A Review on Classification of Satellite Image Using Artificial Neural Network (ANN)

Nur Anis Mahmon[1] and Norsuzila Ya'acob [1,*]

[1]Faculty of Electrical Engineering,Universiti Teknologi MARA,

404500 Shah Alam Selangor, Malaysia

[*]Wireless Communication Technology (WiCoT),
Faculty of Electrical Engineering, Universiti Teknologi MARA,
404500 Shah Alam Selangor, Malaysia
nuranismahmon@yahoo.com, norsuzilayaacob@yahoo.com, azita968@salam.uitm.edu.my

***Abstract*—Artificial Neural Networks (ANNs) have been useful for decades to the development of image classification algorithms applied to several different fields. Image classification is the major component of the remote sensing to extract some of the important spatially variable parameters, such as land cover and land use (LCLU). The aim of this study is to investigate the capability of Artificial Neural Network system (ANNs) for classifying the satellite images using different algorithm which are back-propagation algorithm and K-means algorithm with different approaches. ANN's classifier is compared with two classification techniques of conventional classifier which are Maximum Likelihood (ML) and unsupervised (ISODATA). Neural network classification is based on the training data set and it the proper classification. ML and ISODATA classifiers are broadly used in many remote sensing applications. Overall classification accuracy and Kappa Coefficient were calculated to get the comparison of the performance the image classification. The optimal performance would be identified by validating the classification results with ground truth data. The accurate classification can produce the correct LU/LC map that can be used fir variety.**

***Key Words*— Remote Sensing, Artificial Neural Network, Land Use and Land Cover.**

I. INTRODUCTION

Image classification of thematic land use and land cover (LU/LC) inventories by using remotely sensed data to generate a various applications, especially for control the environment including agriculture crop characteristic and forest ecosystem classification [1]. Image classification is the major component of the remote sensing [2]. Image classification depends on the spectral distinctness of classes and also on the context the classification. Nevertheless, it is important to refine techniques to improve the accuracy image classification of remote sensing for deriving the land use and land cover information.

Nowadays, Artificial Neural Network (ANN) has obtained increasing popularity as an alternative to statistical methods for classification of Remote Sensing Data. This paper is focusing on two algorithms of ANN which is back-propagation and k-means to classify the land use and land cover of Landsat Image. These algorithms investigated with several approaches for the image classification effort.

II. REMOTE SENSING

Remote sensing is the science and art of obtaining the object, phenomenon or area through the analysis of data acquired by a device that is not touched with the object, area or phenomenon under investigation that can be done at any time [3].

Usually the remote sensing scope is when the farmer observes or ranchers observe their field to look the condition without touching just make the observation. The observation is done when their looking the colour or overall appearances of plants which reflect the plant's condition. But in remote sensing application the plant's condition illustrated by satellite image without physically touching. The satellite image is taken from satellites and aircraft from a point of view high above the field [4]. Remotely sensed data are widely used in (LU/LC) classification

Image processing is the one technique to collect for manipulation of digital images using a computer. Image classification normally comprises in four steps: 1. Pre-processing which is to reduce haze, atmospheric correction and finding the band ratio, etc. 2. Training sample: Process to do selection of particular criteria feature for describes the pattern, 3. Decision: Select the suitable technique for comparing the image pattern according with the target, 4. Assessing the accuracy of the classification [5]. After the satellite image classification was done the LU/LC mapping will be produced with the accurate classification.

III. Image Classification Using ANN

A. Image Classification

Image classification is very important task for many aspects of the environmental application. This study emphasizes on the analysis and usage of different algorithm of neural network which are back-propagation and k-means algorithm. Finally the study will compare the neural network classification with another method such as Maximum Likelihood and ISODATA classifier. Present work categorizes the LU/LC classes are typically different types to cover the area of the study area. Table I shows the categories for three groups of classes has been used in this study.

