wrapper approach incorporated the Correlation Ranking Filter (CRF) for spatial features to remove unimportant features. Results suggest that GA based neural classifiers and using CRF for spatial features are robust and effective in finding optimal subsets of features from large data sets.

Keywords: Satellite Imagery, Feature Selection, Feature Extraction, Wrapper Approach, Genetic Algorithm.

المنارة للاستشارات
www.manaraa.com

ملخص الدراسة

اختيار الخواص الامثل للأجسام الملتقطه من الأقمار الصناعية للتعرف عليها بإستخدام الخوارزميات الجينية.

على مدى السنوات القليلة الماضية شهدت الحاجة إلى استخدام بيانات الاستشعار عن بعد لإنجاز المهام المعقدة في استخراج المعالم من الصور. تعتبر استخراج معالم رسم الخرائط من الصور هي مهمة صعبة لأن الصور الجوية صاخبة بطبيعتها، ومعقدة، وغامضة. استخراج المعالم من الصور تعتبر مهمة جدا للعديد من أنشطة نظم المعلومات الجغرافية GIS مثل التحديث، والارجاع الجغرافي وكذلك تكامل البيانات الجغرافية المكانية.

اختيار الخواص هي عملية اختيار أقل عدد من الخواص بحيث يحقق أعلي نسبة من الدقة فى تصنيف البيانات وتقليل حجم البيانات المستخدمة فى التصنيف. ولها العديد من الفوائد مثل تقليل الوقت اللازم لعملية تصنيف البيانات, تخفض من تعقيدات خوارزميات التصنيف, توفر أيضا الوقت وتحسن من دقة التصنيف.

فى هذه الأطروحة تم اقتراح منهجية التجميع بالإعتماد على الخوارزميات الجينية للبحث عن جميع إحتمالات الخواص الأمثل, مع استخدام ثلاث خوارزميات للتصنيف وهي الشبكات العصبية, شجرة القرارت و أقرب جار, تم تنفيذ التجارب بمساعدة برنامج WEKA على مجموعة بيانات تحتوي على 38 من الخواص المستخرجة من صور الأقمار الصناعية بإستخدام برنامج ENVI. بعد إجراء العديد من التجارب تم إقتراح استخدام فلترة الخواص المكانية لحذف الخواص الغير ضرورية والتي تؤثر بشكل سلبي علي دقة التصنيف. وقد تبين لنا ان استخدام الشبكات العصبية للتصنيف مع فلترة الخواص المكانية بالإعتماد على الخوارزميات الجينية تكون فعالة فى إيجاد الخواص الأمثل.

الكلمات المفتاحية: صور الأقمار الصناعية, اختيار الخواص, استخلاص الخواص,منهجية التجميع, الخورازميات الجينية

المنارة للاستشارات
www.manaraa.com

ACKNOWLEDGEMENT

First, I thanks **Allah** for guiding me and taking care of me all the time. My life is so blessed because of his majesty.

I would like to thank **my parents** and my entire family for providing me unconditional support and encouragement throughout my time in postgraduate. Special thanks must also go to my brother **Eng. Wesam** for his help in collecting satellite imagery.

My heartiest gratitude to my wonderful wife, **Doaa**, for her patience and forbearance through my studying and preparing this thesis, and my son, **Ahmed**, whom I do all of this for him.

I kindly thank my supervisor **Prof. Nabil M. Hewahi** for his constant guide, challenging discussions and advices. I am grateful to him for working with me. I learned so much, it has been an honor.

I would like to express my appreciation to the academic staff of information technology program at the Islamic University-Gaza.

Eyad A. Alashqar
December 2014

المنارة للاستشارات
www.manaraa.com

Table of Contents

المنارة للاستشارات
www.manaraa.com

المنارة للاستشارات
www.manaraa.com

LIST OF TABLES

www.manaraa.com

LIST OF FIGURES

www.manaraa.com

المنارة للاستشارات
www.manaraa.com

List of Abbreviations

GIS	Geographic Information System
FOV	Field of View
IFOV	Instantaneous Field of View
DN	Digital Number
GCPs	Ground Control Points
FSS	Feature subset selection
GA	Genetic Algorithm
ANN	Artificial Neural Network
KNN	K Nearest Neighbor
ENVI®	The Environment for Visualizing Images
WEKA	Waikato Environment for Knowledge Analysis
J48	An open source Java implementation of the C4.5 decision tree algorithm
FS	Feature selection
LR	Learning Rate
HL	Hidden Layer
Epochs	An epoch is a measure of the number of times all of the training vectors are used once to update the weights.
CF	Confidence Factor
Shapefile	The shape file format is a digital vector storage format for storing geometric location and associated attribute information.
CRF	Correlation Ranking Filter

المنارة للاستشارات
www.manaraa.com

CHAPTER 1: INTRODUCTION

This chapter describes historic overview of remote sensing technology and its development stages. It discusses the characteristics of satellite sensors as well as the most of the common image processing available in image analysis systems. Moreover, discuss the feature selection based on wrapper approach, with more details about genetic algorithm and classification algorithms.

1.1 Principles of Remote Sensing

Remote sensing, also called earth observation, is the science (and to some extent, art) that can be broadly defined as any process whereby information is gathered about an object, area or phenomenon without being in contact with it .This is done by sensing and recording reflected or emitted energy and processing, analyzing, and applying that information. Our eyes are an excellent example of a remote sensing device. We are able to gather information about our surroundings by gauging the amount and nature of the reflectance of visible light energy from some external source (such as nature light as the sun or industry light bulb) as it reflects off objects in our field of view [52]. For more details see Appendix A.1.

1.2 Feature subset Selection

The goal of the Feature Subset Selection (FSS) is to detect irrelevant and/or redundant features as they harm the learning algorithm performance [36]. A good FSS algorithm can effectively remove irrelevant and redundant features and take into account feature interaction. This not only leads up to an insight understanding of the data, but also improves the performance of a learner by enhancing the generalization capacity and the interpretability of the learning model [18]. In other words, no new feature is created, the features that are considered irrelevant or redundant are discarded, and we ideally would end up with the best possible feature subset, that is, the subset with minimum size and which leads to the minimum classification error rate. Feature selection with subset evaluation requires defining how to search the space of feature subsets (search method) and what measure to use when evaluating a feature subset (evaluation criterion) as well as the initial feature set and a termination condition.

المنارة للاستشارات
www.manaraa.com

Selecting a good subset of relevant attributes can improve not only the speed of the classifier but also its accuracy and the dimensionality of data [18, 12, 19, 31]. Another important advantage of feature selection is that it allows a better insight on the process that produced data [18, 19].

FSS methods fall into two broad categories: Wrapper and Filter [32, 29]. The Wrapper approach uses the error rate of the classification algorithm as the evaluation function to measure a feature subset as shown in Figure 1-1, while the evaluation function of the Filter approach is independent of the classification algorithm. The accuracy of the Wrapper approach is usually high; however, the generality of the result is limited, and the computational complexity is high. In comparison, Filter approach is of generality, and the computational complexity is low. Because the Wrapper approach is computationally expensive [56], the Filter approach is usually a good choice when the number of features is very large. Thus, we focus on the Wrapper method in our experiment, because we have only 38 features.

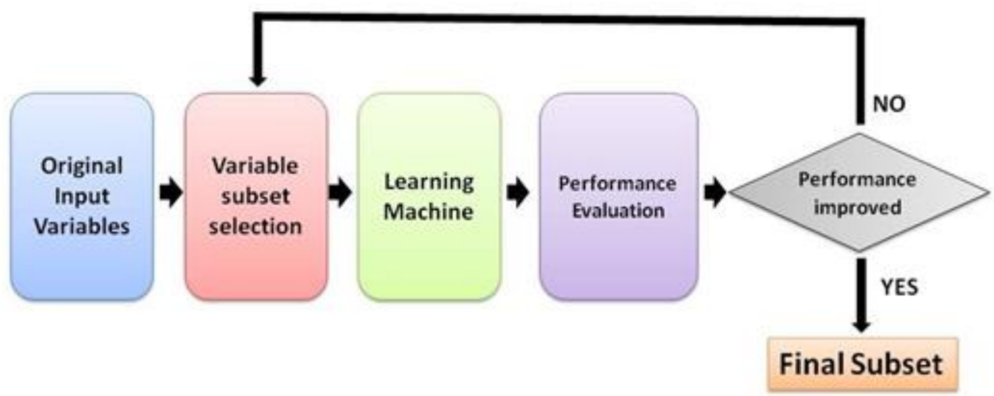


Figure 1-1: Feature Subset Selection algorithm, Wrapper approach

We can evaluate the performance of an FS algorithm; depends on three criteria:

1. **The classification accuracy**: We use the classification accuracy for selected features to measure how well the selected features describe a classification problem.

www.manaraa.com

2. **The runtime:** We use the runtime to measures the efficiency of an FSS algorithm for picking up the useful features. It is also view as a metric to measure the cost of feature selection.
3. **The number of selected features**: We use the selected features to measure the simplicity of the feature selection results, and the dimensionality of data.

Feature subset selection aims to improve the performance of learning algorithms, which usually is measure with classification accuracy. The FSS algorithms with higher classification accuracy are in favor. However, the runtime and the number of selected features cannot be ignoring. This can be explained by the following two considerations [18]:

1 Assume there are two different FSS algorithms Ax and Ay, and a given data set D. If the classification accuracy with Ax on D is slightly greater than that with Ay, but the runtime of Ax and the number of features selected by Ax are much greater than of Ay, then Ay is often choose.

2 Usually, we do not prefer to use the algorithms with higher accuracy but longer runtime, so is those with lower accuracy but shorter runtime. Therefore, we need a tradeoff between classification accuracy and the runtime of feature selection/the number of selected features. For example, in real-time systems, it is impossible to choose the algorithm with high time-consumption even if its classification accuracy is high.

As previously mentioned, we focused on the Wrapper method in our experiment, we need to use search algorithm to find best subset of features and classifier to evaluate the features subset. A number of search procedures had proposed for feature selection, thus, we focus on the Genetic Algorithm (GA) in our experiment, because it is generally known that GA is better in large populations.

1.2.1 Genetic Algorithm (GA)

Genetic algorithms (GA), a general adaptive optimization search methodology based on a direct analogy to Darwinian natural selection and genetics in biological systems, is a promising alternative to conventional heuristic methods. GA work with a set of candidate

المنارة للاستشارات

www.manaraa.com

solutions called a population. GA work based on ‘survival of the fittest’, the GA obtains the optimal solution after a series of iterative computations. GA generates successive populations of alternate solutions that are representing by a chromosome, i.e. a solution to the problem, until acceptable results are obtaining. Associated with the characteristics of exploitation and exploration search, GA can deal with large search spaces efficiently, and hence has less chance to get local optimal solution than other algorithms [17].

If we are solving some problem, we are usually looking for some solution, which will be the best among others. The space of all available solutions, it means objects among those the desired solution is called search space. Each object in the search space represents one feasible solution. Each available solution can be "marked" by its value or fitness for the problem.

An initial population is created containing a predefined size (number of chromosomes), each represented by a genetic string. Each chromosome has an associated fitness value, typically representing an accuracy value. The concept that fittest (or best) individuals in a population will produce fitter offspring to be used in the next produced population. Selected individuals are choosing for reproduction (or crossover) at each generation; with an appropriate mutation factor to random modify the genes of an individual, in order to develop the new population as shown in Figure 1-2.

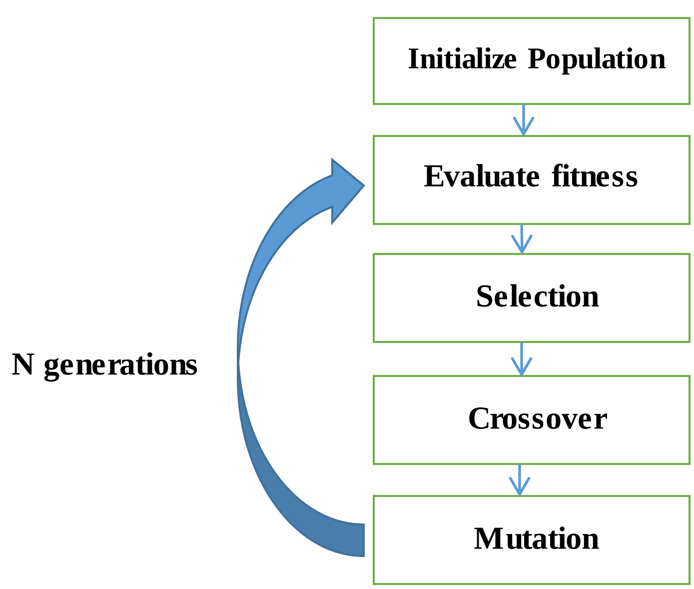


Figure 1-2: Overview of simple genetic algorithm

www.manaraa.com

Figure 1-3 shows idea of the basic genetic algorithm. Each of the L subset of features in the population in generation k is representing by a string of bits of length N, called a chromosome. Each classifier is scoured according to its accuracy on a classification task, giving L scalar values.

The chromosomes are then ranked according to this accuracy. The chromosomes are considered in descending order of score, and operated upon by the genetic operators of replication, crossover, and mutation to form the next generation of chromosomes of the offspring. The cycle repeats until a classifier exceeds the higher accuracy.

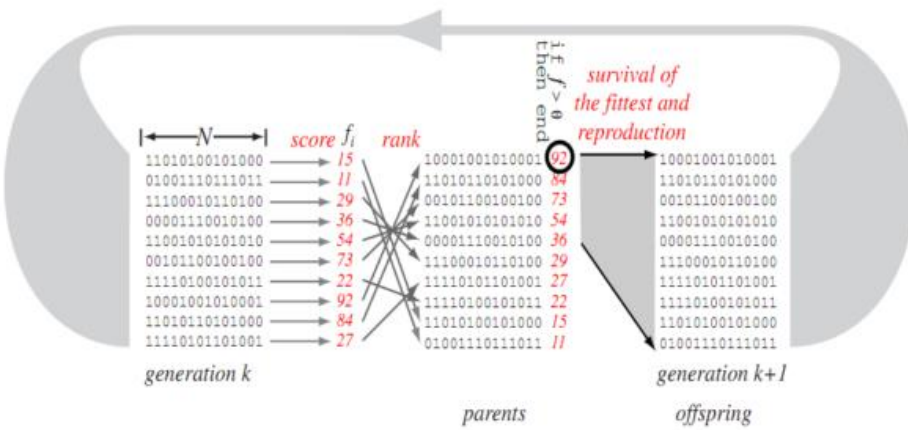


Figure 1-3: A basic genetic algorithm is a stochastic iterative search method [17]

The GA consists of three main stages: selection, crossover and mutation.

1. **Selection (survival of the fittest)**

Selection is a genetic operator that chooses a chromosome from the current generation's population for inclusion in the next generation's population based on fitness value. For maintained the good results the best chromosomes should survive and create new offspring. To select the best chromosomes, there are many methods for that, such as roulette wheel and rank selection.

2. **Crossover**

After the selection of the best chromosomes, we will create new population to perform crossover. Crossover selects sub-string (genes) from parent chromosomes and creates a

www.manaraa.com

new offspring. The simplest way to do this is to choose randomly some crossover point and everything before this point copy from a first parent and then everything after a crossover point copy from the second parent, as shown in Figure 1-4.

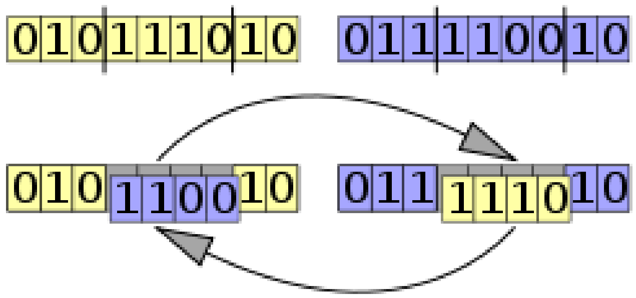


Figure 1-4: The crossover operation in GA [51]

3. Mutation (random modifications)

After a crossover is performed, mutation operator that changes one or more bit values in a chromosome from its initial state. Mutation operator prevent populations to falling into local optimum solutions. For bit-string encoding, we can switch a few randomly chosen bits from 1 to 0 or from 0 to 1. Mutation can then be following, as shown in Figure 1-5:

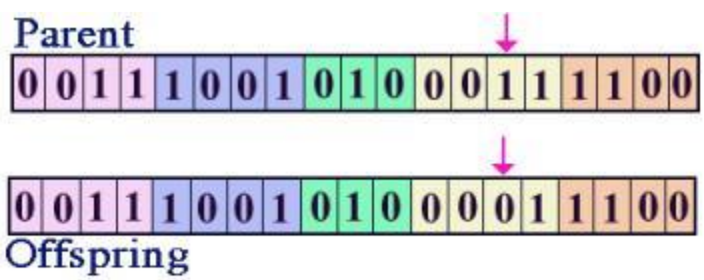


Figure 1-5: The mutation operation in GA [51]

At the end of the discussion about genetic algorithm improvements, we will list some of the attractive advantages and some disadvantages of genetic algorithms:

Advantages:

- Using chromosome-encoding GA can solve every optimization problem.
- It solves problems with multiple solutions.

المنارة للاستشارات
www.manaraa.com

- Easy to incorporate with other methods.
- Can easily run in parallel.

Disadvantages:

- There is no absolute assurance that a GA will find a global optimum.
- Often computationally expensive, i.e. slow.
- Sometimes it is difficult to find an encoding and a good fitness function.
- The quality of a result is often hard to validate.

1.2.2 Classification Algorithms

The wrapper approach was applied as black box using three classifiers, Artificial Neural Network (ANN), K-Nearest Neighbors (KNN) and J48 Decision tree within optimize search algorithm (Genetic Algorithm).

1.2.2.1 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are an attempt to model the power of the brain [5]. The brain has evolved many efficient ways to store and process information that we attempt to model through artificial neural networks.

ANN had their start relatively recently in the 1940's. The basic processing unit of a neural network is the neuron. McCullough and Pitts published the first model of the neuron in 1943 [49]. At the highest level, a neuron receives a series of inputs and depending upon the strength of the input and the connection determines whether the neuron will fire or not. The inputs are multiplying by their synaptic connection and summed. This sum is then using as input for a transfer function, which calculates the output of the neuron. This function is represented by Equation 1-1. The basic conceptual framework for a single neuron is show in Figure 1-6.

$$y_k = \varphi\left(\sum_{j=0}^{m} w_{kj} x_j\right) \qquad (1\text{-}1)$$

المنارة للاستشارات
www.manaraa.com

Where, "w" represents the weight of the synaptic connection between the input and the neuron, "x" represents the input value, and φ represents the transfer function of the neuron.

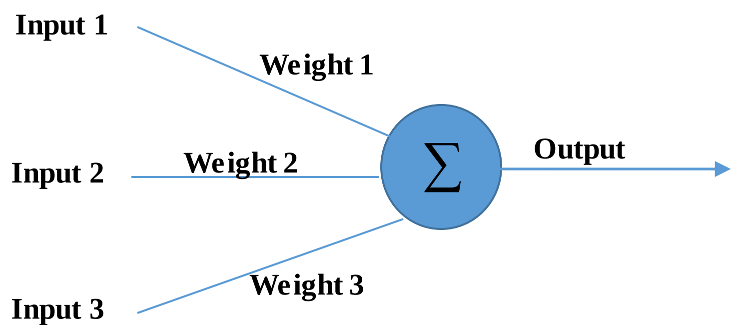


Figure 1-6: A Simple diagram of a perceptron. Lines represent connections to other neurons (synapses).

The structure of a feed-forward artificial neural network (i.e. multi-layer perceptron) includes input, hidden and output layers see (Figure 1-6). The input layer introduces the distribution of the data for each class to the network. Each input layer node represents one of the input objects features; we will be extracting them from satellite imagery. The output layer is the final processing layer that has a set of values to represent the classes such as (Roads, Buildings, and Rivers).

