MAXIMUM LIKELIHOOD CLASSIFIER AND ARTIFICIAL NEURAL NETWORKS FOR LAND USE AND LAND COVER CLASSIFICATION BASED ON TEXTURE ANALYSIS USING THEOS, CASE STUDY OF CHOK CHAI DISTRICT, NAKHON RATCHASIMA PROVINCE OF THAILAND

Sasikarn PLAIKLANG[a] Yaowaret JANTAKAT[b] and Suwit ONGSOMWANG[c]

[a]Graduate student, School of Remote Sensing, Institute of Science, Suranaree University of Technology,
111 University Avenue, Suranaree Sub-district, Muang. Nakhon Ratchasima 30000, Thailand;
Tel. +66(0) -44-224652; Fax. +66(0) -44-224316
E-mail: prettynoom@hotmail.com

[b]Lecturer, Department of Applied Information and Communication Technology,
Faculty of Sciences and Liberal Arts, Rajamangala University of Technology Isan,
744 Suranarai Road Muang District, Nakhon Ratchasima 30000, Thailand;
Tel. +66(0) -44-233000; Fax. +66(0) -44-233052
E-mail: yjantakat@gmail.com

[c]Asst. Prof., School of Remote Sensing, Institute of Science, Suranaree University of Technology,
111 University Avenue. Suranaree Sub-district, Muang. Nakhon Ratchasima 30000, Thailand;
Tel. +66(0) -44-224652; Fax. +66(0) -44-224316
E-mail: suwit.ongsomwang@gmail.com

KEY WORDS: Maximum Likelihood Classifier (MLC), Artificial Neural Networks (ANN), Land use and land cover classification, Texture measures, THEOS

Abstract: The aim of the study is to evaluate suitable algorithm and datasets for land use and land cover (LULC) classification in Chok Chai district of Nakhon Ratchasima province in Thailand. This study prepared 10 datasets (1 multispectral data and 9 texture measures data) that were used for LULC classification using supervised classification with Maximum Likelihood Classifier (MLC) and Artificial Neural Networks (ANN). Herein, all datasets were classified into 10 classes that consisted of (1) urban and built-up area, (2) paddy field, (3) cassava, (4) sugarcane, (5) eucalyptus, (6) orchard, (7) forest land, (8) water body, (9) scrub, and (10) abandon land. In addition, accuracy assessment of LULC classification in each dataset was performed with Overall Accuracy (OA) and Kappa hat coefficient of agreement ($\widehat{K}$). As a result, MLC is suitable algorithm for combined datasets of multispectral data and texture measures because accuracy of all combined datasets was higher than ANN. Herewith multispectral data with mean dataset provided the highest OA and $\widehat{K}$ of 83.75% and 81.17%, respectively. However, if only multispectral dataset was considered, ANN was more suitable than MLC since it provided higher accuracy than MLC. This provided OA and $\widehat{K}$ of 86.25% and 84.10%, respectively. Furthermore, $\widehat{K}$ for each LULC types showed that applying texture measures with multispectral data of THEOS can increase the accuracy of LULC classification.

1. INTRODUCTION

Remote Sensing is the efficient tool for acquiring the collection of data in term of spatial and temporal to widely study in land use and land cover change (Reis, 2008; Rogana and Chen, 2004). Therefore, remote sensing has been precious for environmental research and planning (Powell et al., 2008 and Joshi et al., 2006). Researches in satellite image classification have long attracted the interest of the remote sensing community since most environmental and socioeconomic applications are based on the classification results (Perumal and Bhaskaran, 2010; Lu and Weng, 2007). A number of algorithms for supervised classification have been developed over the past decade to cope with both the increasing demand for these products and the specific characteristics of a variety of scientific problems (Samaniego and Schulz, 2009). Supervised classification has been developed to tackle the multispectral data classification (Landgrebe, 2002). In addition, texture measures can be increase efficiency in the process of per-pixel classification (Lobe, 1997). Four groups of techniques have been used to extract texture information from remote sensing images. These include: (1) the first-order statistics, (2) the second-order statistics, (3) the third-order statistics and (4) fractal (Berberoglu and Curran, 2006).