TABLE I. Types of Classes for Image Classification

Number of classes	Type of Classes
5	(i) agriculture (including forest and vegetation); (ii) waterbodies; (iii) industrial; (iv) open space; (v) road
4	(i) agriculture (including forest and vegetation); (ii) waterbodies; (iii) industrial; (iv) road
2	(i)Developed area; (ii) undeveloped area

B. Back-Propagation Algorithm

Image classification will be conducted using supervised techniques feed-forward neural network which is back-propagation algorithm that corresponding in Fig 2.

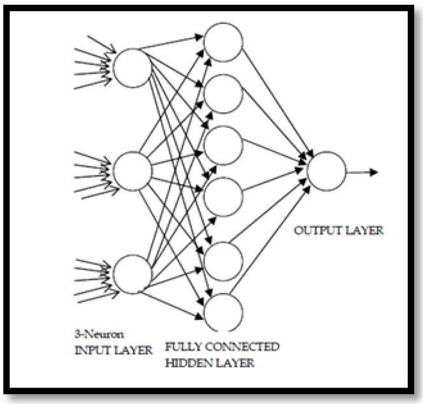


Fig. 1. Back-propagation Architecture for the feed-forward network of single hidden layer case[9]

Before training and classify LU/LC of image satellite, the normalized process of training sample has been performed. This process is to avoid the saturation in the process of network broadcasting. Equation below shows the normalized formula [8]:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

According to the formula above, x' mean value of normalized input, x equal input vector (original value), x_{min} and x_{max} represents the original entire training samples set for minimum and maximum values.

In BPNN process multiple hidden layers for feed-forward will be used, and the number of hidden layers can be changed based on discretion. The number of neurons in output layer will be equal to the number of classes (N), which is based on coding, followed the output. The number of hidden layer neurons is proposed according the some criteria include the number of hidden neutrons should be in the range between the size of the input layer and size of the output layer. From previous study the number of hidden layer generally proposed to be following the formulations [9]:

$$Nh = Np\left(r\left(Ni + No\right)\right) \quad (2)$$

Where:

Np = Number of training samples
Ni = Number of input neurons
No = Number of output neurons

Back-propagation is the most popular method which has already programmed to create the network model and to teach the networks. Nowadays have other modern methods of trained the data which is conjugate gradient method and the Lavenberg-Marquardt method. These methods have their own advantage which is they are faster. But such advantage occurs only in case when the problem should be solved by the neural network with finding out the method of its solution on the basis trained process. This project prefers using back-propagation algorithm compares with modern algorithm because BPNN is the method that works independently from what so ever theoretical assumptions. It means, contrary to other clever algorithm which sometimes works, the back-propagation methods always works.

Selection of the training set is important stages to get the best performance of classification.

C. Training Data Set

The training data provided to complete a representative description of each land use and land cover category as possible, the size of the set will increase significantly along with an increase the spectral variability of desired classes [6]. The training set is trained using samples from either ground truth data and the output of a K-means clustering image output. This showed the training set is trained from different training samples to certain representative parts of the given data set. Before starting trained the data samples the image texture important to interpret each type of ground object category. According to research work [7], the different sizes of training set can give different impact for the classification output (Fig 1). This project selects the training data sizes randomly.

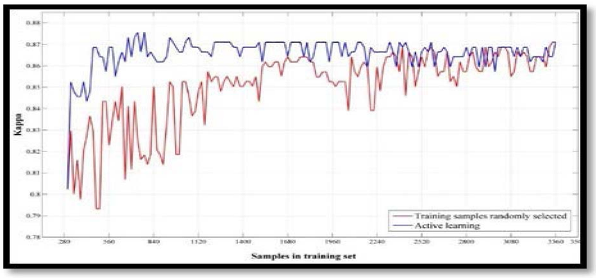


Fig. 2. Example of active learning results and result based on training samples randomly selected and this achieve the highest accuracies when there are more than 200 samples per class [7].

According to Fig 3, the image classified five classes using back-propagation algorithm.