Training is an iterative process that seeks to modify the network through numerous presentations of data. There are many different methods to train neural networks, the two main distinctions are unsupervised and supervised learning [4]. An unsupervised neural network only uses the input data to adjust its synaptic weights. Supervised learning however relies on a set of training data with known target values. In other words, the training data consists of a set of input patterns and output values. The goal of training is to optimize a function that will map the inputs to the outputs that can be used to correct approximate unseen inputs.

Constructing an ANN using a supervised learning methodology requires the initialization of a network with random synaptic weights between neurons. At this point, an input signal presented to the network would result in no meaningful output. To derive a meaningful output the network synapses must be adjusted. The method to adjust the many weights of the network requires a calculation of error of the network for an input pattern at each epoch. An epoch represents an iteration of measuring the output error and

www.manaraa.com

updating the synaptic weights in response. A learning rate is often used to control how quickly the weights are updated. If a large value is used the weights of the network will oscillate wildly if set too low it will take more epochs to adjust the weights.

After training is completed, usually signaled by a lack of further decrease in the error or after a set number of epochs, the weights of the network are set and testing of new samples begins. During testing, the testing data is presented to the network to obtain a measure of performance. This performance is measured by a similar method that is using to determine the error of the network during training.

1.2.2.2 K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is the most basic instance-based method [15, 28]. KNN is also a lazy learning method where it does not decide how to generalize beyond the training examples until each new input is encountering. In its basic form, the learning phase in IBL algorithms consists of simply saving the normalized feature values of all training instances. With KNN, the classification phase is conducting for a given sample by calculating its pair-wise similarity with all training instances. The similarity is defined by a given similarity function, for example the additive inverse of the Euclidean distance, this function is represented in Equation 1-2. Given a new instance to be classified, its class membership is determining by the most common class of its k nearest neighbors in terms of pair-wise similarities. Because the computation is doing in the classification phase rather than in learning, IBL algorithms are relatively fast at learning but slower at classification [21].

$$\text{Euclidean} \qquad \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (1\text{-}2)$$

Nearest neighbor, algorithms in general are susceptible to the curse of dimensionality [15]. For an instance to be classified, the predicting region is defined to be the sub-region of the input space containing its k nearest training instances. This formulation leads to a problem when the number of dimensions, n for example, in the input space is large. Because of the geometry of the Euclidean spaces, the radius of the prediction region

المنارة للاستشارات
www.manaraa.com

grows in the proportion of the nth root of the volume whereas the number of training points in the region varies linearly with the volume. Therefore, with large number of features, the variance of the similarities in the predicting regions is high to the proportions that can make the similarity measures misleading.

To overcome this problem, a crucial choice is provided to the k-value. A small k-value can reduce the growth of the volume of the predicting region while a big k-value can reduce the effect of noise in the data [15].

In addition, feature selection as a means to avoid the problem can be effective with the nearest neighbors' classifiers. Because each feature alone is giving the same weight in classification, redundant and irrelevant features can distort the performance of the classifier. An irrelevant feature introduces misleading bias to the similarities and redundant feature causes a particular background concept behind several features to dominate [21].

1.2.2.3 J48 Decision tree

Decision Trees are a popular family of supervised learning algorithms. Decision Trees origin from the field of decision and statistics theory [45].
Decision trees are directed graphs with a root, internal nodes, branches and leaves (also known as terminal nodes or decision nodes). All internal and terminal nodes have exactly one incoming branch. The root and the internal nodes have two or more branches leading to their child nodes.

The process of building a tree model from the training set is knows as tree induction or tree growing. The most commonly used approach is the greedy top–down method. The basic idea is to recursively "test on attributes to partition the training data into smaller and smaller subsets until each subset contains instances that belong to a single class" [43].

The general algorithm starts with the entire training set and an empty model. It selects a "best" attribute and generates a node for it. The algorithm performed a test on the attribute's values and based on the outcome of this test; it partitions the instances at that

المنارة للاستشارات
www.manaraa.com

node in two or more subspaces that are associated to newly created child nodes. This process iterates recursively at each node. The tree induction stops when all instances in a node belong to the same class or if it is not worth to continue partitioning the training data further. Each leaf node has associated a class label, which is the (majority) class of the instances that are associated to that node.

The choice of the best attribute at each node is mainly based on the class distribution of the records before and after the test [39]. Most of the measures used are based on the difference between the degree of impurity at the parent node and the weighted sum of the degrees of impurity at the child nodes after splitting. The relative proportion of instances at the child nodes gives the weights. One common measure of impurity at node t is the entropy, defined as:

$$Entropy\ (t) = -\sum_{i=1}^{c} p(i|t) \log_2(p(i|t)) \qquad (1\text{-}3)$$

Where p (i|t) is the proportion of instances at node t that belong to the class i (i=1,..,c). Other impurity measures are Gini Index and Classification error [35]. When the measure of impurity is entropy, gain is also knows as information gain.

To classify a new instance, this is propagating down the tree and it is labelling accordingly to the class label in the leaf it reaches.

Pruning decision trees is a fundamental step in optimizing the computational efficiency as well as classification accuracy of such a model. Applying pruning methods to a tree usually results in reducing the size of the tree (or the number of nodes) to avoid unnecessary complexity, and to avoid over-fitting of the data set when classifying new data.

There are several decision trees algorithms, such as CHAID [27], CART [6], ID3 [42], C4.5 [41].

المنارة للاستشارات
www.manaraa.com

1.3 Digital Image Processing

Today's with high advanced technology most remote sensing data are recorded and saved in digital format. Digital image processing may involve several procedures including formatting and correcting of the images data, digital enhancement to facilitate better visual interpretation, or even automated classification of targets and features entirely by computer. A digital image that contains graphical information instead of text or a program. Pixels or cells are the basic building blocks of all digital images. Pixels are small adjoining squares in a matrix across the length and width of your digital image as shown in Figure 1-7 [48]. Each cell contain a digital number (DN) this value of each cell is related to the brightness, color or reflectance at that point.

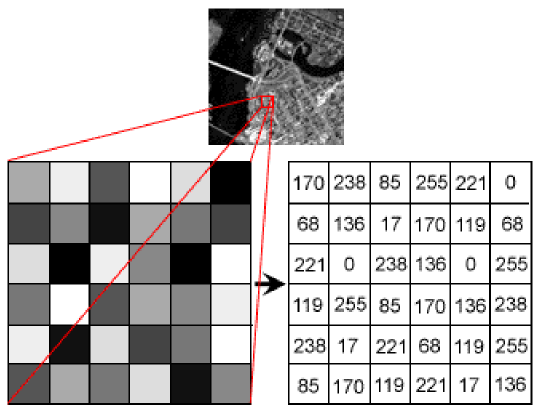


Figure 1-7: Digital image pixels [66]

Most of the common image processing functions available in image analysis systems, which categorized into the following five categories:

1. Preprocessing
2. Image Enhancement
3. Image Transformation
4. Image Segmentation
5. Feature Extraction

1.3.1 Preprocessing

Preprocessing includes data operations, which normally precede further manipulation and analysis of the image data to extract specific information. These operations sometimes

المنارة للاستشارات
www.manaraa.com

referred to as image restoration and rectification, which intended to correct for sensor and platform-specific radiometric and geometric distortions of data [52].

1.3.2 Image Enhancement

Image enhancement is the modification of an image to make it easier for visual interpretation and understanding of imagery. The advantage of digital imagery is that it allows us to manipulate the digital pixel values in an image. Most enhancement operations distort the original digital values [33].

1.3.3 Image Transformation

Digital Image Processing offers a limitless range of possible transformations on remotely sensed data. Image transformations typically involve the manipulation of multiple bands of data, whether from a single multispectral image or from two or more images of the same area acquired at different times (i.e. multitemporal image data) basic image transformations apply simple arithmetic operations to the image data [52]. For more details see Appendix A.2.3.

1.3.4 Image Segmentation

Image segmentation is the primary technique that using to convert a scene or image into multiple objects [33]. Applying the object-based paradigm to image analysis refers to analyzing the image in object space rather than in pixel space, and objects can be used as the primitives for image classification rather than pixels, so image segmentation is the process of partition an image into segments by grouping neighboring pixels with similar feature values (brightness, texture, color, etc.).

1.3.5 Feature Extraction

Feature Extraction uses an object-based method to classify the objects, where an object (also called segment) is a group of pixels with similar spectral, spatial, and/or texture attributes.

After feature extraction, we have three categories of features: spectral feature, spatial feature and texture feature, thus we have 38 features for all categories with three bands

المنارة للاستشارات
www.manaraa.com

for each feature in spectral and texture. We can divide the number of features attributes to 12, 14, and 12 for spectral, spatial and texture respectively.

1.4 Statement of the Problem

In satellite imagery we have 38 features for objects classification and recognition obtained from different features categories (i.e. spectral, texture and spatial). Obtaining the optimum set of features based on genetic algorithm, maintain the classification accuracy and reduce data dimensionality is the main problem of this research. Contrast to previous research, in our work, we need to use extracted features such as spectral, spatial and texture all together.

1.5 Objectives

1.5.1 Main Objective

Increase classification accuracy and reduce data dimensionality for satellite imagery by using GA to select the optimum feature subset.

1.5.2 Specific Objectives

- Literature review
- Use several satellite imagery to take advantage of object features and have more generalization.
- Design and implement selection of chromosome structure and calculate fitness function in GA.
- Automated selection of the optimum features subset.
- Evaluate the proposed approach according to classification accuracy.

1.6 Significance of the Thesis

Finding out an optimum set of features for satellite imagery will definitely minimize the computation time and improves the classification accuracy. This helps the experts and specialized software in the field of object recognition to determination an optimal subset of features.

www.manaraa.com

1.7 Scope and Limitations

The Satellite imagery contains too many objects such as roads, buildings, trees, rivers and vehicles…etc. Therefore, in this research we intend to extract features only for roads, building and rivers as a data set.

1.8 Methodology

The methodology that will be followed to achieve the study aim can be outlined through the following points.

- **Data collection:** Aerial photos and satellite images from number of sources that provide these images (Landsat, IKONOS, Spot, and Quick Bird satellites). We downloaded 15 imagery for training and 10 imagery for testing.
- **Image processing:** Today's with high advanced technology most remote sensing data are recording and saved in digital format. Digital image processing may involve several procedures including formatting and correcting of the images data [52]. In this stage, we need to use ENVI software for image processing.
    1. **Image Enhancement:** Image enhancement is the modification of an image to make it easier for visual interpretation and understanding of imagery. The advantage of digital imagery is that it allows us to manipulate the digital pixel values in an image. Most enhancement operations distort the original digital values.
    2. **Image Transformation:** Digital Image Processing offers a limitless range of possible transformations on remotely sensed data. Image transformations typically involve the manipulation of multiple bands of data, whether from a single multispectral image or from two or more images of the same area acquired at different times.
- **Image segmentation:** The aim of image segmentation is domain-independent partitioning of imagery into a set of visually distinct regions based on properties such as intensity (grey-level), texture, or color [69]. In this stage, we need to use ENVI software for image segmentation.

المنارة للاستشارات
www.manaraa.com

- **Feature Extraction:** After image segmentation, we need to extract features for each object (Spatial, Texture, and Spectral); also in this stage, we need to use ENVI software.

- **Feature Subset Selection (Genetic Algorithm):** After extracting the features as a data set, we will use GA as an optimization algorithm to select the best subset of features.

- **Evaluation:** In this stage, we have many steps to evaluate this work.

    1. Extract spatial features only and perform classification accuracy.
    2. Extract spectral features only and perform classification accuracy.
    3. Extract texture features only and perform classification accuracy.
    4. Extract spatial, spectral and texture features all together and perform classification accuracy.
    5. Execute correlation-ranking filter for spatial features only and perform classification accuracy.
    6. After generating features subset using GA, perform classification accuracy and compare it with others accuracies.

1.9 Outline of the Thesis

The thesis is organized as follows. Chapter 2 present some related works. Chapter 3 includes the methodology and proposed model. In Chapter 4, we present and analyze our experimental results. Chapter 5 will draw the conclusion and summarize the research achievement and future directions.

المنارة للاستشارات
www.manaraa.com

CHAPTER 2: Related works

2.1 Introduction

In the last years, an attention about the feature selection problem has been increasing. In fact, new applications dealing with huge amounts of data have been developing, such as data mining, medical data processing and satellite imagery processing. This chapter intends to give an overview for approaches related to the main topics of this thesis.

Generally, when the number of features are large but the number of training samples are small, features that have little or no discriminative information weaken the performance of classifiers. This situation is typically called the curse of dimensionality [58], in this situation we have to choose a feature subset yielding the highest performance.

It is very difficult to predict which features or features combinations will achieve better in classification accuracy. We will have different performances as a result of different features combinations. In addition, using excessive features may degrade the performance of the algorithm and increase the complexity of the classifier. Relatively few features used in a classifier can keep the classification performance robust [16]. Therefore, we have to select an optimized subset of features from a large number of available features.

2.2 Feature Selection Methods

Two major approaches for feature selection, wrapper and filter approach [29, 18, 59, 53]. Many researchers have used wrapper-filter as a hybrid approach [62, 22, 25]. In this thesis, we use wrapper approach for feature selection. In the wrapper approach, the features selection are done using the classification algorithm as a black box. The feature selection algorithm conducts a search for a good subset using the classification algorithm itself as part of the evaluation function. The accuracy of the induced classifiers is estimated using accuracy estimation techniques.

المنارة للاستشارات
www.manaraa.com

2.2.1 Filter Methods

Filter approach evaluate the goodness of the feature subset by using the intrinsic characteristic of the data. As name suggests, filters are algorithms, which filter out insignificant features that have little chance to be useful in analysis of data. Filter methods are computationally less expensive and also more generic than wrappers or furthermore hybrid methods because they do not consider underlying classifier.

Authors in [35] provide an effective feature selection for tree species classifiers in mixed-species of boreal forest. They have one dataset contains 35 input features were the 5 input spectral bands, 9 contextual features and 21 segment-wise features, they have three classes for tree species (pine, spruce and deciduous), and 4 classes for non-tree like shadow, open area (clearance), bare ground and green vegetation. The dataset was splitted in 1/3 for independent testing and 2/3 for model design, with randomly split within each class. Authors provide sequential feature selection with variable ranking and KNN classifier as evaluation technique, which means that measure the correlation between features and classes, this method reduce features from 35 to 10.

2.2.2 Wrapper Methods

Wrapper methods select a feature subset using a learning algorithm as part of the evaluation function. The learning algorithm is used as a kind of "black box" function to guide the search. The evaluation function for each candidate feature subset returns an estimate of the quality of the model that is induced by the learning algorithm, which therefore causes better estimate of accuracy. Wrapper approach based on search algorithms fall into two major categories optimal and suboptimal features subset, such as Sequential Forward Selection, Sequential Backward Selection [23, 55], Sequential Forward Floating Selection and Sequential Backward Floating Selection [40], Steepest Ascent and the Fast Constrained Search [50]. These feature selection techniques have limitations in optimal subset selection for satellite imagery due to strong correlation between features [61].

In recent years, heuristic optimization algorithms such as, genetic algorithm (GA) method [13, 53, 26, 63, 47], ant colony algorithm [61] and swarm intelligent [14, 44], have

المنارة للاستشارات
www.manaraa.com

attracted many attentions in wide range of satellite imagery classification. Many researches works on hyperspectral image, which contain a wealth of data, but interpreting them requires an understanding of exactly what properties of ground materials we are trying to measure, these images contain hundreds of bands and features, many researches work on Hyperspectral images [63, 30].

In [63], authors proposed a GA based wrapper feature selection method "GA-SVM" for hyperspectral imagery, which contains up to 200 bands. Authors used ENVI/IDL as a programming language to implement "GA-SVM", and they used two criteria to design the fitness function, namely classification accuracy and the number of selected features, to evaluate features subset. For experiments, they create training sets and testing sets using ENVI software labeled with five classes namely built-up area, water body, grassland, forest and unused land. After the experiment, the number of bands used for classification was reduced from 198 to 13, while the classification accuracy increased from 88.81% to 92.51%.

New criterion function called Thornton's separability index has been successfully deployed for the optimization of feature selection for classification satellite imagery [1, 13, 16]. Thornton's separability index is defined as the fraction of data points whose classification labels are the same as those of their nearest neighbors. Thus, it is a measure of the degree to which inputs associated with the same output tend to cluster together [16].

Anthony and Ruther in [1] tries to find the optimum combination of bands for every class. They used separability index as evaluation function with Exhaustive Search (ES) and Genetic Algorithm (GA) with SVM as a classification technique, for experiments, they used two datasets with 7 bands and contains six land cover classes were sought namely: wetlands, water (lakes and rivers), Bush/shrub/trees, Grasslands, "bare ground" and Roads. Instead of using classification accuracy to evaluate features subset, they used separability index, to evaluate every band combination. After the experiment, the result can be showed as Roads used two bands (2 & 5), while the classification accuracy increased from 67.22% to 75.32%.

www.manaraa.com

Haapanen and Tuominen [24], evaluated the potential of the combination of satellite image (spectral) and aerial photograph (spectral and texture) features to increase classification accuracy for forest inventory. In addition, authors tried to reduce the dimensionality of these features by removing unnecessary or adverse features using two feature selection GA and sequential Forward Selection (FS) with K Nearest Neighbor. Firstly, they use GA and FS to select best features from each image separately are used. Secondly, select best features from combination. Results said the accuracy of the estimation with all features was better than either the satellite image or the aerial photograph features alone.

In [57], authors proposed a method with a three-step object-oriented classification routine that involves the integration of 1) image segmentation, 2) feature selection by GAs and 3) joint Neural Network (NN) based object-classification. For feature extraction, 89 features were extracted using eCognition 3.0 software tool based on IKONOS imagery. After applying feature selection, the dimensionality of the input space is reduced from 89 to 23 and classification accuracy increased from 87.41% to 90.10%.

In [30] authors proposed wrapper approach based on GA as random search technique for subset generation with different classifiers/ induction algorithms namely decision tree C4.5, NaïveBayes, Bayes networks and Radial basis function as subset evaluating criteria on four standard datasets. Experimental results show employing feature subset selection enhanced the classification accuracy in most of the cases. Moreover, results show that no one wrappers among the four wrappers experimented is best for all the datasets experimented.

Ant colony algorithm (ACA) is a cooperative search technique that mimics the foraging behavior of real life ant colonies. Authors in [61] proposed ant colony algorithm for feature selection from hyperspectral imagery. There experiments show that the proposed method reduce the features from 200 to 20.

The goal of authors in [2] is to detect the best spectral band using particle swarm optimization with ANN for supervised classification. For experiments, they used multispectral image with six bands and four classes: road, river, vegetation and urban

المنارة للاستشارات
www.manaraa.com

area. After experiments, the results show that among the red, green and blue bands any one is getting selected in different run of the algorithm.

This paper [38], the impact of genetic search on classification accuracy for rule induction algorithms is studied. Seven rule induction algorithms: JRip, ConjuctiveRule, DecisionTable, OneR, PART, Ridor and ZeroR are used based on wrapper approaches. For experiments, 16 input features with 2 output classes are used. After the experiments, genetic search selected four attributes used in rule induction algorithms. Results show that the classification accuracy with genetic search improves or maintains the classifications with the seven rule induction algorithms. Genetic search improves the accuracy of four classifiers: JRip, Ridor, DecisionTable and PART and maintains the accuracy of tree classifiers: ConjuctiveRule, OneR and ZeroR.

2.2.3 Hybrid Filter-Wrapper Methods

The hybrid model attempts to take advantage of the two models by exploiting their different evaluation criteria in different search stages.