This study used three groups of texture measures as follows: (1) the first-order statistics: texture measures are statistics calculated from the original image values, like variance, and do not consider pixel neighbor relationships, (2) the second-order statistics: texture measures is based on brightness values spatial-dependency Gray-Level Co-occurrence Matrices (GLCM), (3) The third-order statistics: semivariogram is used for spatial variation analyze, by which it is a function relating one-half the squared differences between points to the directional distance between two samples and it separated distance of lag (Jensen, 2005).

The aims of the study to evaluate optimum algorithm and dataset from multispectral data and texture measures data for land use and land cover classification based on THEOS data. Herewith such prepared datasets are classified into 10 LULC types (i.e. urban and built-up area, paddy field, cassava, sugarcane, eucalyptus, orchard, forest land, water body, scrub and abandoned land) using supervised classification with MLC and ANN. Then accuracy assessment of LULC classification from each datasets is performed to identify optimum algorithm.

2. OBJECTIVES

1. To classify LULC types using supervised classification with MLC and ANN algorithms base on multispectral data and texture measures data from THEOS
2. To compare accuracy of LULC classification
3. To evaluate optimum algorithm and dataset from multispectral data and texture measures data for LULC classification based on THEOS

3. STUDY AREA

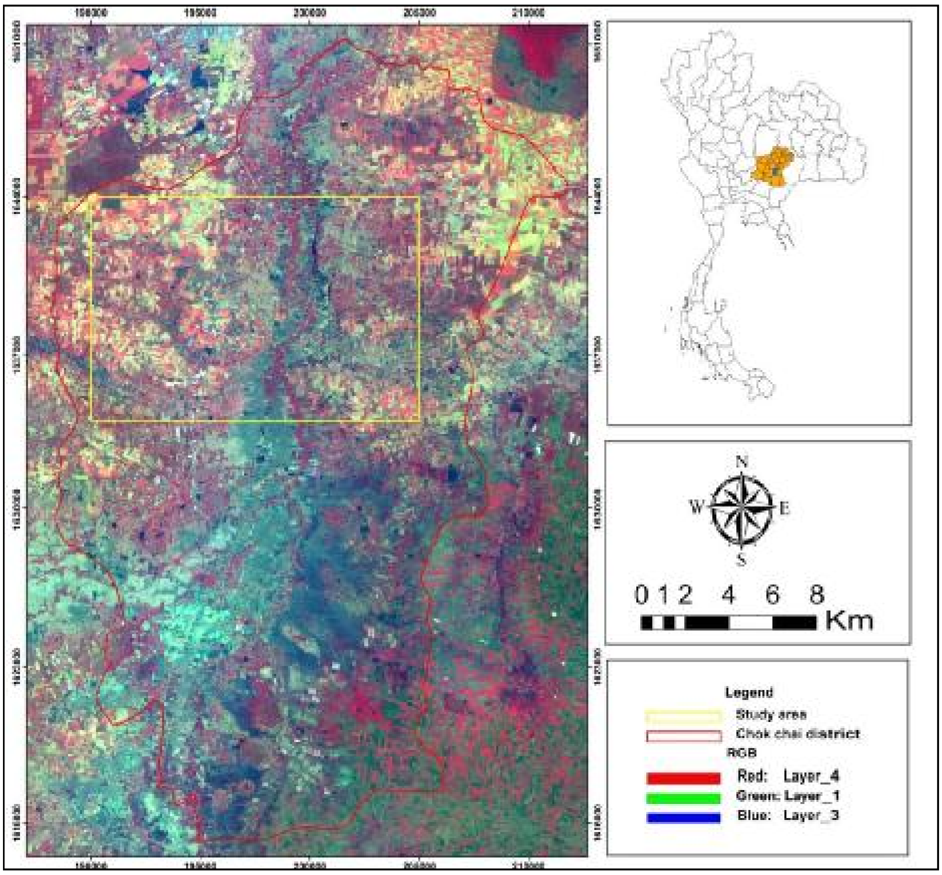

Figure 1: Study area

The study area is a part of Chok Chai district where locates in Nakhon Ratchasima province of southeast Thailand (Figure 1). The situation is on UTM coordinate system between 190000E - 205000E and 1633995N - 1644000N. The study area is approximately 150 sq. km. where topographic characteristic is the undulating plateau with rivers flow through the area. This provides the area is suitable for farming and soil is mostly sandy loam.