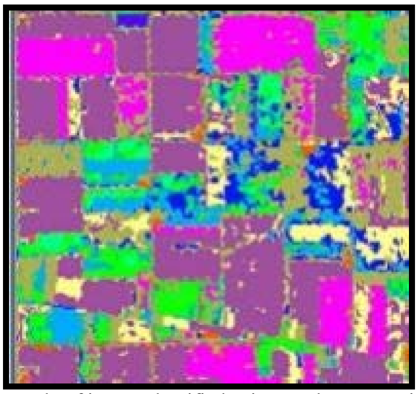


Fig. 3. Example of image classified using Back-propagation algorithm [10].

D. K-means Algorithm

K-means algorithm as unsupervised techniques is used; this algorithm starts with some clusters of pixel in feature space. Data clustering will be conducted before continuing with data classification. Data clustering is one process which large data set of data is grouped into clusters of a smaller set in similar data. K-means algorithm used in this study is one of the most non-hierarchical for data clustering. This algorithm is the process that the user can specify the number of clusters to be the classification of the satellite image. In the present study the clustering algorithm chosen is given by [11].

$$J = \sum_{i=1}^{k} \sum_{x_j \in s_i} \left\| x_j - m_i \right\|^2 \tag{3}$$

Where:

k = Number of clusters

S_i, i=1, 2, 3…

x = Feature sample

m_i = sum of the cluster

$\left\| x_j - m_i \right\|^2$ = Distance that measure

The output of K-means algorithm is used in two tasks, where (a) as the input for image sample for training the data; and (b) for testing the accuracy of classification to be compared with another algorithm in artificial neural network system (Fig 4.).

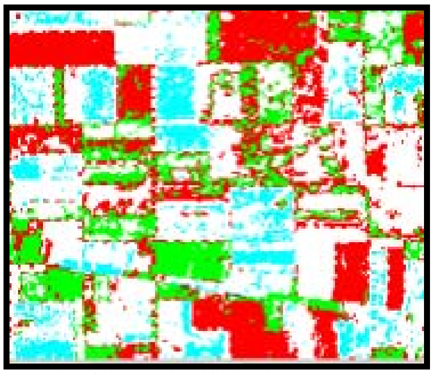


Fig. 4. Example of image classified using K-means algorithm [10]

E. K-means Clustering

This processing is carried out with different numbers of clusters. The relationship between the number of cluster with the output k-means clustering will be provided whether can give effect in trained the sample data. This study provided the number of cluster is depending on the number of classes (N) for classification. Eq. 4 shows the proposed algorithm of the k-means clustering.

$$J = \sum_{i=1}^{k=N} \sum_{x_j \in s_i} \left\| x_j - m_i \right\|^2 \tag{4}$$

From previous study the number cluster for K-means algorithm was randomly selected. Based on Fig. 5, the program will make group or cluster the data by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Each dot represents an object and the coordinates (X, Y) represents two attributes of the object. The colours of the dot and label number represent the cluster. In this paper, the number of cluster is equal to the number of classes.

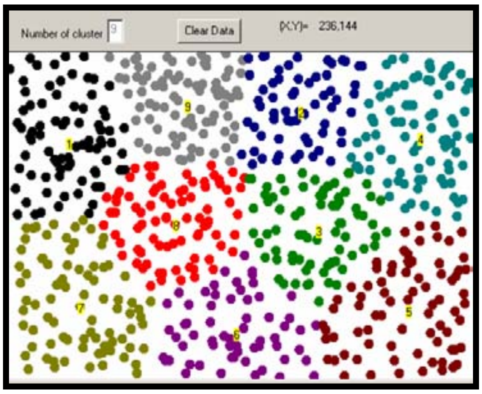


Fig. 5. Example of cluster the data by minimizing the sum of squares of distances between data and the corresponding cluster centroid [12].