In [11], hybrid approach was proposed with Self-adaptive differential evolution (SADE) for searching feature subset and Fuzzy KNN classifier used to calculate the classification accuracy as evaluation criterion. Before doing experiment, authors used ReliefF algorithm for removing the redundancy and noisy of features. After the experiments, the results shown that the SADE based method requires less memory and computation cost than the other searching methods. Authors used GA and Ant Colony Optimization (ACO) based methods for compassion with proposed methods, and the results shown the proposed methods outperforms others.

In [53] authors proposed GA based hybrid feature selection with classification technique called a supervised Nearest Neighbour Distance Matrix (NNDM). WEKA software used for implementation the experiments, which conducted using 9 datasets. The initial population for the feature selection is generated based on Information Gain (IG), which used to generate correlated subset of features, the NNDM classifier used as the evaluation function to evaluate the fitness of the new population. The experiment results show that

www.manaraa.com

the proposed method can reduce estimation time needed to optimize the subset feature selection.

2.3 Classification Algorithms

Integrating the GA with other classifiers has been used to produce several feature selection algorithms such as GA-ANN, GA-KNN and GA-J48 Decision tree.

Artificial Neural Network (ANN) used in [22, 30, 1]. Some of the advantages of using ANN, is well suited to problems in which the training data corresponds to noisy and complex sensor data such as satellite imagery. It maintains non-linearity and it could deal with biggest problems. However, it is suffering from multiple local minima; the problem of local minima could be solved by using techniques such as: stochastic gradient descent and k-fold.

K-Nearest Neighbors (KNN) used in [5, 18], KNN classifier is a very simple classifier, it simply uses the training data itself for classification, however it can be slow for real-time prediction if there are a large number of training examples and is not robust to noisy data.

Decision tree is used as evaluation classifier in [46, 30, 4]. The main advantage of using decision tree is its simplicity and less sensitivity to errors. However, it is the results candidates to over-fit the training data, especially when the used training data is too small or have noise.

From the above survey, it is noticed that feature selection is of considerable importance, particularly when too many features are used. There are many research works on feature selection using many algorithms and methods, but all of the previous researches never used all the spectral, spatial and texture features all together with 38 features with spectral and texture features with 3 bands. Many of the previous work used one classifier to measure the performance of the new obtained short list of features. In our case, we apply various classifiers to ensure that the results are improving regardless of the classifier type. In addition, some research papers use ACO for feature selection, we preferred to have GA as an optimization algorithm because it is generally known that GA is better in large populations. Moreover, we selected three main objects to consider in our

المنارة للاستشارات
www.manaraa.com

research, i.e., buildings, roads and rivers, roads and rivers haven chosen among the objects because they may look to be very similar and feature selection would be very crucial. To confirm the realistic of our results, we use CRF for spatial features to remove unimportant features and this did not use in previous researches. Overall, the methodology we use is different from all other previous methods as we are going to show in the next chapter.

المنارة للاستشارات
www.manaraa.com

CHAPTER 3: Methodology and Proposed Model

This chapter contains detailed description of the steps of the methodology of our research. The proposed followed methodology is presented below and shown in Figure 3-1.

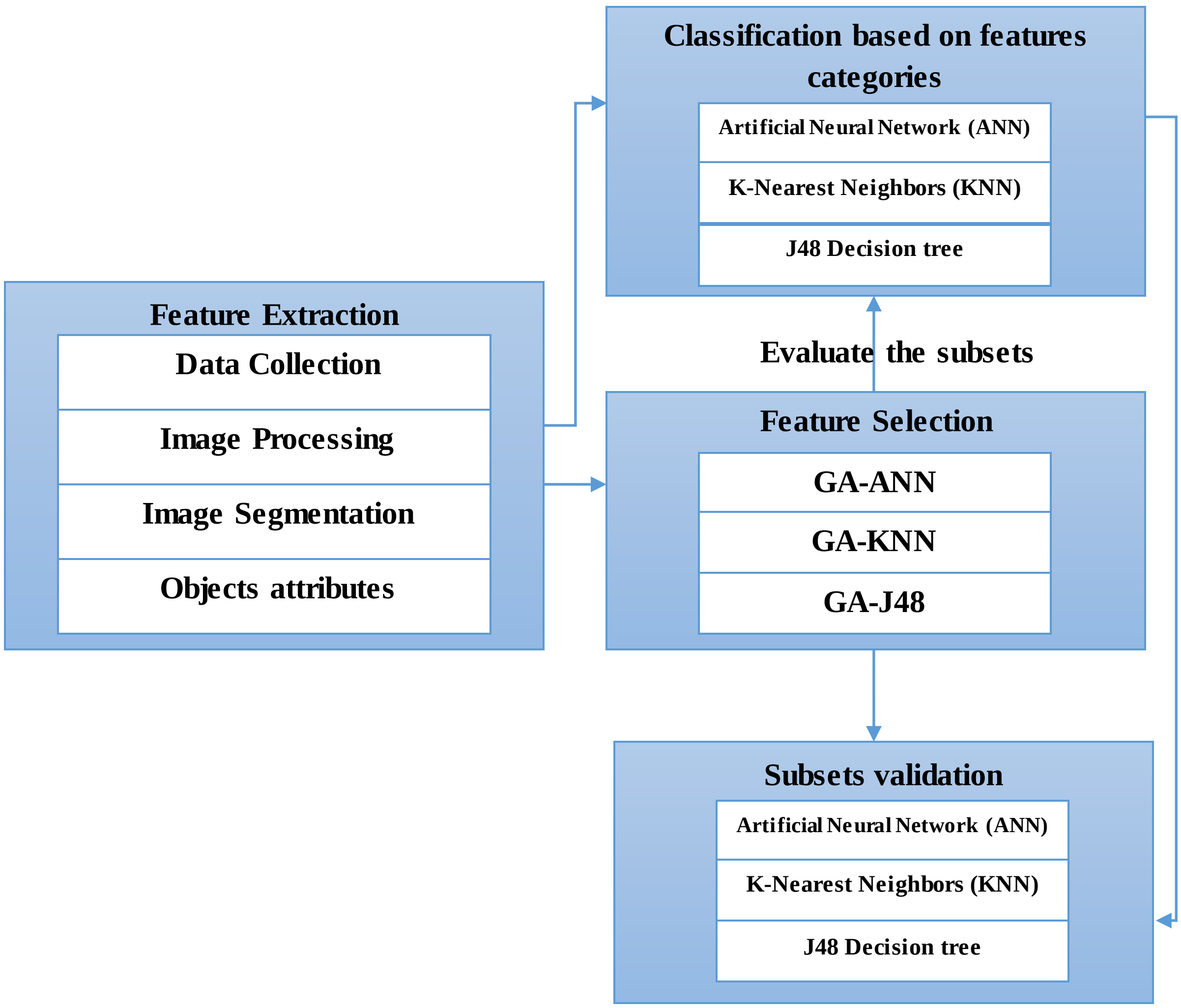


Figure 3-1 : Methodology flowchart

3.1 Data Collection and Preparation

Maps have been the main source of data for geographic analysis for many years. Raster data is commonly obtained by scanning maps or collecting aerial photographs and

المنارة للاستشارات
www.manaraa.com

satellite images. These images are produced from processing high-resolution commercial panchromatic satellite imagery, such as IKONOS, Quickbird2, and Landsat.

3.1.1 Data Collection

Many free sources offer free aerial photos and satellite images in the internet as USGS (U.S. Geological Survey) [65] and NASA website [64].

Many attempts to get suitable aerial and satellite images from various sources to apply feature extraction method have been tried. Some of criteria used to select a case study are diversity of features such as (buildings, trees, roads, and rivers):

1. **Number of objects:** as shown in Figure 3-4, we chose image contains only Roads and buildings.
2. **Contrasting colors:** as shown in Figure 3-2, we chose image contains River as a blue line, which contrasting with the green background.
3. **Spatial resolution:** as shown in Figure 3-5, we have provide high-resolution image from Gaza municipality for Gaza city.
4. **Complexity:** as shown in Figure 3-3, we chose image contains asphalt road and land road with convergent colors.

Figure 3-2: Sample (1) of satellite image describes river as a blue line

المنارة للاستشارات
www.manaraa.com

Figure 3-3: Sample (2) of satellite image describes asphalt road and land road

Figure 3-4: Sample (3) of satellite image describes asphalt road between buildings

Figure 3-5: Sample (4) of satellite image describes asphalt road between buildings and agricultural area

3.1.2 Image Preprocessing

Preprocessing of an image often include radiometric correction and geometric correction. The following subsections illustrate all needed steps of automatic feature extraction based on sample (4) as shown in Figure 3-5.

www.manaraa.com

3.1.2.1 Geometric Correction

To correct the geometric distortions as we described in Appendix A.2.1, one should apply two steps, geo-referencing and resampling using ARCGIS 10.1 or ERDAS 2013 as shown in Figure 3-6 [33].

The geographic space of each dataset is a reference according to four known coordinates corresponding to the minimum x and y values, the minimum x and maximum y values, the maximum x and minimum y values, and the maximum x and y values. Georeferencing is the process of assigning geographic information to an image. Knowing where an image is located in the world allows information about features contained in that image to be determined. This information includes location, size and distance.

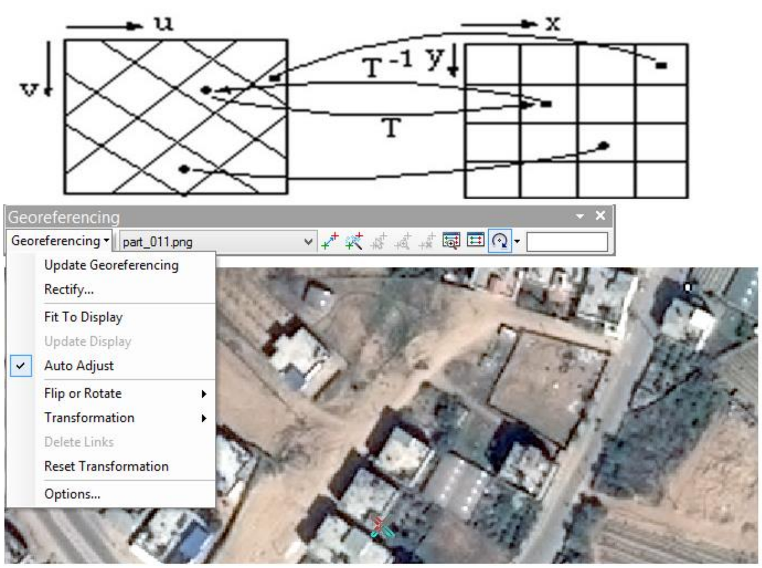


Figure 3-6: Geo-referencing method & toolbar in ARCGIS 10.1

After correcting the coordinate system, the spatial characteristics of pixels may be changed. So resampling should be applied to obtain a new image more pronounced in which all pixels are correctly positioned within the terrain coordinate system to more accurate feature extraction methods.

3.1.2.2 Radiometric Correction

Radiometric correction involves the processing of digital images to enhance the accuracy of the brightness value magnitudes. Any imagery contains radiometric errors will be

المنارة للاستشارات
www.manaraa.com

referred to as "noise". These errors should be corrected before the post-processing enhancement, extraction, and analysis of information from the image [3].

The sources of radiometric noise and the appropriate types of radiometric corrections partially depend on the sensor and mode of imaging used to capture the digital image data such as aerial photography, optical scanners, sensors and others.

Improvement quality of images, which used in sample (4), radiometric noise reduction, is performed using ERDAS 2013 as shown in Figure 3-7.

Before noise reduction After noise reduction

Figure 3-7: Noise reduction of sample (4)

3.1.2.3 Image Enhancement

Histogram processing is used in image enhancement. A histogram can tell you whether or not your image has been properly exposed, whether the lighting is harsh or flat, and what adjustments will work best [9], for more details see Appendix A.2.2.

Figure 3-8 showing the image and histogram for study area (sample 4). The histogram shows that the vast majority of the pixels are of medium intensity. Mostly everything in this image is a shade of dark gray. There are, however, several buildings with high intensity.

المنارة للاستشارات
www.manaraa.com

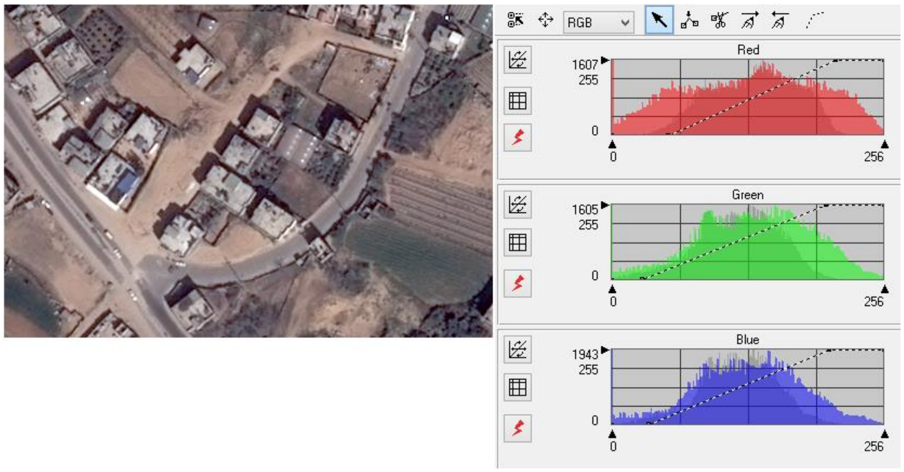


Figure 3-8: Histogram of study area sample (4)

3.2 Feature Extraction Methods

Feature Extraction is a combined process of segmenting an image into objects of pixels, computing attributes for each object, classifying the objects to classes and extract it, for more details see Appendix A.2.5.

Digitizing is a way of conversion of information from analogously produced graphical maps to machine readable vector or raster formats. Many methods are used for the vectorizing process and feature extraction [48]. Automated methods are adopting in this study to extract features from imagery based on object recognition. Figure 3-9 shows the methods and programs which have been used in this study.

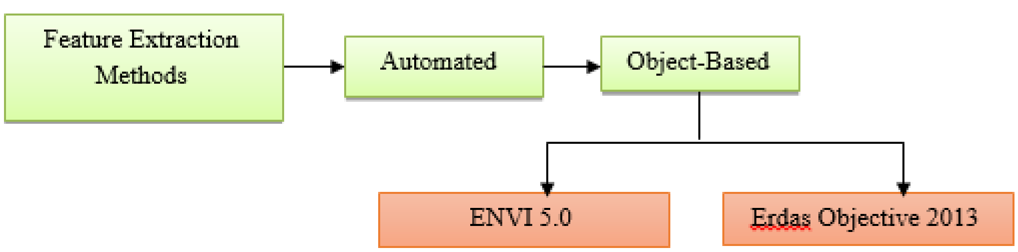


Figure 3-9: Feature extraction methods and programs

Commercial programs are introduced with new tools and developed new algorithms to extract feature from images such as (ERDAS Imagine 2013, ENVI 5.0, Feature analyst

المنارة للاستشارات
www.manaraa.com

5.2, and Feature extraction 11, FETEX 2.0). The processing that applied to the case study image using one programs (ENVI 5.0).

3.2.1 Feature Extraction Using ENVI 5.0

ENVI® (the Environment for Visualizing Images) is a revolutionary image processing system. From its inception, ENVI was designed to address the numerous, specific needs of those who regularly use satellite and aircraft remote sensing data.

ENVI feature extraction consists of a combined process of segmenting an image into objects of pixels, then computing attributes for each object. The workflow consists of two primary steps, find objects and extract features as shown in Figure 3-10. To find objects, the task is divided into four steps: segment images, merge segments, refine segments, and compute attributes. Once this task is completed, the feature extraction task can be performed. The feature extraction task consists of supervised or rule-based classification and exporting classification results to shape files and/ or raster images.

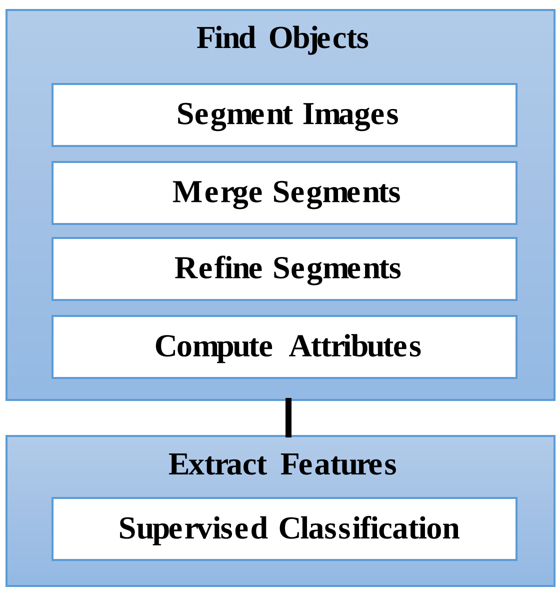


Figure 3-10: Feature extraction workflow of ENVI 5.0

In our experimentations, we use ENVI as a tool for Feature Extraction (the process Example Based Workflow under the category of Feature Extraction) as shown in Figure 3-11.

www.manaraa.com

3.2.1.1 Image Segmentation

Image segmentation is the primary technique used to convert a scene or image into multiple objects. Applying the object-based paradigm to image analysis refers to analyzing the image in object space rather than in pixel space, and objects can be used as the primitives for image classification rather than pixels, so image segmentation is the process of partition an image into segments by grouping neighboring pixels with similar feature values (brightness, texture, color, etc.).

Image segmentation can be performed automatically by employing an edge-based segmentation algorithm, which is very fast. It needs a familiar end user and only requires one input parameter (scale level). Adjust the scale level as necessary, values range from 0.0 (finest segmentation) to 100 (coarsest segmentation; all pixels are assigned to one segment).

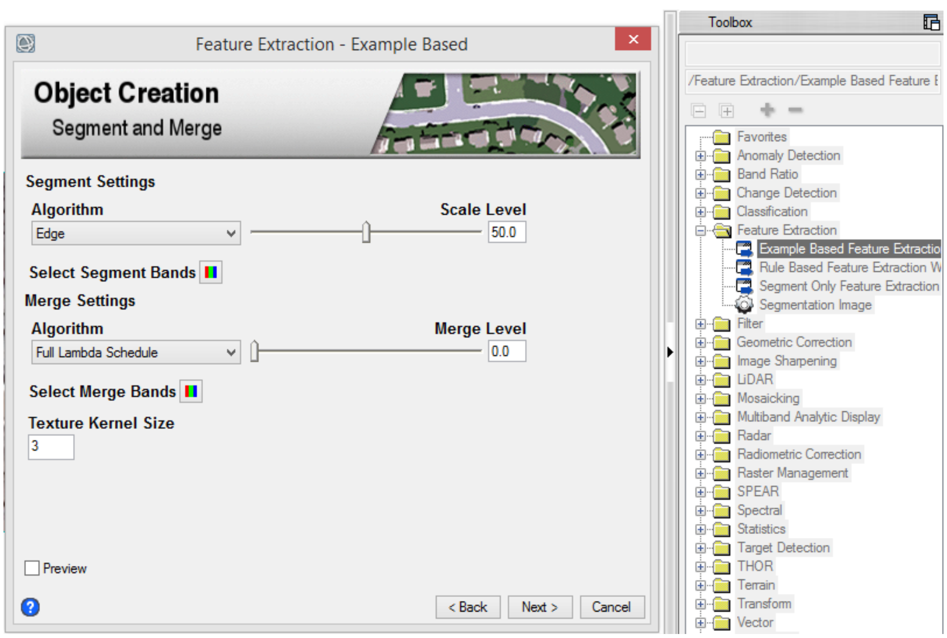


Figure 3-11: Object based feature extraction toolbox

Figure 3-12 shows boundary detection of (School building in Gaza) using edge-based segmentation algorithm at different levels of segmentation.

www.manaraa.com

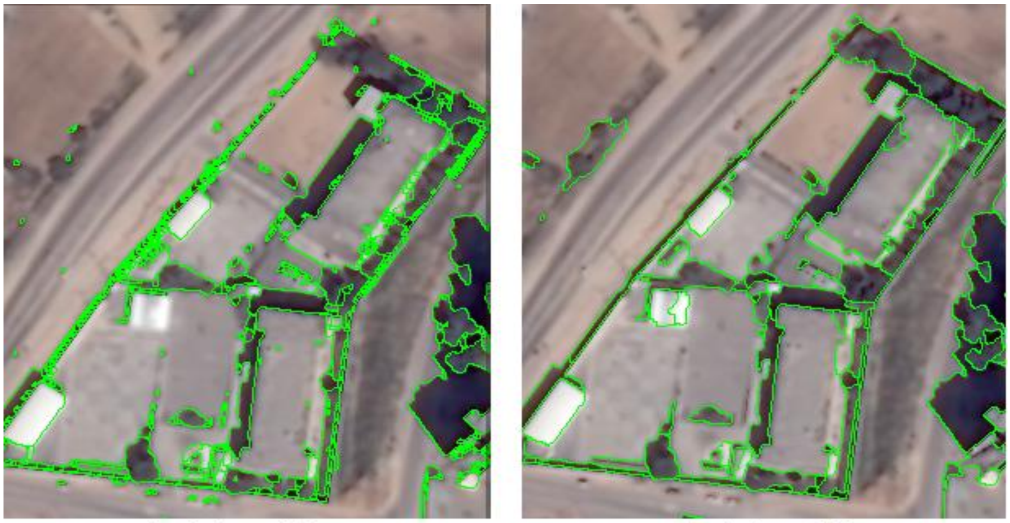
Figure 3-12: Image segmentation result at different levels

3.2.1.2 Merging Segments

Merging combines adjacent segments with similar spectral attributes. Some of features on the image are larger, textured areas such as trees and building. Merging Segments used to aggregate small segments within these areas where over-segmentation may have a problem. Scale level for merging is a useful option for improving the delineation of roads, buildings, and rivers boundaries, as it is clearly shown in Figure 3-13. To obtain better merging, there are some factors that may affect the quality of images.