4. METHODS

The steps of the process are showed in Figure 2. The details of each process were summarized in the following sections.

1. **Data Collection**
   1.1 Satellite data: THEOS multispectral data with band 1, 2, 3 and 4 are acquired on 19 February 2010.
   1.2 GIS data: GIS data collection (e.g. land use layer, topographical map, and administration boundary layer).
2. **Preprocessing consists of three processes as follows:**
   2.1 Geometric correction: operating with image to image registration in UTM coordinate system and datum WGS 1984 zone 48 based on ground control points (GCPs) collecting from orthophotomap years 2000 - 2002 of Land Development Department (LDD).
   2.2 Optimum Index Factor (OIF): three bands combination are extracted with maximum of OIF value.
   2.3 Principal Component (PC): using the output from 2.2 to create the first PC (PC1) for texture measures.
3. **Data Extraction and preparation consist of two processes as follows:**
   3.1 Texture measures calculation: using PC1 from 2.3 to calculate texture measures.
   3.2 Dataset preparation: using the output from 3.1 to create ten datasets: (1) dataset of multispectral (MS), (2) dataset of MS and mean, (3) dataset of MS and variance, (4) dataset of MS and contrast, (5) dataset of MS and angular second moment, (6) dataset of MS and correlation, (7) dataset of MS and homogeneity, (8) dataset of MS and entropy, (9) dataset of MS and dissimilarity and (10) dataset of MS and semivariogram.
4. **Data classification:** using the output dataset from 3.2 to classify LULC using supervised classification with MLC and ANN algorithm.
5. **Post processing operation:** using LULC classification output from 4 to spatial filtering by Majority filtering algorithm.
6. **Ground verification and accuracy assessment:** performed accuracy assessment and reported OA and $\widehat{K}$.
7. **Optimum classification method and dataset:** using the output from 6 to evaluate optimum algorithm and dataset for LULC classification (Figure 2).