IV. Ground Truthing

Ground truth or field observation helps to identify the true type of classes in the study area or as the sample training data set to train each algorithm for image classification. In this work Global Positioning System (GPS) has been used to get the observation of the study area which observe in 15 points for each ground categories to cover the study area. Total ground point by divided in two sections, which is half of the set is for training sample and the other will used as a datum or reference point for validation stages.

For validation stages the location of every pixel of classes will be compared with classification results from ANNs and conventional methods. The location of the classification results and the datum is obtained from the GIS process using ArcGIS software. These comparisons also will analyses using confusion matrix table.

V. ACCURACY ASSESSMENT

A. Confusion Matrix

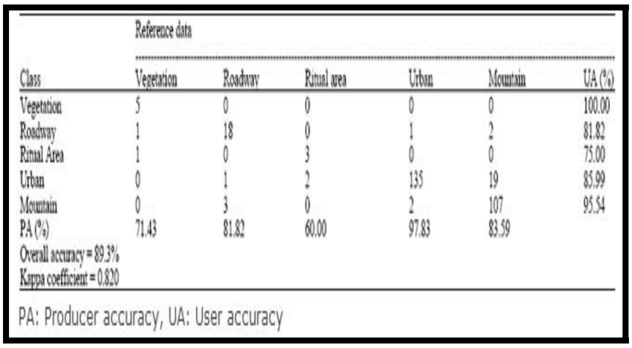
The results of the classification are depending on the accuracy assessment and Kappa coefficient value. The percentage of accuracy of classification result for all classifiers was calculated by analysed with confusion matrix and also called error matrix (Fig. 5). Beside this, there is some indicator that used to show the classification results such as overall accuracy, producer accuracy, user accuracy and Kappa coefficient value [13].

$$k = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+}.x_{+i}}{N^2 - \sum_{i=1}^{r} x_{i+}.x_{+i}} \tag{5}$$

Where:

X_{ii} = Sum of diagonal input of error matrix
X_{i+} = Sum of row I of error matrix
X_{+I} = Sum of column I of error matrix
N = No. of elements in error

Based on [14], producer accuracy was calculated by dividing the number of correct objects of a specific class with the actual number of reference data objects for that class. While the user accuracy was determined by dividing the number of correct objects of a specific class by the total number of objects assigned to that class. To perform the producer accuracy the proportion of labelled object in the reference data was informed correctly. User accuracy, however, quantifies the proportion of objects assigned to a specific class that agree with the objects in the reference data. User accuracy indicates the probability that a specifically labelled object also belongs to that specific class in reality. It can show the commission errors.



Class	Reference data: Vegetation	Roadway	Ritual area	Urban	Mountain	UA (%)
Vegetation	5	0	0	0	0	100.00
Roadway	1	18	0	1	2	81.82
Ritual Area	1	0	3	0	0	75.00
Urban	0	1	2	135	19	85.99
Mountain	0	3	0	2	107	95.54
PA (%)	71.43	81.82	60.00	97.83	83.59	

Overall accuracy = 89.3%
Kappa coefficient = 0.820

PA: Producer accuracy, UA: User accuracy

Fig. 6. Example the table of confusion matrix of the best results: comparison between ANN classifier, ML classifier and ISODATA classifier with ground truth data [15].

B. Land Use and Land Cover Mapping

Land use and land cover maps can play an important role in the environment management. The five classes of image classification from the land use and land cover classifications can describe the environment of the study area. So, this project obtained classified of LU/LC classification with analysis of LANDSAT 8 satellite image. Based on Fig 6 the classified map produced according to the best result of image classification.

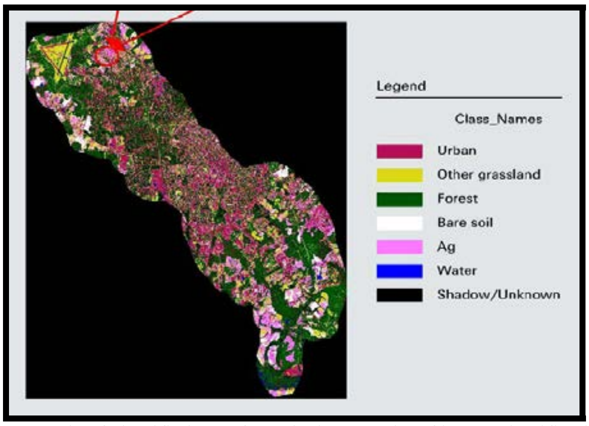


Fig. 7. Example of classified map from the best results of image classification [16].