- **Shadow:** In high spatial resolution satellite images, elevated objects such as buildings, bridges, trees and towers, especially in urban region, usually cast shadows. Shadows may cause loss of feature information, false color tone and shape distortion of objects, which seriously affect the quality of images.
- **Contrast:** Defined as the separation between the darkest and brightest areas of the image. Increase contrast, you increase the separation between dark and bright, making shadows darker and highlights brighter.
- **Texture:** Texture characteristics of the high-resolution satellite images, often used to describe texture are smooth (uniform, homogeneous), intermediate, and rough (coarse, heterogeneous).

www.manaraa.com



Scale Level 10 Scale Level 90

Figure 3-13: Merging segments result at different levels

By trial and error, we found that the best results as shown in Figure 3-14.

Figure 3-14: Optimal segmentation level 62 and merge level 90

After image segmentation and merging, a supervised classification will be performed using samples for the different classes (buildings, roads, and rivers). The classifier used is a K nearest neighborhood classifier that defines set of classes, which can be separated automatically. The K nearest distances are used as a majority vote to determine which class the target belongs to. The K Nearest Neighbor method is much less sensitive to outliers, noise in the dataset and generally produces a more accurate classification result

المنارة للاستشارات
www.manaraa.com

compared with traditional nearest-neighbor methods. Finally, we select the school building (as example) and identify the class name to export the features, as shown in Figure 3-15.

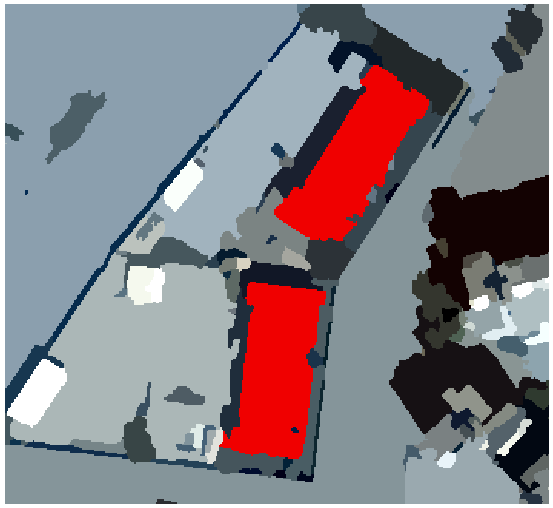


Figure 3-15: Building extraction and shape file exported

3.2.1.3 Objects Attributes

As mentioned before, we have three categories of features: spectral feature, spatial feature and texture feature. This will yield to 38 features for all categories with three bands for each feature in spectral and texture. The features attributes are divided to 12, 14, and 12 for spectral, spatial and texture respectively as shown in Table 3-1.

Table 3-1: List of object Attributes, Copyright 2014 by ENVI sofware

List of Attributes	
• **Spectral Attributes (12)**	
Attribute	**Description**
Spectral_Mean (in 3 band)	Mean value of the pixels comprising the region in band x
Spectral_Max (in 3 band)	Maximum value of the pixels comprising the region in band x
Spectral_Min (in 3 band)	Minimum value of the pixels comprising the region in band x

المنارة للاستشارات
www.manaraa.com

Spectral_STD (in 3 band)	Standard deviation value of the pixels comprising the region in band x
• **Texture Attributes (12)**	
Attribute	**Description**
Texture_Range (in 3 band)	Average data range of the pixels comprising the region inside the kernel (whose size you specify with the Texture Kernel Size parameter in segmentation)
Texture_Mean (in 3 band)	Average value of the pixels comprising the region inside the kernel
Texture_Variance (in 3 band)	Average variance of the pixels comprising the region inside the kernel
Texture_Entropy (in 3 band)	Average entropy value of the pixels comprising the region inside the kernel
• **Spatial Attributes (14)**	
Attribute	**Description**
Area	Total area of the polygon, minus the area of the holes. If the input image is pixel-based, the area is the number of pixels in the segmented object. For a segmented object with 20 x 20 pixels, the area is 400 pixels.
Length	The combined length of all boundaries of the polygon, including the boundaries of the holes. This is different than the Major_Length attribute. If the input image is pixel-based, the length is the number of pixels. For a segmented object with 20 x 20 pixels, the length is 80 pixels.
Compactness	A shape measure that indicates the compactness of the polygon. A circle is the most compact shape with a value of 1 / pi. The compactness value of a square is 1 / 2(sqrt (pi)).

المنارة للاستشارات
www.manaraa.com

Convexity	Polygons are either convex or concave. This attribute measures the convexity of the polygon. The convexity value for a convex polygon with no holes is 1.0, while the value for a concave polygon is less than 1.0.
Solidity	A shape measure that compares the area of the polygon to the area of a convex hull surrounding the polygon. The solidity value for a convex polygon with no holes is 1.0, and the value for a concave polygon is less than 1.0.
Roundness	A shape measure that compares the area of the polygon to the square of the maximum diameter of the polygon. The "maximum diameter" is the length of the major axis of an oriented bounding box enclosing the polygon. The roundness value for a circle is 1, and the value for a square is 4 / pi.
Form_Factor	A shape measure that compares the area of the polygon to the square of the total perimeter. The form factor value of a circle is 1, and the value of a square is pi / 4.
Elongation	A shape measure that indicates the ratio of the major axis of the polygon to the minor axis of the polygon. The major and minor axes are derived from an oriented bounding box containing the polygon. The elongation value for a square is 1.0, and the value for a rectangle is greater than 1.0.
Rectangular_Fit	A shape measure that indicates how well the shape is described by a rectangle. This attribute compares the area of the polygon to the area of the oriented bounding box enclosing the polygon. The rectangular fit value for a rectangle is 1.0, and the value for a non-rectangular shape is less than 1.0.
Main_Direction	The angle subtended by the major axis of the

المنارة للاستشارات
www.manaraa.com

	polygon and the x-axis in degrees. The main direction value ranges from 0 to 180 degrees. 90 degrees is North/South, and 0 to 180 degrees is East/West.
Major_Length	The length of the major axis of an oriented bounding box enclosing the polygon. Values are map units of the pixel size. If the image is not georeferenced, then pixel units are reported.
Minor_Length	The length of the minor axis of an oriented bounding box enclosing the polygon. Values are map units of the pixel size. If the image is not georeferenced, then pixel units are reported.
Number_of_Holes	The number of holes in the polygon. Integer value.
Hole_Area/Solid_Area	The ratio of the total area of the polygon to the area of the outer contour of the polygon. The whole solid ratio value for a polygon with no holes is 1.0.

3.3 Feature Selection

The main goal of feature selection is to reduce the dimensionality by eliminating irrelevant features and selecting the best discriminative features. Many search methods are proposed for feature selection [50, 61, 8, 53].

In our study, we use wrapper approach for feature selection. Wrapper methods evaluate subset of attributes based on their usefulness to a given classifier. Wrappers are conceptually very simple. To use this feature selection technique, one needs to decide: 1) how to search the space of all possible subsets of variables and how to halt it, 2) how to estimate the accuracy of the classifier used called by the wrapper, and 3) which classifier to use as a black box [19]. The accuracy of the classifier used as a black box is usually estimated using the holdout method or cross-validation.

Figure 3-16 illustrates the feature selection process. First, the data are splitted into training and testing sets. The train set is used in the feature selection while keeping the

المنارة للاستشارات
www.manaraa.com

test set only for the final evaluation of the performance of the induction algorithm. Then, the search is conducted using a chosen search method and by evaluating each candidate subset with respect to the performance of the learning algorithm. The performance is assessed usually either through cross-validation or using a validation set that is separate from the train and test sets. Once the terminating condition is met, the learning phase is conducted on the train set represented by the selected feature subset. Last, the output model is used to evaluate the test set.

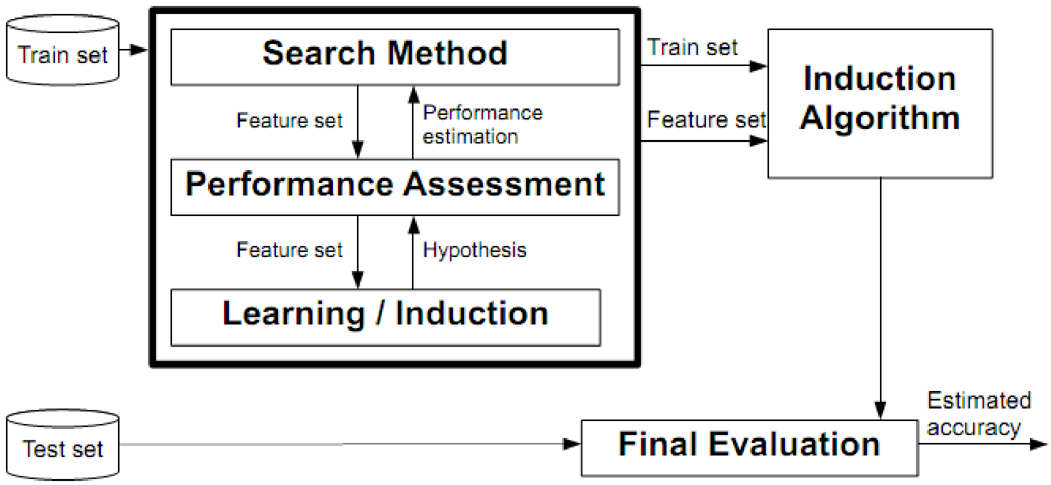


Figure 3-16: Feature Selection based on Wrapper method [29]

The experiments were conducted using the Experimenter tool in WEKA. WEKA is a collection of machine learning algorithms and data preprocessing tools written in Java and distributed under the terms of the General Public License (GNL). The software offers both graphical users interface for data processing and visualization as well as a possibility to use WEKA via scripts or Java code [58].

WEKA implements the wrapper selection by the function "WrapperSubsetEval". The function allows choosing the learning and search methods used in selection as well as whether to use cross validation (in this case, the number of folds can be chosen) or a separate validation set to assess the performance of the candidate subsets. WEKA offers implementations of a wide variety of learning and search methods used in the selection.

المنارة للاستشارات
www.manaraa.com

3.3.1 Feature Selection Optimization

We need to search the whole feature space to find the optimal subset of features. If the feature set contains N features, the number of possible subsets is 2^N. This makes the problem NP-hard and an exhaustive search method, which involves searching through all possible subsets, becomes prohibitively expensive as the number of features increases. Therefore, a method using random subset generation would be the most proper approach that is genetic search algorithm [56]. Although the search space with these methods is O (2N), in practice the space is reduced by defining the maximum number of iterations.

3.3.1.1 Genetic Search

The overall architecture of our wrapper approach based on GA is given in Figure 3-17.

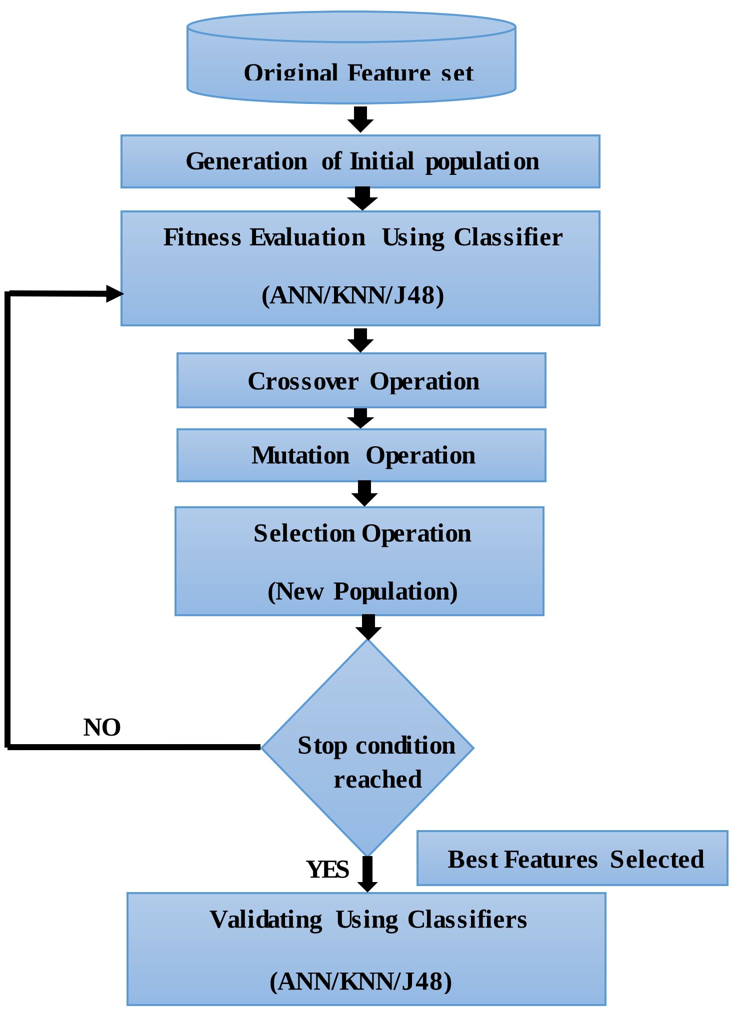


Figure 3-17: Flowchart of our wrapper method based on GA and classifier for evaluation

www.manaraa.com

Based on the previous steps and after feature extraction stage, our target is to find out and select the optimum minimized feature set that will make the classification even better than when the full set of features are used. As indicated, GAs are used to explore the space of all subsets of a given feature set. Each of the selected feature subsets is evaluated (its fitness measured based on accuracy) by invoking classifiers.

The first step in applying GAs to the problem of feature selection is to map the search space into a representation suitable for genetic search. Since we are only interested in representing the space of all possible subsets of the given feature set, the simplest form of representation is to consider each feature in the candidate feature set as a binary gene "0" or "1".

Then, each individual consists of fixed-length binary string representing some subset of the given feature set. An individual of length 'n' corresponds to an n-dimensional binary feature vector 'F', where each bit represents the elimination or inclusion of the associated feature. For example, Fi=0 represents elimination and Fi=1 indicates inclusion of the ith feature, as shown in Figure 3-18. Hence, a feature set with five features can be represented as <F1 F2 F3 F4 F5>. Then, an individual of the form <11111> indicates inclusion of all the features, and <11010> represents the subset where the third and the fifth features are eliminated.

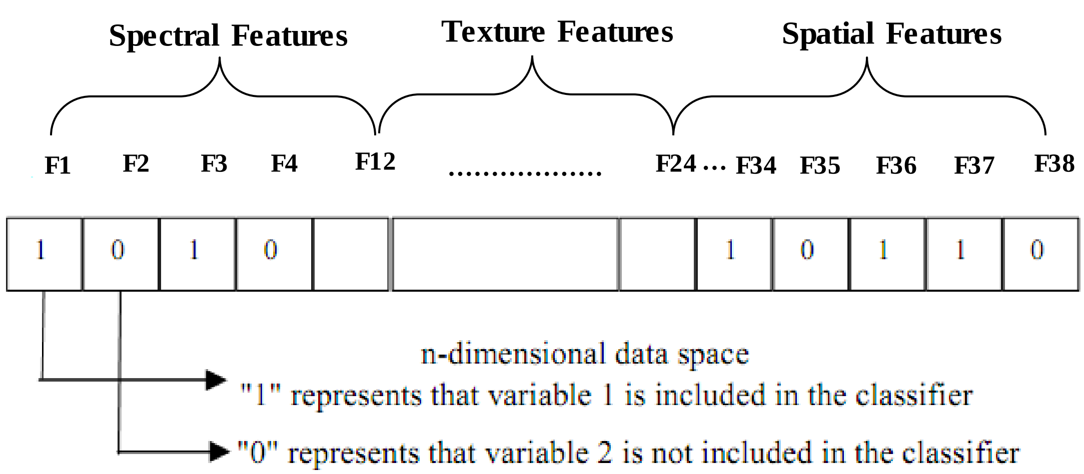


Figure 3-18: Encoding of features into a n-bit chromosome string

المنارة للاستشارات
www.manaraa.com

Once the fitness values of all individuals of the current population have been computed, the GA begins to generate next generation as follows [34, 20]:

- **Crossover:** See section 1.3.1, a crossover operator selects a crossover point randomly then interchanges bit-string of parents at this point to produce two new offsprings. If we cannot perform crossover, offspring will be the exact copy of parents.

  Crossover is made in hope that new chromosomes will have good parts of old chromosomes and maybe the new chromosomes will be better. However, it is good to leave some part of population, survive to next generation. As shown in Figure 3-19, one-point crossover is performed between parent A and parent B, and produced two offsprings C and D.