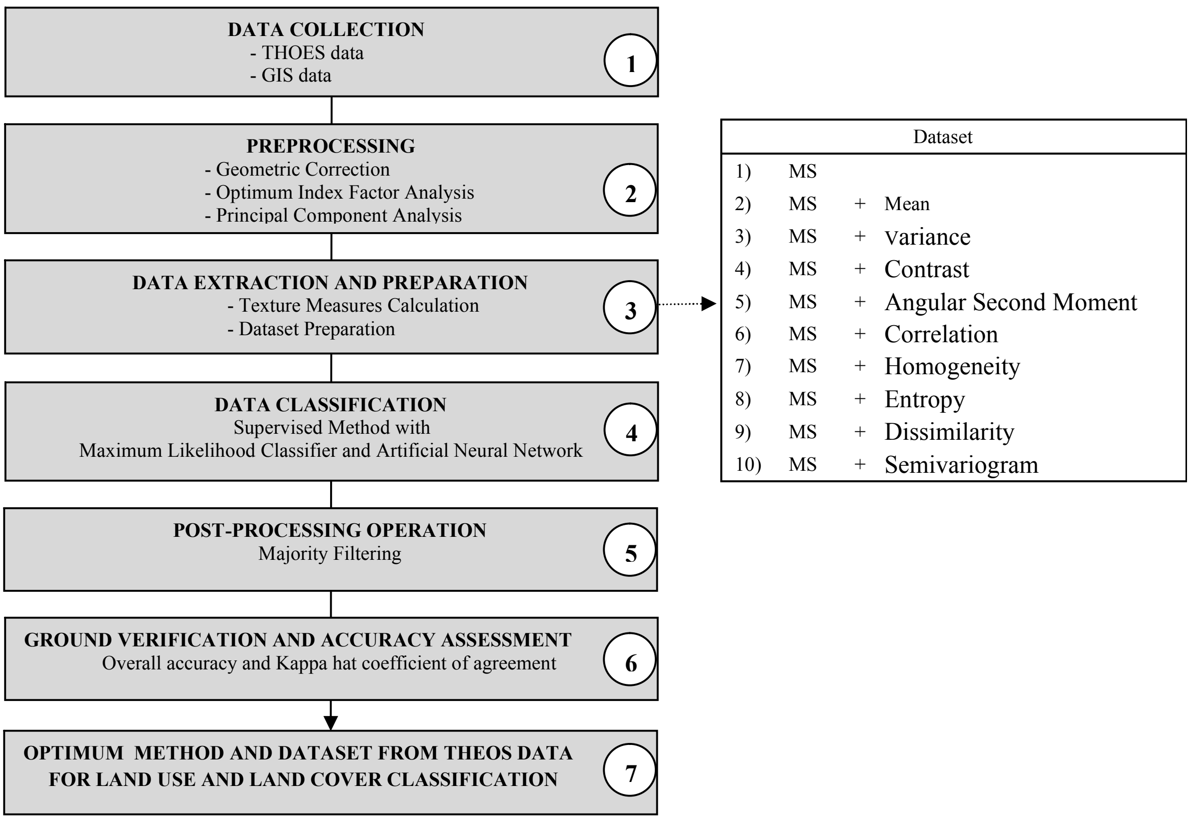

Figure 2: The steps of the process

5. RESULTS

The results of LULC classification based on texture analysis using THEOS can be explained as follows:

5.1 Texture measures

The output of PC1 was used for calculating nine texture measures (Figure 3) and creating ten datasets for LULC classification.