VI. CONCLUSION

This study reviews the capability ANNs to classify image satellite with various ANNs approaches. The classification takes difference time to classify because the numbers of data training set is depending on the algorithm neural network model and number of classes. This paper also reviewed various techniques applicable for classification of satellite images to be compared with ANNs according to the value of accuracy assessment. The comparison of these techniques also will be validated with the datum, this study approaches the best techniques to classify the land use and land cover with various approaches of classification techniques. At the last output the LU/LC map of the study area can indeed be applied for a variety of purposes such as deforestation, archeology, weather forecasting, urban planning, development, damage assessment etc.

ACKNOWLEDGMENT

The authors would like to thank Faculty of Electrical Engineering for their valuable support. The authors also would like to thank Agency Remote Sensing Malaysia (ARSM) for providing the satellite imagery and UiTM Library-Tun Abdul Razak Library for supplying topographic map of the study area.

REFERENCES

[1] Khorram, S. "Comparison of Landsat MSS and TM data for urban land-use classification", IEEE Trans. Geoscience Remote Sensing. 1987, 25, 238-243.

[2] Mukherjee, J. "MRF Clustering for segmentation of color images" in Pattern Recognition Letters. 2002, pp. 917-929.

[3] Lillesand, T.M., and R. W. Kiefa "Remote Sensing & Image interpolation". John & Sons. 2000

[4] Karry Kyllo "Agricultural Remote Sensing Basics" in NDSU: North Dakota State University. 2004

[5] M.Seetha, I.V.Muralikrishna "Artificial neural network and other methods of image classification" in Journal of theoretical and Applied Information Technology.

[6] Principe, J.C., Euliano, N.R., Lefebvre, W.C. "Neural and adaptive system: fundamentals through simulations", John Wiley & Sons, Inc.: New York, NY, USA. 1999, pp. 100-222.

[7] Congcong Li et al "Comparison of classification algorithm and training sample sizes in urban land classification with Landsat thematic Mapper imagery" in ISSN Remote Sensing. 2014, pp. 968-983.

[8] Jiefeng Jiang etl. "Application of Back Propagation Neural Network in the Classification of High Resolution Remote Sensing Image". Geographic Information Information research.

[9] Mayank Toshniwal "Satellite image classification using neural network" in 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications. 2005, pp. 27-31.

[10] Smriti Seghal "Remotely sensed LANDSAT image classification using neural network approaches" in International Journal of Engineering research and Applications (IJERA), vol 2, pp.043-046, 2012.

[11] Y.V. Venkatesh and S. Kumar Raja "On the classification of multispectral satellite images using the multilayer perceptron" In pattern recognition, The Journal of Pattern Recognition Society, 2002.

[12] Kardi Teknomo "K-Means Clustering Tutorial" of the Teknomo, Kardi. K-Means Clustering Tutorials. 2007.

[13] Sultan AlSultan "Innovative Satellite Image Map Of R.Alkhabrta Area, Saudi Arabia Using High Resolution Image", School of Physics.

[14] Tso, B. and P.M. Mather. "Classification Methods for Remotely Sensed Data". 1st Edn., Taylor and Francis, London. 2001.

[15] Congcong Li et al "Comparison of classification algorithm and training sample sizes in urban land classification with Landsat thematic Mapper imagery" in ISSN Remote Sensing. 2014, pp. 968-983.

[16] Siamak Khorram. "Application of Neural Network-based Classification for Watershed Land Cover Mapping", North Carolina State University (NCSU).