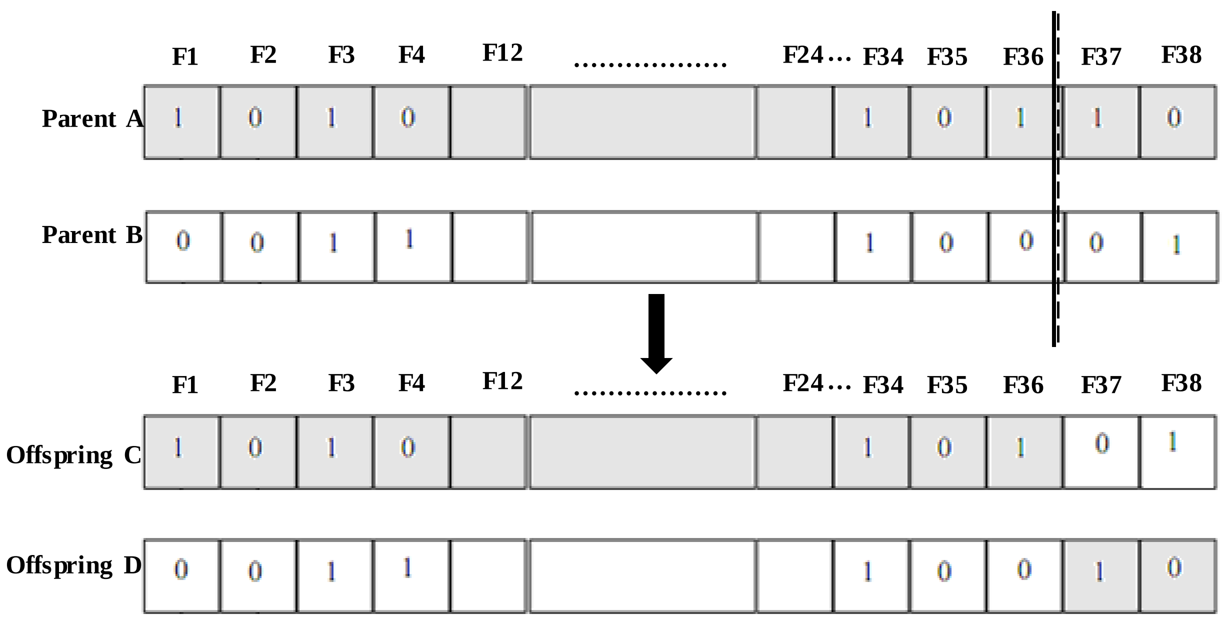


Figure 3-19: Bit-String Crossover of Parents A & B to new Offspring C & D

- **Mutation:** See section 1.3.1, if we cannot perform mutation, offspring will take after crossover without any change. Flip Bit is a mutation operator that alters the value of the chosen gene (0 turn into 1 and 1 turn into 0). This mutation operator can only be used for binary genes. As shown in Figure 3-20, value of F4 in Spectral features is changed from 0 to 1.

www.manaraa.com

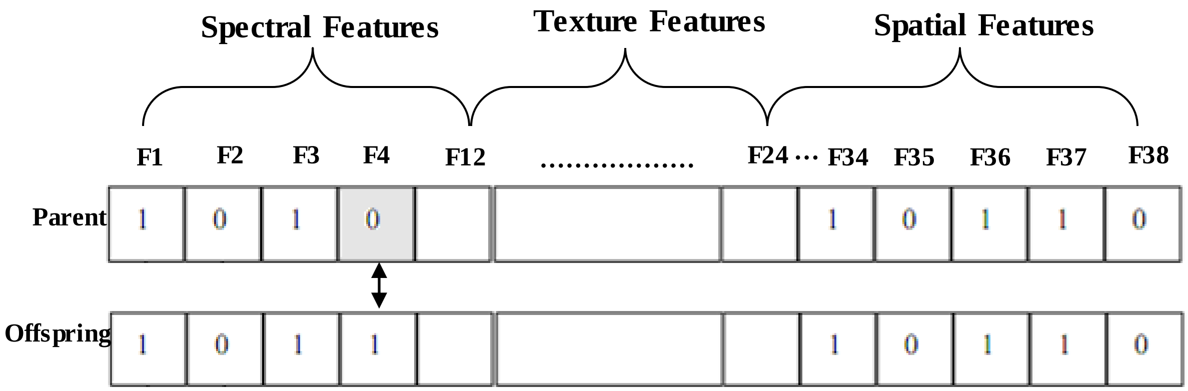


Figure 3-20: Bit-Flipping Mutation of Parent to new Offspring

The procedure above is iteratively executed until the maximum number of generations is reached. The advantage to this representation is that the classical GA's operators as described before (binary mutation and crossover) can easily be applied to this representation without any modification. This eliminates the need for designing new genetic operators, or making any other changes to the standard form of genetic algorithms.

Choosing an appropriate evaluation function is an essential step for successful application of GAs to any problem domain. As before, the process of evaluation involved the steps presented in Figure 3-17. Evaluation functions provide GAs with the feedback about the fitness of each individual in the population. GAs then use this feedback to bias the search process to provide an improvement in the population's average fitness.

We use three families of classification algorithms as a basis for comparisons. These are the neural network, decision-tree J48 and k-Nearest Neighbors. Classifiers that were used are well known in the machine learning community and represent three completely different approaches to learning, hence we hope that our results are of a general nature and will generalize to other classification algorithms.

3.3.2 Classification Algorithms

Wrapper methods evaluate subset of attributes based on their usefulness to a given classifier. It was required that the used classifiers had been effective and widely used in

www.manaraa.com

the previous studies in the field. Thus, three classifiers are chosen: Neural Network, k-Nearest Neighbors, and J48 Decision Tree.

3.3.2.1 Artificial Neural Network (ANN)

The basic architecture of an artificial neural network is shown in Figure 3-21. Each circle in input layer represents an objects attribute (Features), where each circle in output layer represents output class such as Roads, buildings and rivers. This network topology is determining by the user and is based on the type and complexity of the problem space.

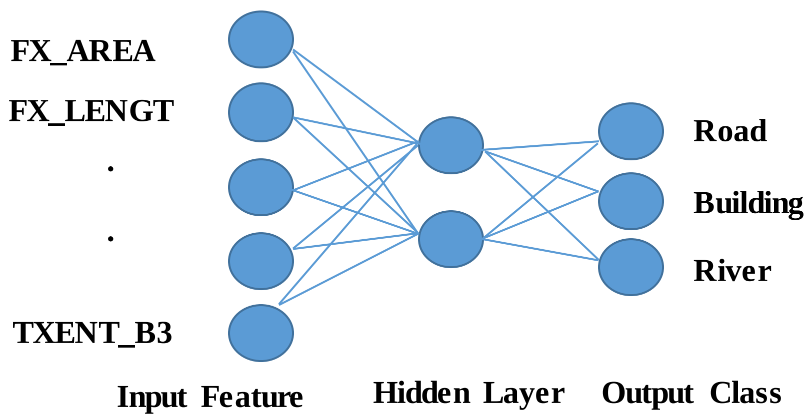


Figure 3-21: Basic architecture of an artificial neural network. Input neurons represent object feature and output layer represent object class

The size and training parameters of artificial neural networks have a critical effect on their performance. Building of a back-propagation network involves the specification of the number of hidden layers, number of learning cycles (Epoch) and learning rate. Thus, we perform multiple training runs to obtain the best ANN model parameters. In our experiments, after several attempts we choose Learning Rate = 0.1, Hidden Layer = 11 and number of Epoch (range between 400 and 500).

In our experimentations, we use WEKA as a tool for ANN classifier (the classifier **"MultilayerPerceptron"** under the category of functions).

3.3.2.2 K-Nearest Neighbors (KNN)

A new object is classified by a majority vote of its neighbors. The new object is assigned to the class most common among its K nearest neighbors measured by a distance

www.manaraa.com

function, as shown in Figure 3-22. If K = 6, then the object is simply assigned to the class of its nearest neighbor.

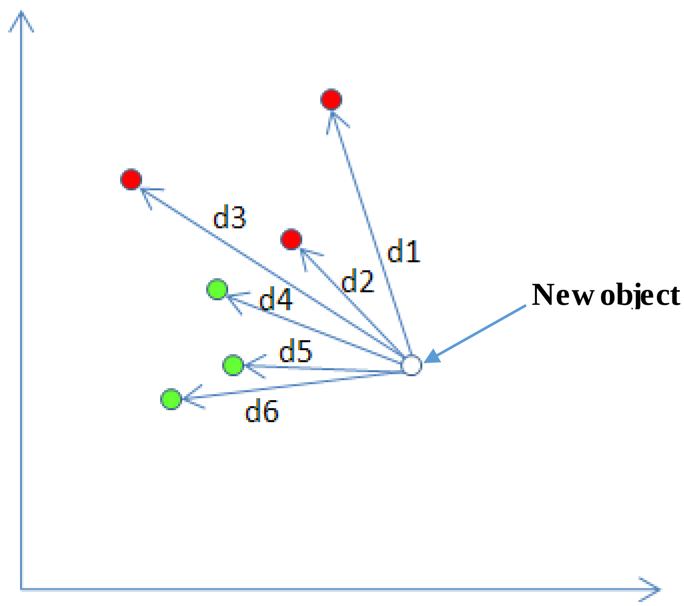


Figure 3-22: K nearest neighbors measured by a distance function

Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee [54].

In our experiments, we choose K value randomly ranging from 1 to 15 by an increment of 1, we found that the best value of k = 8.

For our experiments, we use WEKA as a tool (the classifier IBK under the category of lazy learners) with Euclidean distance as a similarity measure.

3.3.2.3 J48 Decision tree

Decision tree is one of the inductive learning algorithms that generate a classification tree to classify the data. Decision tree is based on the "divide and conquer" strategy.

The basic architecture of a decision tree is depicted in Figure 3-23. Each node represents an objects attribute (Features) with a decision rule or a class such as Road, building and River.

www.manaraa.com

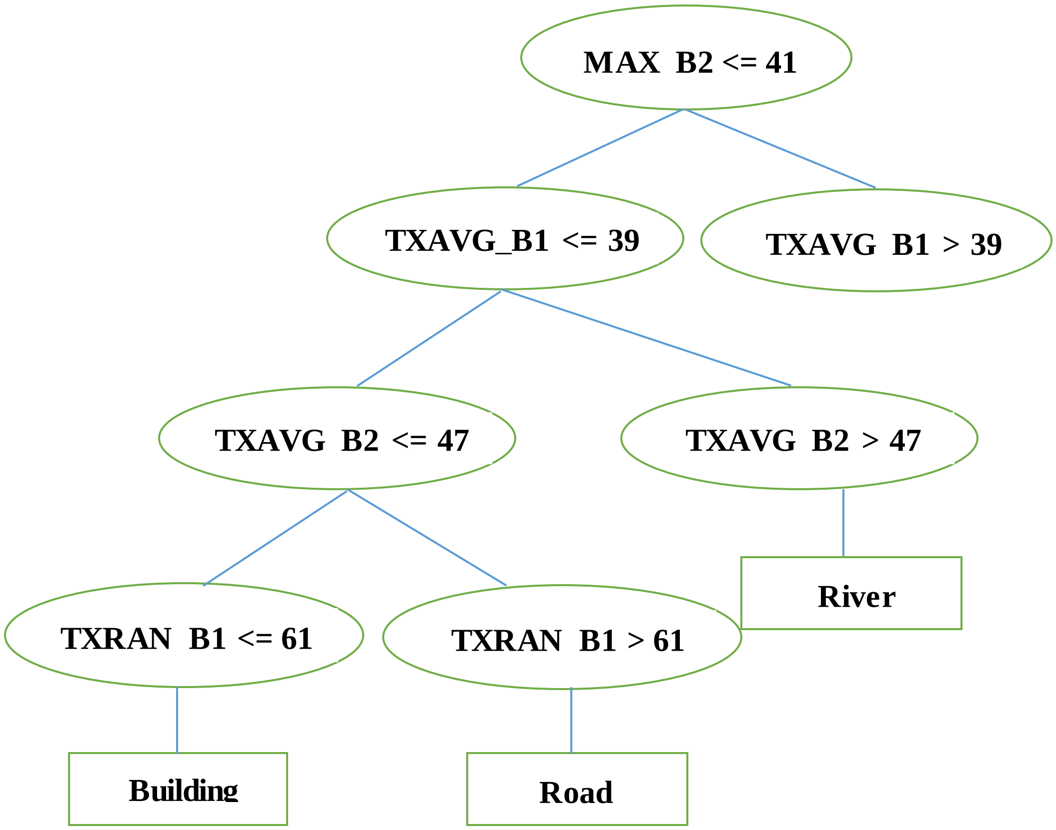


Figure 3-23: Example of decision tree using J48 classifier

To prune our decision trees, we use post-pruning that labeled by WEKA as the confidence factor. In the WEKA J48 classifier, lowering the confidence factor decreases the amount of post-pruning. We tested the J48 classifier with confidence factor ranging from 0.01 to 1.0 by an increment of 0.1 and cross validation folds for the testing set was held at 10 during confidence factor testing.

In our experiments, we focus on J48. Decision tree J48 is the implementation of algorithm C4.5 developed by the WEKA project team [7].

المنارة للاستشارات
www.manaraa.com

CHAPTER 4: Experimentation and Results

In this chapter, we shall present our experiments on our approach for selecting the subset of features from satellite imagery.

4.1 Experimental Environment and Tools

All experiments are implemented on a dell server of Intel Xeon(R) Processing power of 2.40 GHz CPU with 16GB RAM. The following are the used tools:

- **WEKA:** we use WEKA for our experimentation (GA-ANN, GA-KNN, GA-J48).
- **ENVI:** The software used to process the satellite imagery including image segmentation and feature extraction.
- **ERDAS Imagine:** This software is used for preprocessing the satellite imagery.
- **ARCGIS:** This software is used to open sahpefile and export the features.
- **Microsoft Word:** the program is used for document typing.
- **Microsoft Excel:** we use excel to partition, organize and store datasets in tables. In addition, it is used for some simple preprocessing and analyzing the results.

4.2 Dataset

We download our images from different sources with high-resolution and 3-band spectral imagery; we have 15 satellite imagery contains (roads, building, rivers) for training and 10 imagery for testing. We process these images using ENVI and ERDAS software to perform feature extraction described in section 3.2. Table 4-1 and Table 4-2 show the dataset structure and the extracted features.

Table 4-1: The experiments are done with two datasets.

Dataset	# of Roads	# of Building	# of Rivers	Totals
Training dataset	1317	1288	1481	4087
Testing dataset	658	644	740	2043

المنارة للاستشارات
www.manaraa.com

Table 4-2: List of features extracted form ENVI software

List of Features		Data Type	# of Features
Spectral Features	AVG_B1, STD_B1, MAX_B1, MIN_B1, AVG_B2, STD_B2, MAX_B2, MIN_B2, AVG_B3, STD_B3, MAX_B3, MIN_B3	Numerical	12
Texture Features	TXRAN_B1, TXAVG_B1, TXVAR_B1, TXENT_B1, TXRAN_B2, TXAVG_B2, TXVAR_B2, TXENT_B2, TXRAN_B3, TXAVG_B3, TXVAR_B3, TXENT_B3	Numerical	12
Spatial Features	FX_AREA, FX_LENGTH, FX_COMPACT, FX_CONVEX, FX_SOLID, FX_ROUND, FX_FORMFAC, FX_ELONG, FX_RECT_FI, FX_MAIN_DI, FX_MAJAXLN, FX_MINAXLN, FX_NUMHOLE, FX_HOLESOL	Numerical	14

4.3 Feature Selection Based Wrapper Method

The experiments in this context involve running the wrapper method with chosen classifier algorithm and search algorithm. As mentioned before we use GA as a randomized feature selection and choose three classifiers (ANN, KNN, Decision trees J48) to be classifiers for our experiments. To test our system, we use the "WrapperSubsetEval" function in WEKA. The function allows choosing the classifier and search method used in selection. Thus, we have five experiments in this context using GA with every classifier alone, correlation ranking filter for spatial features and optimal features subsets validation, as shown in Table 4-3.

GA will return an optimum subset of features and then the classifier will evaluate the obtained subset. The basic idea is to compare the accuracy of a classifier on the original dataset having the complete set of features with the newly obtained dataset containing

www.manaraa.com

only the subset of features returned by the feature selection method. This procedure will allow us to evaluate the importance of the obtained subset and its effect on the classifier.

In our experiments, we use 10-fold cross-validation, by which the data set is divided into 10 subsets, one of them used as a test and the rest is used for training.

As mentioned before we use GA as a randomized feature selection method, which prevents falling in local minima. We use the following parameters:

- **The population size (P):** This is the number of chromosomes in each generation, where each chromosome is an individual of randomly generated 38 features.
- **Max_Generations:** Positive integer specifying the maximum number of iterations before the algorithm halts.
- **Crossover Probability:** Crossover randomly selects a point within the strings representing the parents and swaps all the bits after that point between the two, section 3.3.1.1 introduced more details.
- **Mutation Probability:** Mutation randomly changes one bit or more of an individual to introduce perturbation in the population, section 3.3.1.1 introduced more details.

Table 4-3: List of the five main experiments

Experiment 1	GA-ANN
Experiment 2	GA-KNN
Experiment 3	GA-J48
Experiment 4	Correlation Ranking Filter for Spatial Features
Experiment 5	Optimal features subsets validation

المنارة للاستشارات
www.manaraa.com

We implement the classifier for five times, each time we use the feature category alone as shown in Table 4-4.

Table 4-4: Experiments for evaluation based on features categories using (ANN,KNN,J48)

EXP. 1	Spatial Features (# of features is 14)
EXP. 2	Spectral Features (# of features is 12)
EXP. 3	Texture Features (# of features is 12)
EXP. 4	Spectral and Texture (# of features is 24)
EXP. 5	All Features (# of features is 38)

4.3.1 Experiment 1: GA-ANN

In the first experiment, we use the Feed-Forward ANN with back-propagation, which is one of the most popular techniques, as a classifier. Section 1.3.2.1 introduced more detail about ANN.

We use three levels to represent the forward neural network, the input layer with a number of neurons equal to the number of selected features, the output layer with a number of three nodes to represent the target classes "Road, Building and River", and a hidden layer. We tried to solve this problem to get a better estimate of the performance by 10-fold cross validation.

Table 4-5 shows the obtained classification results with the best ANN parameters.

Table 4-5: Classification accuracy based on Features categories using ANN

No.	Input Features	# of Features	ANN parameter	Accuracy	Time (Seconds)
1	Spatial	14	LR = 0.1,Epochs = 400,Hidden Layer (HL) = 11	45.11%	17.86s

www.manaraa.com

2	Spectral	12	LR = 0.1,Epochs = 400, Hidden Layer (HL) = 11	84.46%	16.41s
3	Texture	12	LR = 0.1,Epochs = 400, Hidden Layer (HL) = 11	86.25%	18.47s
4	Spectral and Texture	24	LR = 0.1,Epochs = 400, Hidden Layer (HL) = 11	87.49%	18.47s
5	All Features	38	LR = 0.1,Epochs = 450, Hidden Layer (HL) = 11	88.37%	40.65s

From Table 4-5 and Figure 4-1, it is clear that considering all features is having the highest accuracy with 88.37%. It is also to be noted that the texture features are more important than spatial and spectral features despite they are 12 features. In addition, it is to be noted that the less important features are the spatial features despite they are 14 features.

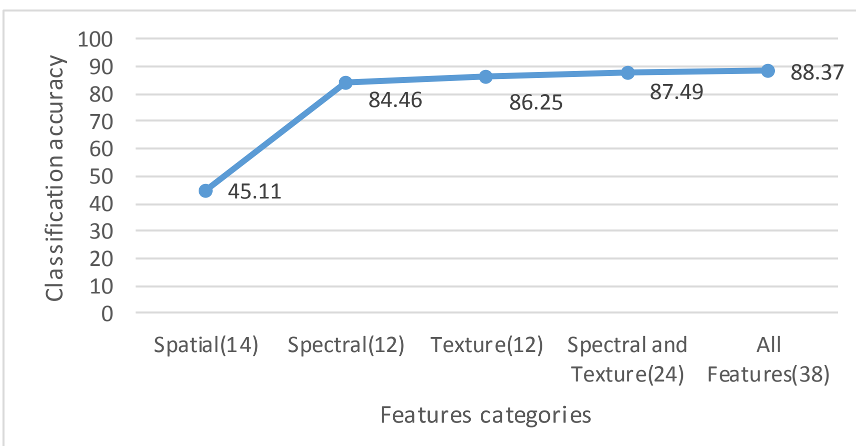


Figure 4-1: Classification accuracy based on Features categories using ANN

www.manaraa.com

Now we use GA for Feature Selection (FS) to select the best subset of features for 38 features. By trial, we found the best parameter for GA-ANN are as in Table 4-6 and Table 4-7.

Table 4-6: Best parameter for GA-ANN

Genetic Algorithm (GA)	
MAX GENERATION	180
POPULATION SIZE	40
CROSSOVER PROBABILITY	0.6
MUTATION PROBABILITY	0.033

Table 4-7: Best parameter for ANN

Artificial Neural Network (ANN)	
HIDDEN LAYERS	11
LEARNING RATE	0.1
NUMBER OF EPOCHS	450

The wrapper using ANN and employing GA returned a subset of only **17** features, as shown in Table 4-8.

Table 4-8: Optimal subsets returned by wrapper employing GA-ANN

List of Attributes		**# of Features**
Spectral Features	AVG_B1, STD_B1, AVG_B2, STD_B2, MIN_B2, AVG_B3, STD_B3	7
Texture Features	TXRAN_B1, TXAVG_B1, TXAVG_B2, TXVAR_B2, TXAVG_B3, TXVAR_B3, TXENT_B3	7

المنارة للاستشارات
www.manaraa.com

Spatial Features	FX_FORMFAC, FX_RECT_FI, FX_MINAXLN	3

The time taken to find the "optimal" subset with GA-ANN was nearly 48 hours with the overall classification accuracy **89.70%**. It is to be noted that wrapper method improved the accuracy of ANN when genetic algorithm is used. The estimated accuracy averaged over the runs is 1.32% higher than the accuracy when all features are considered in the dataset.

In addition, GA-ANN reduces number of features with 55% (from 38 to 17). This is useful in reducing data dimensionality. The obtained results shown in Table 4-7 confirms the results obtained in Table 4-4. It is clear that GA only selects 3 features out of the 14 spatial features, which means that the spatial features are the least important features. In addition, it is obvious that the texture features are the most important features where 7 features are selected out of 12. This again confirms the results shown in Table 4-4.

4.3.1.1 Training dataset

Table 4-9 and Figure 4-2 illustrate experimental results for training dataset before feature selection and after feature selection. Through the results, we note that the wrapper method based on GA-ANN for feature selection is very useful to reduce data dimensionality, improve classification accuracy and reduce estimation time for classification.

Table 4-9: The results of classification accuracy and estimation time before and after using GA-ANN on training dataset

	# of Features	Accuracy	Time (Seconds)
Before FS	38	88.37%	40.65s
After FS	17	89.70%	23.8s

المنارة للاستشارات
www.manaraa.com

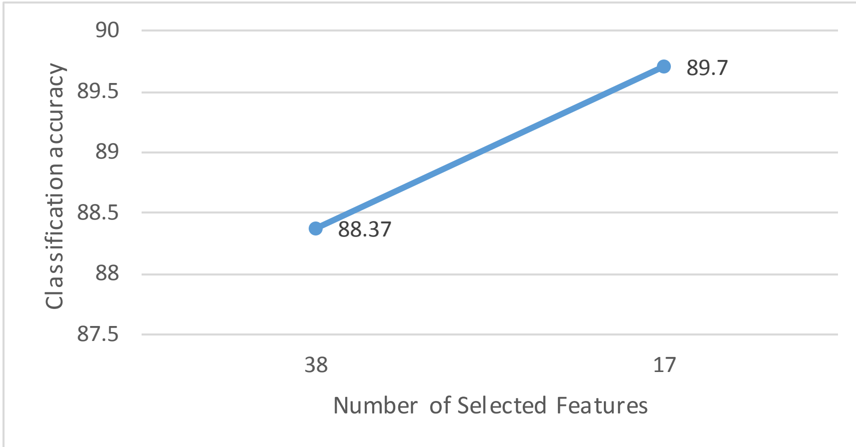


Figure 4-2: The results of classification accuracy and estimation time before and after using GA-ANN on training dataset

4.3.1.2 Testing dataset

After training, we test the optimal features subset using different dataset. As shown in Table 4-10 and Figure 4-3.

Table 4-10: The results of classification accuracy and estimation time before and after using GA-ANN on testing dataset

	# of Features	Accuracy	Time (Seconds)
Before FS	38	88.23%	36.46s
After FS	17	89.43%	21.01s

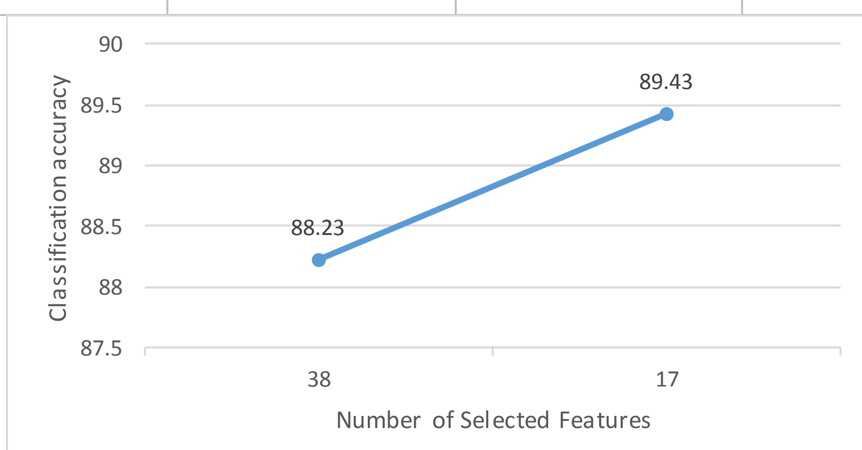


Figure 4-3: The results of classification accuracy and estimation time before and after using GA-ANN on testing dataset

المنارة للاستشارات
www.manaraa.com

4.3.2 Experiment 2: GA-KNN

In the second experiment, we use the K-Nearest Neighbors, which is one of the simplest techniques as a classifier. Section 1.3.2.2 introduced more details about KNN. Table 4-11 and Figure 4-4 shows the obtained classification results with the best k parameter.

Table 4-11: Classification accuracy based on features categories using KNN

No.	Input Features	# of Features	K-NN parameter	Accuracy	Time (Seconds)
1	Spatial	14	K = 20	44.84%	1s
2	Spectral	12	K = 8	83.36%	1s
3	Texture	12	K = 8	85.71%	1s
4	Spectral and Texture	24	K = 10	86.32%	1s
5	All Features	38	K = 8	83.68%	3s

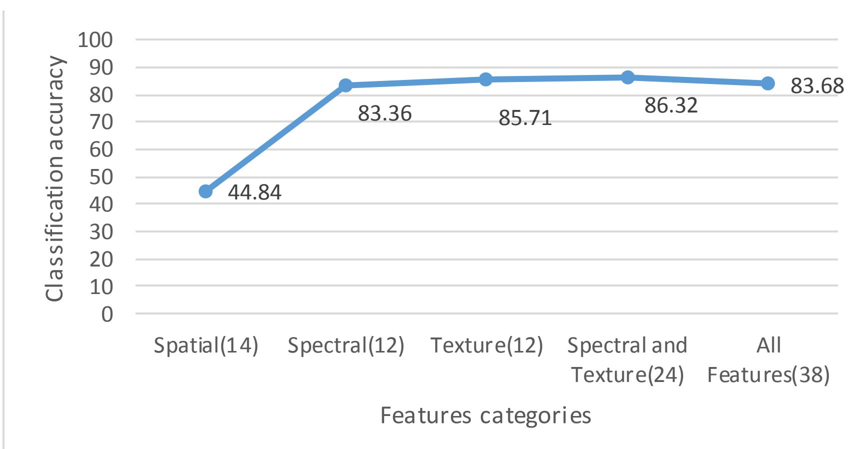


Figure 4-4: Classification accuracy based on features categories using KNN

The results presented in Table 4-11 confirm the results shown in Table 4-5 in which the spatial features are having the less importance but the mixed features of spectral and texture are more important than texture features alone.

المنارة للاستشارات
www.manaraa.com

The wrapper using KNN and employing GA returns a subset of only **14** features as shown in Table 4-12 and Table 4-13.

Table 4-12: Best parameter for GA-KNN

Genetic Algorithm (GA)	
MAX GENERATION	180
POPULATION SIZE	40
CROSSOVER PROBABILITY	0.6
MUTATION PROBABILITY	0.033

Table 4-13: Optimal subsets returned by wrapper employing GA-KNN

List of Attributes		**# of Features**
Spectral Features	AVG_B1, AVG_B2, MAX_B2, MIN_B2, AVG_B3	5
Texture Features	TXAVG_B1, TXVAR_B1, TXAVG_B2, TXVAR_B2, TXENT_B2, TXAVG_B3, TXVAR_B3, TXENT_B3	8
Spatial Features	FX_CONVEX	1

The time taken to find the "optimal" subset with GA-KNN was nearly 9 hours with the overall classification accuracy 87.49%. The accuracy with GA-KNN is becoming higher than the accuracy when all features are considered with a percentage of 3.81% on an average. In addition, GA- KNN reduces the number of features with 63% at least (from 38 to 14). Results shown in Table 4-13 confirm the results obtained in Table 4-8, which shows that spatial features have the least effect whereas the texture features are having the highest effect.

المنارة للاستشارات
www.manaraa.com

4.3.2.1 Training dataset

Table 4-14 and Figure 4-5 illustrate experimental results for training dataset before feature selection and after feature selection. Through the results, we note that the wrapper method based on GA-KNN for feature selection is very useful to reduce data dimensionality, improve classification accuracy and reduce estimation time for classification.

Table 4-14: The results of classification accuracy and estimation time before and after using GA-KNN on training dataset

	# of Features	Accuracy	Time (Seconds)
Before FS	38	83.68%	3s
After FS	14	87.49%	1s

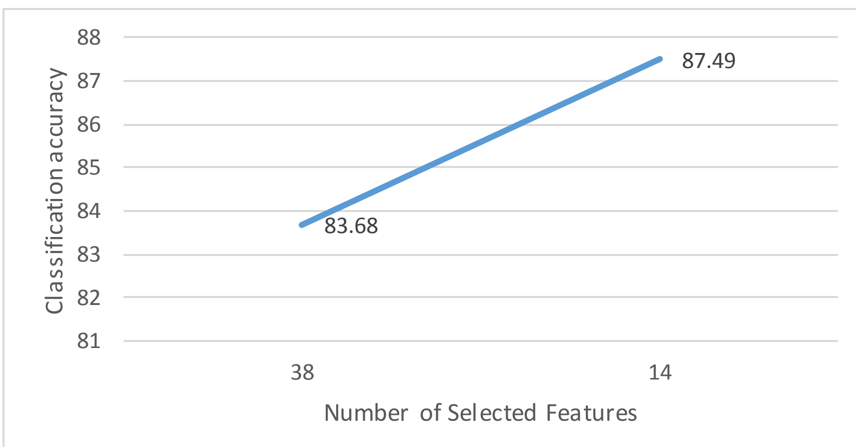


Figure 4-5: The results of classification accuracy and estimation time before and after using GA-KNN on training dataset

4.3.2.2 Testing dataset

After training, we test the optimal features subset using different dataset. As shown in Table 4-15 and Figure 4-6.

المنارة للاستشارات
www.manaraa.com

Table 4-15: The results of classification accuracy and estimation time before and after using GA-KNN on testing dataset

	# of Features	Accuracy	Time (Seconds)
Before FS	38	81.84%	1s
After FS	14	87.35%	0.5s

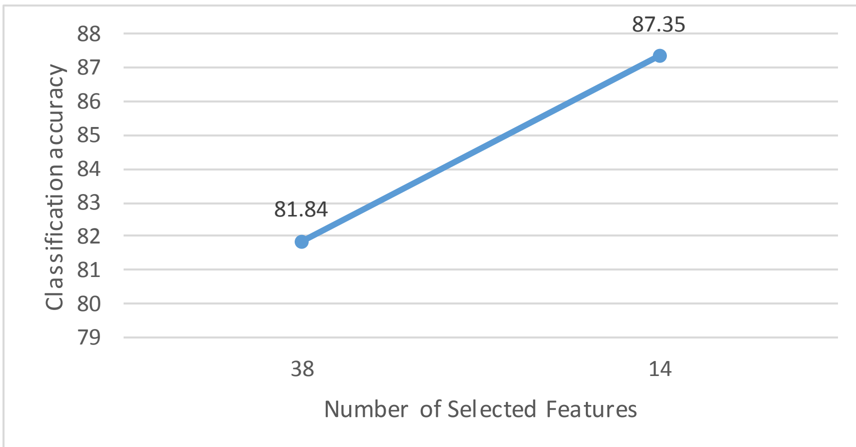


Figure 4-6: The results of classification accuracy and estimation time before and after using GA-KNN on testing dataset

4.3.3 Experiment 3: GA-J48

In the third experiment, we use the Decision tree J48, which is one of the famous classification techniques. Section 1.3.2.3 introduced more details about J48. Table 4-16 and Figure 4-7 show the obtained classification results with the best Confidence Factor (CF) parameter.

Table 4-16: Classification accuracy based on features categories using J48

No.	Input Features	# of Features	J48 parameter	Accuracy	Time (Seconds)
1	Spatial	14	CF = 0.05	52.50%	0.49s
2	Spectral	12	CF = 0.05	78.73%	0.45s

المنارة للاستشارات
www.manaraa.com

3	Texture	12	CF = 0.05	82.55%	0.45s
4	Spectral and Texture	24	CF = 0.05	80.54%	1.01s
5	All Features	38	CF = 0.05	82.85%	1.55s

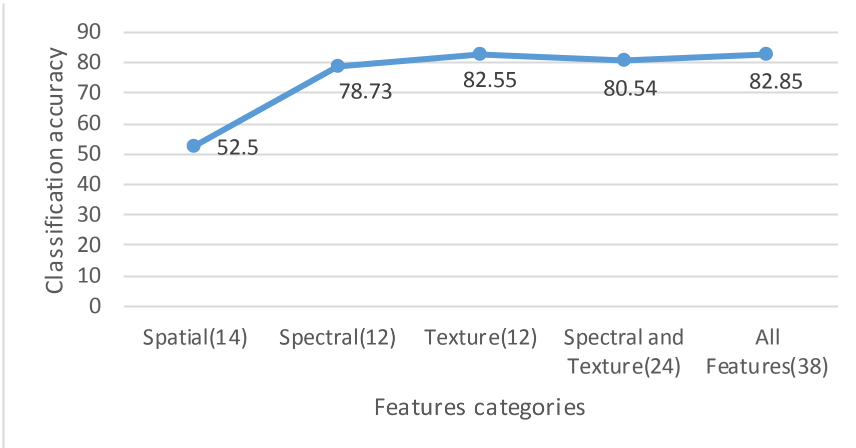


Figure 4-7: Classification accuracy based on features categories using J48

The results shown in Table 4-16 confirm the results shown in Table 4-5 and Table 4-11, in which the spatial features are having the less importance but the texture features are more important than mix of spectral and texture features.

The wrapper using J48 and employing a GA returned a subset of only **16** features as shown in Table 4-17 and Table4-18.

Table 4-17: Best parameter for GA-J48

Genetic Algorithm (GA)	
MAX GENERATION	180
POPULATION SIZE	40
CROSSOVER PROBABILITY	0.6
MUTATION PROBABILITY	0.033

www.manaraa.com

Table 4-18: Optimal subsets returned by wrapper employing GA-J48

List of Attributes		# of Features
Spectral Features	AVG_B1, AVG_B2, AVG_B3, MIN_B3	4
Texture Features	TXAVG_B1, TXRAN_B2, TXVAR_B2, TXRAN_B3, TXAVG_B3, TXVAR_B3, TXENT_B3	7
Spatial Features	FX_LENGTH, FX_SOLID, FX_RECT_FI, FX_MINAXLN, FX_HOLESOL	5

The time taken to find the "optimal" subset with GA-J48 was nearly 11 hours with the overall classification accuracy 85.24%. The accuracy with GA-J48 is becoming higher than the accuracy when all features are considered with a percentage of 2.39% on an average. In addition, GA-J48 reduces number of features with 57% (from 38 to 16).

On the contrary, of previous results, results shown in Table 4-17 does not confirm results obtained in Table 4-13 and Table 4-8, although the number of spatial features isn't the highest in the optimal subset, they still have an important effect (5 out of 16 selected features).

4.3.3.1 Training dataset

Table 4-19 and Figure 4-8 illustrate experimental results for training dataset before feature selection and after feature selection. Through the results, we note that the wrapper method based on GA-J48 for feature selection is very useful to reduce data dimensionality, improve classification accuracy and reduce estimation time for classification.

Table 4-19: The results of classification accuracy and estimation time before and after using GA-J48 on training dataset

	# of Features	Accuracy	Time (Seconds)
Before FS	38	82.85%	1.55s
After FS	16	85.24%	0.67s

المنارة للاستشارات
www.manaraa.com

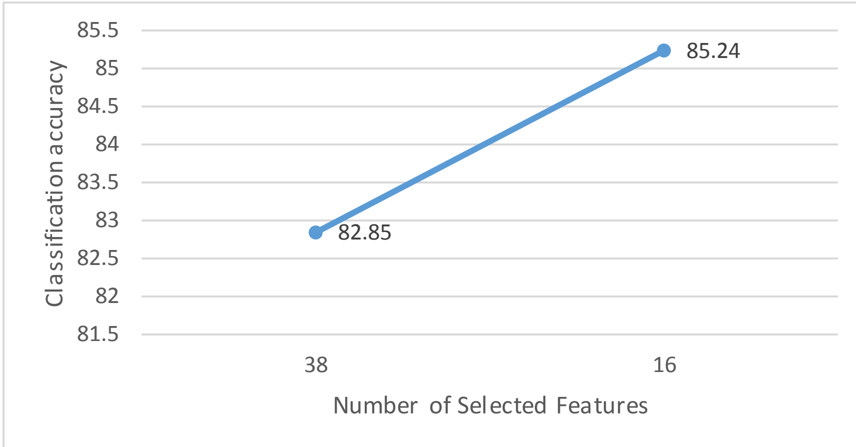


Figure 4-8: The results of classification accuracy and estimation time before and after using GA-J48 on training dataset

4.3.3.2 Testing dataset

After training, we test the optimal features subset using different dataset. As shown in Table 4-20 and Figure 4-9.

Table 4-20: The results of classification accuracy and estimation time before and after using GA-J48 on testing dataset

	# of Features	Accuracy	Time (Seconds)
Before FS	38	79.54%	0.71s
After FS	16	81.16%	0.24s

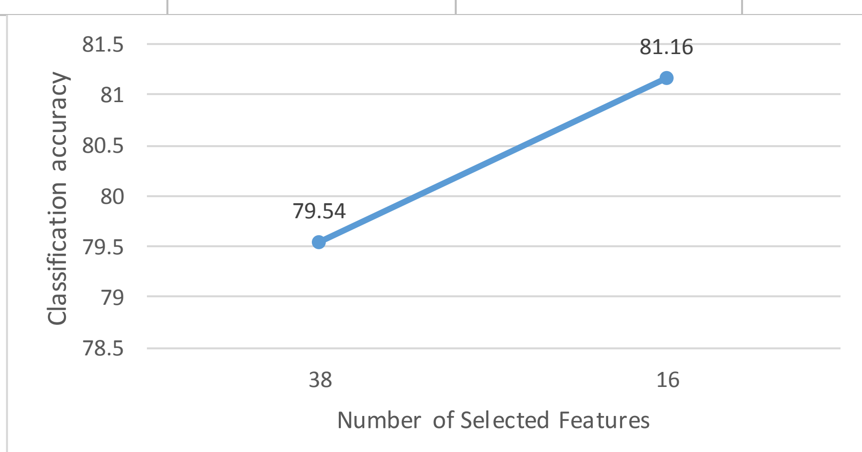


Figure 4-9: The results of classification accuracy and estimation time before and after using GA-J48 on testing dataset

المنارة للاستشارات
www.manaraa.com

4.3.4 Experiment 4: Correlation Ranking Filter for Spatial Features

The obtained results in Table 4-5, Table 4-11 and Table 4-16 show that the spatial features are the least important features, and it was reflected on optimal subsets features. As shown in Table 4-8, Table 4-13 and Table 4-18, spatial selected features are the least among of other selected features. It is also noted from the results of the previous experiments that only 7 spatial features out of the 14 features are having the highest effect in optimal subset as shown in Table 4-21. We propose to use the Correlation Ranking Filter (CRF) for measuring the correlation between spatial features and target classes to reduce the number of spatial features. As shown in Table 4-21, we found only 6 spatial features having the highest correlated filter. The results show the spatial features which have been selected with CRF are the same as those selected in optimal subsets except the last one (FX_LENGTH).

Table 4-21: Spatial Features which Selected in optimal subsets and Correlation Ranking Filter

#	Spatial Features which Selected in Optimal Subsets	Spatial Features which Selected in CRF
1	FX_RECT_FI	FX_RECT_FI
2	FX_FORMFAC	FX_FORMFAC
3	FX_MINAXLN	FX_MINAXLN
4	FX_CONVEX	FX_CONVEX
5	FX_SOLID	FX_SOLID
6	FX_HOLESOL	FX_HOLESOL
7	FX_LENGTH	

As we mentioned before, we reduce the number of features from 38 to 30 by eliminating 8 spatial features, which are the lowest correlation related to target class. Therefore, GA will be able to find the optimal subset in less time. After rerunning the same experiments with 30 features, we obtained mostly the same optimal subsets with almost the same accuracy.

المنارة للاستشارات
www.manaraa.com

After re-run Experiment 1 **"GA-ANN"**, the result presented in Tables 4-22 show that the accuracy with the correlation spatial is better than with all features, and is less computationally expensive. We have 180 generations as a MAX GENERATION parameter to find optimal subset in 48 hours, but using 30 features, we need 140 generations in 36 hours to find the same optimal subset.

Table 4-22: Classification accuracy based on All Features and correlation spatial using ANN

No.	Input Features	# of Features	ANN parameter	Accuracy	Time (Seconds)
1	All Features	38	LR = 0.1,Epochs = 450, Hidden Layer (HL) = 11	88.37%	40.65s
2	Texture + Spectral+ Corr. Spatial	30	LR = 0.1,Epochs = 400, Hidden Layer (HL) = 11	88.57%	31.22s

After re-run Experiment 2 **"GA-KNN"**, the results presented in Tables 4-23 show that the accuracy with the correlation spatial is better than with all features, and is less computationally expensive. We have 180 generations as a MAX GENERATION parameter to find optimal subset in 9 hours, but using 30 features, we need 140 generations in 7 hours to find the same optimal subset.

Table 4-23: Classification accuracy based on All Features and correlation spatial using KNN

No.	Input Features	# of Features	KNN parameter	Accuracy	Time (Seconds)
1	All Features	38	K = 8	83.68%	3s
2	Texture + Spectral+ Corr. Spatial	30	K = 8	85.43%	2.4s

After re-run Experiment 3 "GA-J48", the result presented in Tables 4-24 and Table 4-25 show the accuracy with the correlation spatial is worse than with all features (with minor difference), and the correlation spatial is less computationally expensive. Using 38

المنارة للاستشارات
www.manaraa.com

features, we have 180 generations as a MAX Generation parameter to find optimal subset in 11 hours, but using 30 features, we need 140 generations in 8 hours to find the optimal subset.

Table 4-24: Classification accuracy based on All Features and correlation spatial using J48

No.	Input Features	# of Features	J48 parameter	Accuracy	Time (Seconds)
1	All Features	38	CF = 0.05	82.85%	1.55s
2	Texture + Spectral+ Corr. Spatial	30	CF = 0.05	81.42%	1.2s

Table 4-25: Comparsion between GA-J48 with all features and correlation spatial features

No.	Input Features	# of Features	"Optimal" Subset	Time	Accuracy
1	GA-J48 with All Features	38	16	11 hours	85.24%
2	GA-J48 with Texture+ Spectral+ Corr. Spatial Features	30	15	8 hours	84.32%

4.3.5 Experiment 5: Optimal features subsets validation

In validation experiment, we used optimal features subset, which obtained using GA-ANN, GA-J48 and GA-KNN, then perform validation with ANN, KNN and J48 as classifiers. The results in Table 4-26 and Figure 4-10 make it clear that the optimal features subsets as identified by the various wrapper have indeed improved the classification accuracy of all the three classifiers used for validation when compared to classification accuracy with all the features.

Table 4-26: Optimal features subsets validation obtained wrapper approach using classifiers

Wrapper Approach for	Number of Features	Classifiers Accuracy (%)		
		Artificial Neural	J48 Decision	K-Nearest

المنارة للاستشارات
www.manaraa.com

Feature selection Method		Network (ANN)	tree	Neighbors (KNN)
GA-ANN	17	89.70%	85.12%	84.35%
GA-KNN	14	87.79%	83.67%	87.49%
GA-J48	16	88.34%	85.24%	85.34%
Texture+ Spectral+ Corr. Spatial	30	88.57%	81.42%	85.43%
With all Features	38	88.37%	82.85%	83.68%

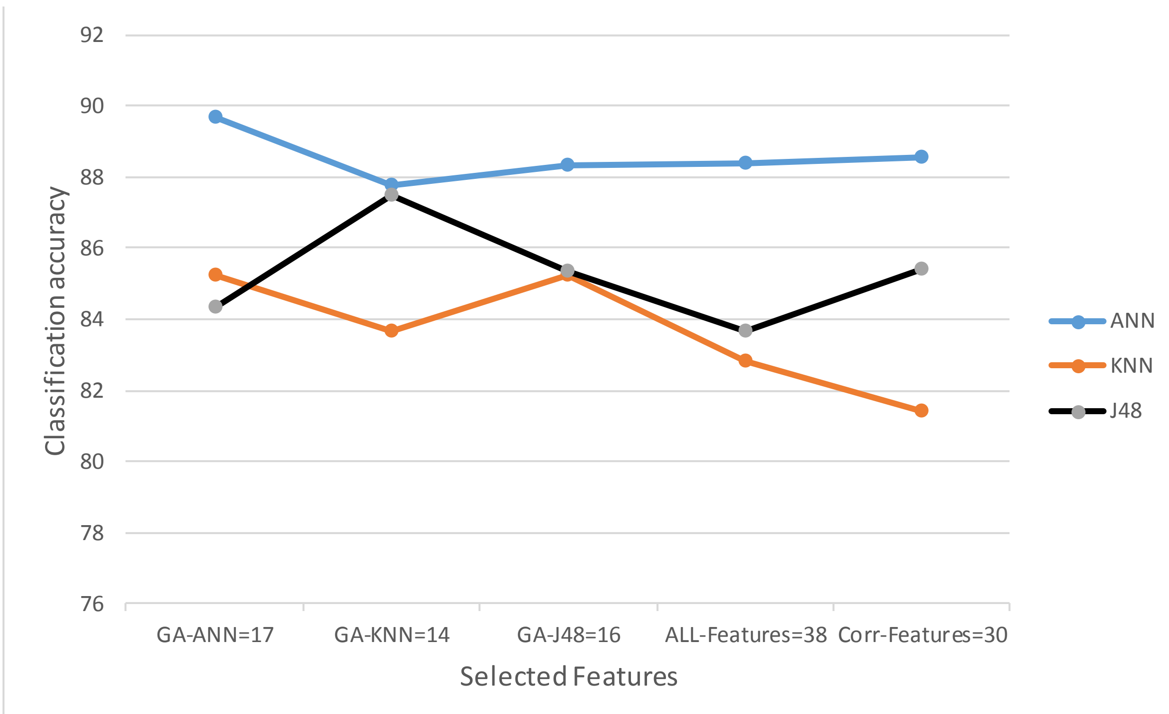


Figure 4-10: Validation optimal features subset obtained wrapper approach using classifiers

4.4 Results Discussion

Feature selection improves calculation efficiency and classification accuracy in classification problems with multiple features. Selecting appropriate features improves the predictive accuracy; on the other hand, selecting inappropriate features compromises

المنارة للاستشارات
www.manaraa.com

the predictive accuracy. Hence, employing appropriate feature selection to select optimal features for a category results in higher classification accuracy.

In Table 4-27, Figure 4-11 and Figure 4-12, we summarize the experiments of wrapper approach based on GA-ANN, GA-KNN and GA-J48.

The best accuracy obtained with GA-ANN features; with training dataset, the accuracy is 89.70%, but with the test dataset, the accuracy was a little bit less with 89.43%. However, the time taken to find the optimal subset features reaches up to 48 hours, which is considered to be a very long time. The main difficulties that might lead to this long time is the variations in satellite images, shadows around the objects such as trees, variation in imaginary resolution, existence of cars in the roads and boats in rivers. The estimation time clearly show that the computation time needed for GA-KNN is shorter than that of GA-ANN and GA-J48.

As mentioned earlier, the spatial features are the least important features among other features. This could be due to the spatial resolution, refer appendix A.1.3 for more details, we downloaded high resolution imagery from different satellites with spatial resolution close to 1 meter, however, that's not enough to recognize objects perfectly. To overcome this problem, we used CRF for spatial features to remove unimportant features.

Table 4-27: Summery of wrapper methods based on (GA-ANN, GA-KNN, GA-J48)

Wrapper methods	"Optimal" Subset	# of Features	Estimation Time to Find Optimal Subset	Accuracy of Training Data set	Accuracy of Testing Data set
GA-ANN	AVG_B1, STD_B1, AVG_B2, STD_B2, MIN_B2, AVG_B3, STD_B3, TXRAN_B1, TXAVG_B1, TXAVG_B2, TXVAR_B2, TXAVG_B3,	17	48 Hours	89.70%	89.43%

المنارة للاستشارات
www.manaraa.com

	TXVAR_B3, TXENT_B3, FX_FORMFAC, FX_RECT_FI, FX_MINAXLN				
GA-KNN	AVG_B1, AVG_B2, MAX_B2, MIN_B2, AVG_B3, TXAVG_B1, TXVAR_B1, TXAVG_B2, TXVAR_B2, TXENT_B2, TXAVG_B3, TXVAR_B3, TXENT_B3, FX_CONVEX	14	9 Hours	87.49%	87.35%
GA-J48	AVG_B1,AVG_B2 ,AVG_B3,MIN_B3 , TXAVG_B1, TXRAN_B2, TXVAR_B2, TXRAN_B3, TXAVG_B3, TXVAR_B3, TXENT_B3, FX_LENGTH, FX_SOLID, FX_RECT_FI, FX_MINAXLN, FX_HOLESOL	16	11 Hours	85.24%	81.16%

للاستشارات المنارة

www.manaraa.com

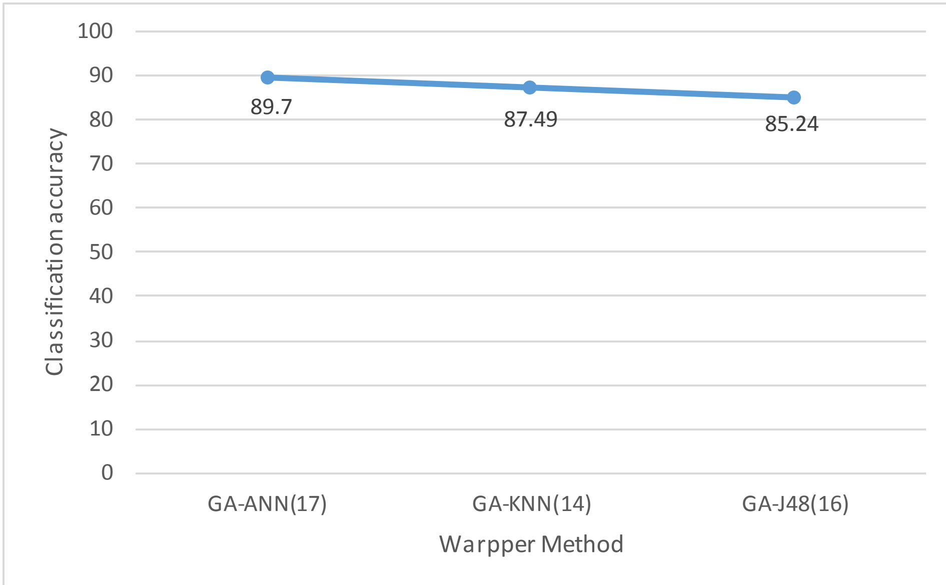


Figure 4-11: Summery of Classification Accuracy and optimum features of wrapper methods based on training dataset

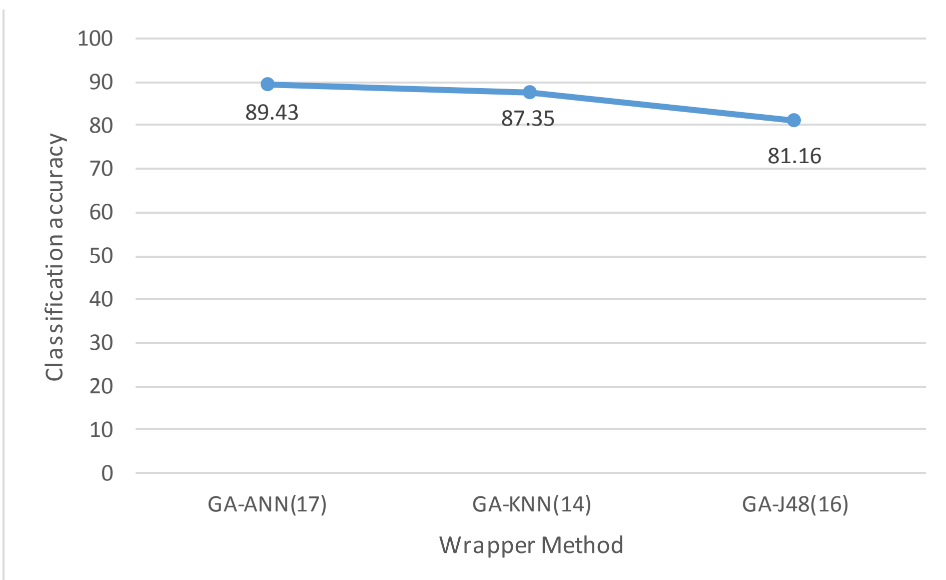


Figure 4-12: Summery of Classification Accuracy and optimum features of wrapper methods based on testing dataset

المنارة للاستشارات
www.manaraa.com

CHAPTER 5: Conclusion and Future Works

5.1 Conclusion

The main objective of this thesis is to improve the accuracy of recognizing objects from satellite imagery based on geospatial features using wrapper approach with a genetic algorithm as an optimization method and neural network, decision tree J48 and K-nearest neighbor as classification and evaluation methods.

ENVI software is used to extract the object features. Our wrapper approach is tested using two datasets for training and testing. Three types of features having 38 features are considered texture, spatial and spectral. Comprehensive experiments are conducted using GA-ANN, GA-KNN and GA-J48, with the help of WEKA software. Experimental evaluation confirms improvement in classification accuracy for all classifiers and the number of features are reduced by at least 55%. The classification accuracy is increased by at least 1.32%. Spatial features are considered to be having the least important features whereas the texture features seems to be having the highest important features. In addition, the correlation ranking filter is used for spatial features and proved that 6 out of the spatial selected features by GA-ANN, GA-KNN and GA-J48 are the same. After removing 8 features from spatial features according to what has been obtained by CRF, the same experiments are conducted using 30 features instead of 38 features and the obtained accuracy and the optimal subsets are almost the same. According to the obtained results among the three approaches GA-ANN, GA-KNN and GA-J48, the GA-ANN is the best with 89.7%. Focusing on GA-ANN results, we found that the largest number of misclassification is between the buildings and roads. This could be due to the similarity of colors between buildings and roads. In contrast, the smallest number of misclassification is between the roads and rivers, which might not be expected due to the similarity between the rivers and roads in shape, especially in satellite images. This result achieved due to the similarity of colors contrast between roads and rivers.

In summary, the proposed wrapper feature selection methods GA-ANN, GA-KNN and GA-J48 can optimize feature subsets and increase classification accuracy at the same time, therefore can be applied in feature selection of the satellite imagery data.

المنارة للاستشارات
www.manaraa.com

4.2 Future Work

The performance could be enhanced more by extracting and selecting the best and the most discriminative features, so for future work we suggest the following:

- Features extraction performance is greatly affected by the segmentation process. In our thesis we use trial and error to choose the best parameters to segment images, it is possible to use GA to choose the best parameters.

- Work to provide very high image resolution to give more accurate results in automatic feature extraction techniques.

- Comparing genetic algorithms with other searching algorithm such as sequential forward selection, sequential backward elimination, and bidirectional selection to find out the optimum subset of features.

- Study different classifiers as evaluation mechanism wrapped with genetic algorithms.

المنارة للاستشارات

www.manaraa.com

References

[1] Anthony, G., and Ruther, H., **"Comparison of feature selection techniques for SVM classification,"** In 10th International Symposium on Physical Measurements and Signatures in Remote Sensing, pp.1-6, 2007

[2] Agrawal, K., and Bawane, G., **"PSO based Selection of Spectral Features for Remotely Sensed Image Classification,"** IJCA Proceedings on National Conference on Innovative Paradigms in Engineering & Technology, pp.22-26, 2013

[3] Aggarwal, S., **"Principles of remote sensing. Photogrammetry and Remote Sensing Division,"** Indian Institute of Remote Sensing, 2003

[4] Bala, J., Huang, J., Vafaie, H., Dejong, K., and Wechsler, H., **"Hybrid learning using genetic algorithms and decision trees for pattern classification,"** In IJCAI (1), pp.719-724, 1995

[5] Baldi, P., and Brunak, S., **"Bioinformatics: the machine learning approach,"** MIT press, 2001

[6] Breiman, R., **"Classification and Regression Trees,"** Belmont, CA: Wadsworth, 1984

[7] Bhargava, N., Sharma, G., Bhargava, R., and Mathuria, M., **"Decision Tree Analysis on J48 Algorithm for Data Mining,"** International Journal of Advanced Research in Computer Science and Software Engineering, pp.1114-1119, 2013

[8] Cui, Y., Wang, J., Bin Liu, S., and Wang, L., **"Hyperspectral Image Feature Reduction Based on Tabu Search Algorithm,"** Journal of Information Hiding and Multimedia Signal Processing, pp.154-162, 2015

[9] Chaudhary, C., and Patil, K., **"Review of Image Enhancement Techniques Using Histogram Equalization,"** International Journal of Application or Innovation in Engineering and Management (IJAIEM), pp.343-349, 2013

[10] Dongare, D., Kharde, R., and Kachare, D., **"Introduction to artificial neural network,"** International Journal of Engineering and Innovative Technology (IJEIT), pp.189-194, 2012

[11] Datta, A., Ghosh, S., and Ghosh, A., **"Wrapper based feature selection in hyperspectral image data using self-adaptive differential evolution,"** In Image

المنارة للاستشارات
www.manaraa.com

Information Processing (ICIIP), 2011 International Conference on pp.1-6, IEEE, 2011

[12] Dash, M., and Liu, H., **"Feature selection for classification,"** Int. J. Intell. Data Anal., pp.131–156, 1997

[13] De Stefano, C., Fontanella, F., and Marrocco, C., **"A GA-based feature selection algorithm for remote sensing images,"** In Applications of Evolutionary Computing, pp.285-294, 2008

[14] Firpi, A., and Goodman, E., **"Swarmed Feature Selection,"** IEEE Pattern Recognition, pp.112-118, 2004

[15] Friedman, H., **"On bias, variance, 0/1—loss, and the curse-of-dimensionality,"** Data mining and knowledge discovery, pp.55-77, 1997

[16] Greene, J., **"Feature subset selection using thornton's separability index and its applicability to a number of sparse proximity-based classifiers,"** In Twelfth Annual Symposium of the South African Pattern Recognition Association, 2001

[17] Goldberg, D., **"Genetic Algorithms in Search, Optimization, and Machine learning,"** Addison-Wesley Professional; 1 st. edition, January 11, 1989

[18] Guangtao, W., Song, Q., Sun, H., Zhang, X., Xu, B., and Zhou, Y. **"A feature subset selection algorithm automatic recommendation method,"** Journal of Artificial Intelligence Research, pp.1-34, 2013

[19] Guyon, I., and Elisseeff, A., **"An introduction to variable and feature selection,"** The Journal of Machine Learning Research, pp.1157-1182, 2003

[20] Goodman, E., **"Introduction to Genetic Algorithms,"** GEC Summit, Shanghai, 2009

[21] Han, J., and Kamber, M., **"Data Mining: Concepts and Techniques,"** 2nd ed. Amsterdam: Elsevier, 2006

[22] Huda, S., Yearwood, J., and Strainieri, A., **"Hybrid Wrapper-Filter Approaches for Input Feature Selection Using Maximum Relevance and Artificial Neural Network Input Gain Measurement Approximation (ANNIGMA),"** Network and System Security (NSS), 2010 4th International Conference on , pp.442,449, 2010

[23] Hossain, R., Oo, T., and Ali, S., **"The combined effect of applying feature selection and parameter optimization on machine learning techniques for**

المنارة للاستشارات

www.manaraa.com

solar Power prediction," American Journal of Energy Research,1(1), pp.7-16, 2013

[24] Haapanen R., and Tuominen S., **"Data Combination and Feature Selection for Multi-source Forest Inventory,"** Photogrammetric Engineering & Remote Sensing, pp.869–880, 2008

[25] Jashki, A., Makki, M., Bagheri, E., and Ghorbani, A., **"An iterative hybrid filter-wrapper approach to feature selection for document clustering,"** InProceedings of the 22nd Canadian Conference on Artificial Intelligence (AI'09), 2009

[26] Jamshidpour, N., Homayouni, S., and Samadzadegan, F., **"Improvement of Hyperspectral Image Classification Using Genetic Algorithm for Feature Selection and SVMs Parameters Optimization,"** pp.1-6, 2008

[27] Kass, V., **"An exploratory technique for investigating large quantities of categorical data,"** Applied statistics, pp.119-127, 1980

[28] Kuncheva, I., and Sanchez, S., **"Nearest Neighbour Classifiers for Streaming Data with Delayed Labelling,"** Data Mining, 2008. ICDM '08. Eighth IEEE International Conference, pp.869-874, 2008

[29] Kohavi, R., and George, J., **"Wrappers for feature subset selection,"** Artificial intelligence, pp.273-324, 1997

[30] Karegowda, G., Jayaram, A., and Manjunath, S., **"Feature subset selection problem using wrapper approach in supervised learning,"** International journal of Computer applications, pp.13-17, 2010

[31] Liu, H., Li, J., and Wong, L., **"A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns,"** Genome Informatics Series, pp.51-60, 2002

[32] Molina, C., Belanche, L., and Nebot, A., **"Feature selection algorithms: a survey and experimental evaluation,"** Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference, pp.306-313, 2002

[33] Murayama, Y., **"Remote Sensing Image Processing,"** Graduate School Life and Environment Sciences, University of Tsukuba, 2010

[34] Mitchell, M., **"An Introduction to Genetic Algorithms,"** A Bradford Book the MIT press, Cambridge: MA, 1998

المنارة للاستشارات

www.manaraa.com

[35] Molinier, M., and Astola, H., **"Feature selection for tree species identification in very high-resolution satellite images,"** In Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, 2011

[36] Moore, W., and Lee, S., **"Efficient Algorithms for Minimizing Cross Validation Error,"** In ICML, pp. 190-198, 1994

[37] Nishida, K., **"Learning and detecting concept drift,"** PhD diss., School of Information Science and Technology, Hokkaido University, 2008

[38] Novakovic, J., Minic, M., and Veljovic, A., **"Genetic Search for Feature Selection in Rule Induction Algorithms,"** 18th Telecommunications forum TELFOR, pp.1109-1112, 2010

[39] Pang-Ning, T. Steinbach, M., and Kumar, V., **"Introduction to Data Mining,"** 1st edition, Addison Wesley, 2005

[40] Pudil, P., Novovicova, J., and Kittler, J., **"Floating Search Methods in Feature Selection,"** International Journal of Remote Sensing, pp.1119-1125, 1994

[41] Quinlan, R., **"C4.5: Programs for Machine Learning,"** San Mateo, Calif.: Morgan Kaufmann, 1993

[42] Quinlan, R., **"Induction of decision trees,"** Machine learning, pp.81-106, 1986

[43] Quinlan, R., **"Probabilistic decision trees,"** In Machine Learning, Morgan Kaufmann Publishers Inc., pp.140-152, 1990

[44] Rashidy-Kanan, H., and Faez, K., **"An Improved Feature Selection Method based on Ant Colony Optimization (ACO) Evaluated on Face Recognition System,"** Elsevier Science, pp.716-725, 2008

[45] Rokach, L., and Maimon, O., **"Top-down induction of decision trees classifiers-a survey,"** Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE, pp.476-487, 2005

[46] Rangaswamy, S., Shobha, G., Satvik, N., and Shivakumar, B., **"Decision Tree Classifiers with GA Based Feature Selection,"** International Journal of Computer Science, pp.75-84, 2013

[47] Sachs, J., **"Image Resampling,"** Digital Light & Color, 2001

[48] Stanley, K., **"Digitizing,"** Virginia University, 2003

المنارة للاستشارات

www.manaraa.com

[49] Trappenberg, P., **"Fundamentals of computational neuroscience,"** New York: Oxford University Press, 2002

[50] Serpico, B., and Bruzzone, L., **"A new search algorithm for feature selection in hyperspectral remote sensing images,"** Geoscience and Remote Sensing, IEEE Transactions on, pp.1360-1367, 2001

[51] Sabuncuoglu, I., Erel, E., and Tanyer, M., **"Assembly line balancing using genetic algorithms,"** Journal of Intelligent Manufacturing, pp.295-310, 2000

[52] Sanderson, S., **"Introduction to remote sensing,"** New Mexico State University, 2001

[53] Sainin, S., and Alfred, R., **"A genetic based wrapper feature selection approach using nearest neighbour distance matrix,"** In Data Mining and Optimization (DMO), 2011 third Conference on IEEE, pp.237-242, 2011

[54] Sutton, O., **"Introduction to K Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction,"** pp.1-10, 2012

[55] Sun, Z., Bebis, G., and Miller, R., **"Object detection using feature subset selection,"** Pattern recognition, pp.2165-2176, 2004

[56] Vafaie, H., and De Jong, K., **"Genetic algorithms as a tool for feature selection in machine learning,"** Tools with Artificial Intelligence, 1992. TAI'92, Proceedings. Fourth International Conference on. IEEE, pp.200-203, 1992

[57] Van Coillie, F., Verbeke, P., and De Wulf, R., **"Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium,"** Remote Sensing of Environment, pp.476-487, 2007

[58] Witten, H., and Eibe, F., **"Data Mining Practical Machine Learning Tools and Techniques,"** 3rd ed. Burlington, Mass.: Morgan Kaufmann, 2011

[59] Yu, L., and Liu, H**., "Feature selection for high-dimensional data: A fast correlation-based filter solution,"** In ICML, pp.856-863, 2003

[60] Yoon, W., Cho, S., Chae, G., and Park, J., **"Automatic land-cover classification of Landsat images using feature database in a network,"** Proceedings of the IGARSS 2005 Symposium, Seoul, Korea, 2005

[61] Zhou, S., Zhang, P., and Su, K., **"Feature selection and classification based on ant colony algorithm for hyperspectral remote sensing images,"** In Image and

المنارة للاستشارات

www.manaraa.com

Signal Processing, 2009, CISP'09, Second International Congress on IEEE, pp.1-4, 2009

[62] Zhu, Z., Ong, S., and Dash, M., **"Wrapper–filter feature selection algorithm using a memetic framework,"** Systems, Man, and Cybernetics, Part B: Cybernetics, Transactions on IEEE, pp.70-76, 2007

[63] Zhuo, L., Zheng, J., Wang, F., Li, X., Ai, B., and Qian, J., **"A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine,"** In Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images, pp.71471J-71471J, 2008

[64] **"NASA,"** NASA. Web. 27 Nov. 2014. <http://www.nasa.gov/multimedia/image-gallery/iotd.html>.

[65] **"Landsat,"** Landsat. Web. 27 Nov. 2014. <http://landsat.usgs.gov/>

[66] **"Topic 10: Image Processing,"** BL-130 Image Processing. Web. 27 Nov. 2014. <http://hosting.soonet.ca/eliris/remotesensing/bl130lec10.html>

[67] **"Characterization of Satellite Remote Sensing Systems,"** Remote Sensing, Satellite Imaging Technology. Web. 27 Nov. 2014. <http://www.satimagingcorp.com/services/ resources/characterization-of-satellite-remote-sensing-systems/>

[68] **"National Aeronautics and Space Administration,"** Science. Web. 27 Nov. 2014. <http://imagine.gsfc.nasa.gov/science/index.html>

[69] **"Concepts and Foundations of Remote Sensing,"** Web. 27 Nov. 2014. < http://dc341.4shared.com/doc/N_3JLY-A/preview.html>

[70] **"Image Interpretation & Analysis,"** Natural Resources Canada. Web. 27 Nov. 2014. <http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9303>

المنارة للاستشارات
www.manaraa.com

Appendix A: Principles of Remote Sensing

A.1 Principles of Remote Sensing

The process of remote sensing involves an interaction between incident radiation and the targets of interest. The process represented by the use of imaging systems where the following seven elements are involved. Note, however that remote sensing also involves the sensing of emitted energy and the use of non-imaging sensors. Figure A-1 shows the essential elements of a remote sensing system, which included the following lines [52].

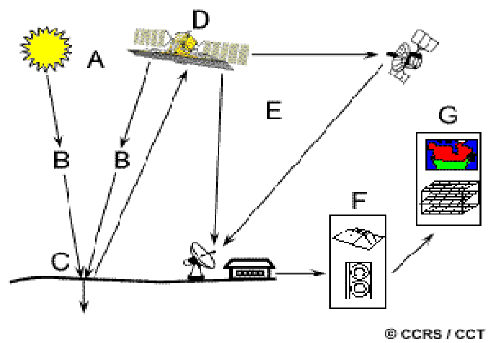


Figure A-1: Elements of remote sensing system [70]

1. ***Energy Source or Illumination (A)*** - energy source which illuminates or provides electromagnetic energy to the target of interest consider the first requirement of remote sensing.
2. ***Radiation and the Atmosphere (B)*** - as the energy travels from its source to the target, it will come in contact with and interact with the atmosphere it passes through. This interaction may take place a second time as the energy travels from the target to the sensor.
3. ***Interaction with the Target (C)*** – after energy pass through atmosphere and reach the target; it interacts with the target depending on the properties of both the target and the radiation.
4. ***Recording of Energy by the Sensor (D)*** -we require a sensor (remotely) to collect and record the electromagnetic radiation after the energy has been scattered by, or emitted from the target.

المنارة للاستشارات
www.manaraa.com

5. ***Transmission, Reception, and Processing (E)*** - the energy recorded by the sensor has to be transmitted, often in electronic form, to a receiving and processing station where the data are processed into an image (hardcopy and/or digital).
6. ***Interpretation and Analysis (F)*** - the processed image is interpreted, visually and/or digitally or electronically, to extract information about the target which was illuminated.
7. ***Application (G)*** – after analysing the raw information from images, the benefits achieved when we apply the information to better understand of issues and solving a particular problem in many fields.

A.1.1 Electromagnetic Radiation

Electromagnetic radiation consists of an electrical field that varies in magnitude, in a direction perpendicular to the direction in which the radiation is traveling, and a magnetic field oriented at right angles to the electrical field. Both these fields travel at the speed of light (c) as shown in Figure A-2 [3].

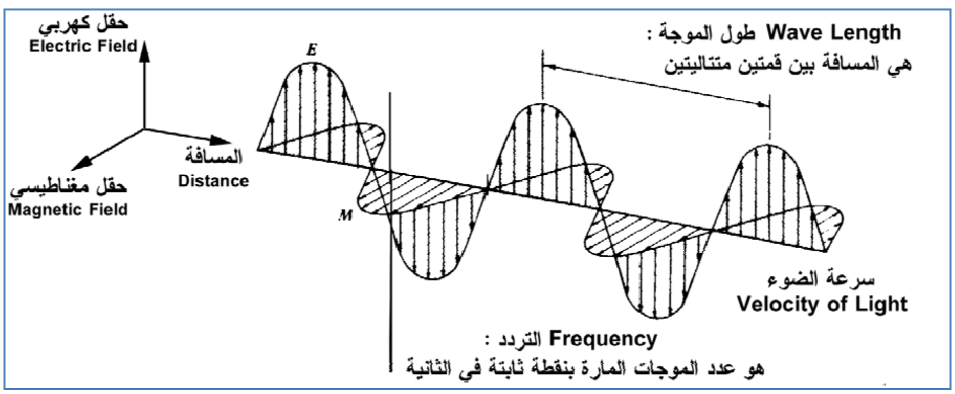


Figure A-2: Electromagnetic radiation components [69]

A.1.2 Electromagnetic Spectrum

The electromagnetic Spectrum is defined as ranges from the shorter wavelengths (including gamma and x-rays) to the longer wavelengths (including microwaves and broadcast radio waves), between this ranges our eyes detect visible spectrum, which consist of three main colors **(RGB) (Red – Green – Blue)** from wavelengths approximately 0.4 to 0.7 μm. Moreover, there are several regions of the electromagnetic

www.manaraa.com

spectrum, which are useful for some remote sensing applications as shown in Figure A-3 [3].

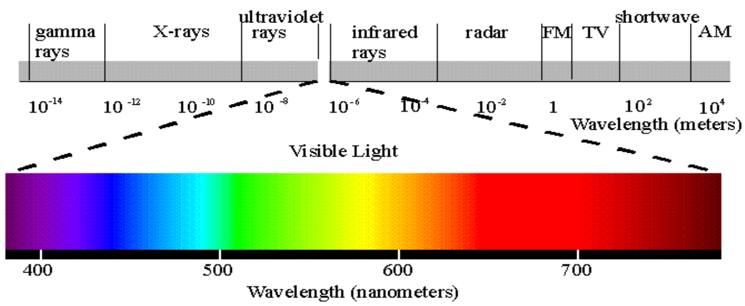


Figure A-3: Electromagnetic spectrum components [68]

A.1.3 Satellite Sensor Characteristics

The principle of most satellite sensors is to gather information about the reflected radiation along a pathway, also known as the field of view (FOV), as the satellite orbits the Earth. The data collected by each satellite sensor can be described in terms of spatial, spectral, radiometric and temporal resolution [47].

- **Spatial Resolution:** The spatial resolution (known as ground resolution) refers to the size of the smallest possible feature that can be detected on ground by sensors, which depends primarily on their Instantaneous Field of View (IFOV).For example the spatial resolution or IFOV of Landsat Thematic Mapper ™ sensor is 30 m [33]. So, the spatial resolution depends on image applications, some of satellites collect data at less than one meter spatial resolution but these are classified military satellites or very expensive commercial systems such as (IKONOS and OUIKBIRD satellites), Figure A-4 shows an example at various spatial resolution (30, 5, 1) meter [33].

www.manaraa.com

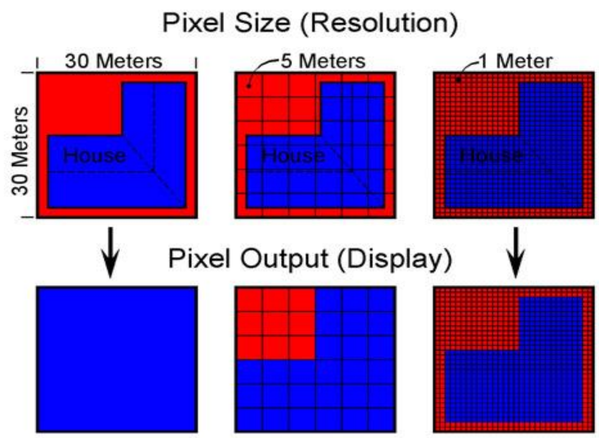


Figure A-4: Spatial resolution [67]

- **Spectral Resolution:** defined as the number and width of spectral bands in the sensing device, also describes the ability of a sensor to define fine wavelength intervals. Imagine with one band is a simplest form of spectral resolution [33].

- **Radiometric Resolution:** The radiometric resolution of an imaging system describes its ability to discriminate very slight differences in energy. The radiometric characteristics describe the actual information content in an image [33].

- **Temporal Resolution:** Temporal resolution is very important in remote sensing system, which refers to the length of time it takes for a satellite to complete one entire orbit cycle. The actual temporal resolution of a sensor depends on a variety of factors, including the satellite/sensor capabilities, the swath overlap, and latitude. With temporal resolution, we are able to monitor changes that take place on the Earth's surface such as (urban development, floods, oil slicks, etc.) Landsat 5 takes 16 day to complete one entire orbit cycle [33].

A.2 Digital Image Processing

A.2.1 Preprocessing

Preprocessing functions mostly fall into categories radiometric and geometric corrections. Radiometric corrections include correcting the data for sensor irregularities and

المنارة للاستشارات
www.manaraa.com

undesirable sensor or atmospheric noise, and converting the data so they accurately represent the reflected or emitted radiation measured by the sensor.

Geometric corrections include correcting for geometric distortions due to sensor-Earth geometry variations, and conversion of the data to real world coordinates (e.g. latitude and longitude) on the Earth's surface. Conversion data to real world coordinates done by analyzing well-distributed Ground Control Points (GCPs). Geometric corrections can do in two steps, Geo-referencing and Geocoding [30].

A.2.2 Image Enhancement

Image enhancement method is called contrast enhancement. In raw imagery, the useful data often populates only a small portion of the available range of digital values (commonly 8 bits or 256 levels). Contrast enhancement involves changing the original values so that more of the available range is used, thereby increasing the contrast between targets and their backgrounds. Linear contrast stretch is considering the simplest type of contrast enhancement.

A.2.3 Image Transformation

Image transformation methods can be classified in two ways, first theoretical transformation methods that used some of calculations such as addition and subtraction, multiplication and division and the application of certain mathematical models. Second empirical transformation methods such as conversion principal components also conversion Gradient color and radiation [52].

A.2.4 Image Segmentation

Image segmentation can be performed automatically by employing an edge-based segmentation algorithm that is very fast, familiar end user and only requires one input parameter (scale level). An example of image segmentation is shown in Figure A-5.

المنارة للاستشارات
www.manaraa.com

Figure A-5: Example of Satellite imagery and image segmentation

A.2.5 Feature Extraction

Figure A-6 shows idea of the basic feature extraction. Traditional classification methods are pixel-based, meaning that spectral information in each pixel is used to classify imagery. With high-resolution panchromatic or multispectral imagery, an object-based method offers more flexibility in the types of features to extract [60].

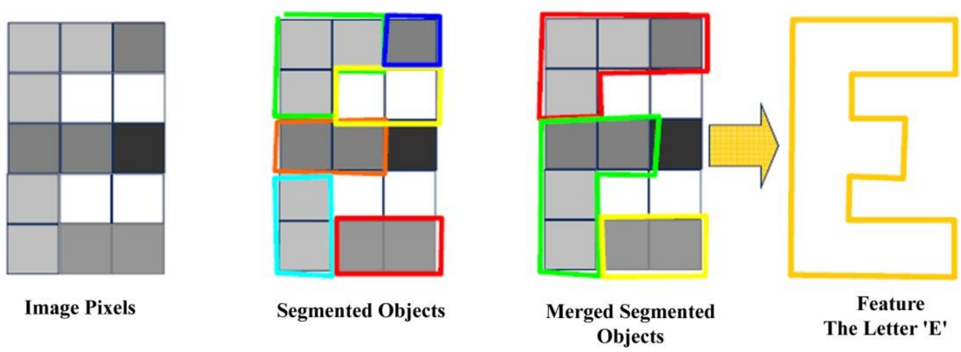


Figure A-6: Concept of object-based feature extraction [67]

The workflow of object based feature extraction involves the following steps:

- Dividing an image into segments
- Computing various attributes for the segments
- Creating several new classes
- Interactively assigning segments (called training samples) to each class
- Exporting the classes to a Shapefile or classification image

www.manaraa.com