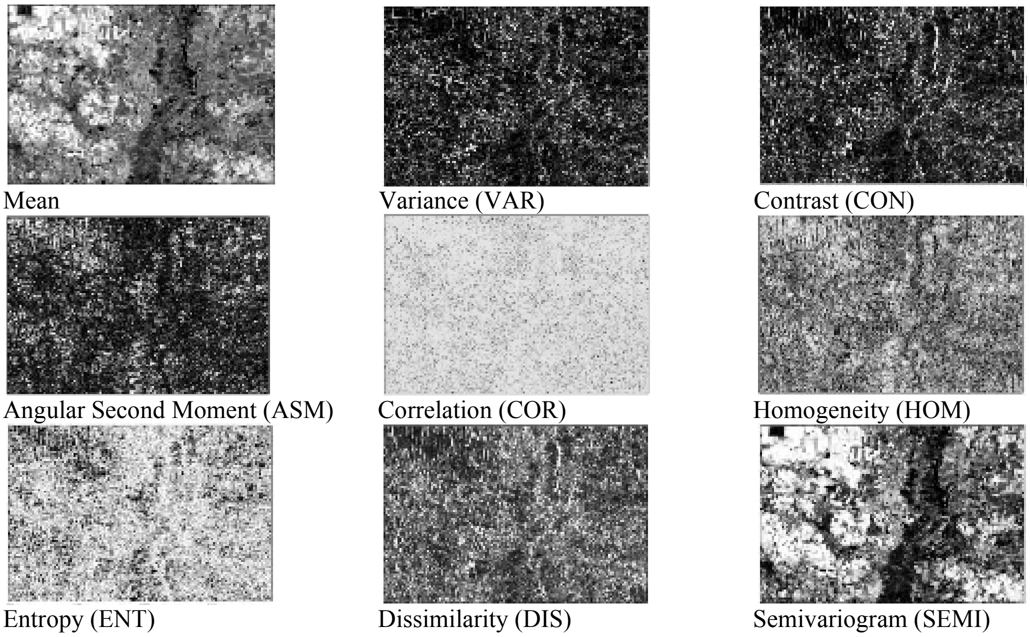

Figure 3: nine texture measures for creating datasets for LULC classification

5.2 LULC classification and post processing operation

LULC classification of ten datasets were analyzed by supervised classification with MLC and ANN algorithm and then performed post processing operation by Majority filtering algorithm as shown in Figure 4 and Figure 5, respectively. The results of LULC classification from ten datasets with MLC and ANN algorithm were shown in Table 1 and Table 2, respectively.

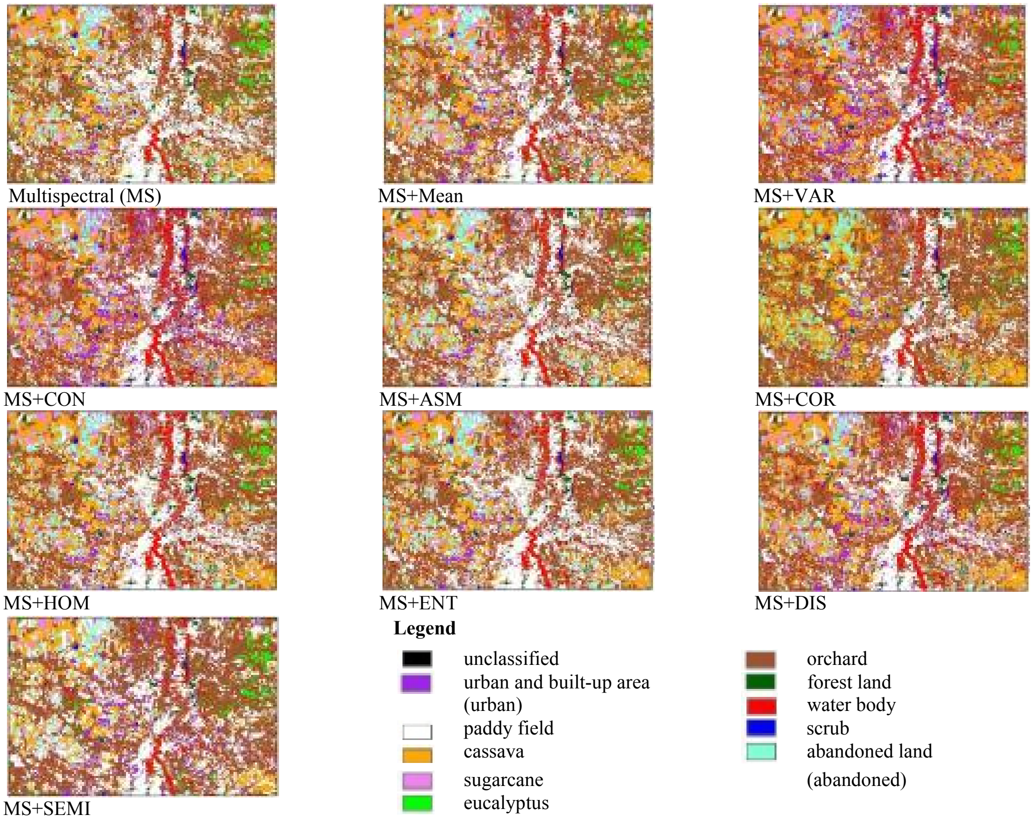

Figure 4: LULC classification from ten datasets with MLC algorithm

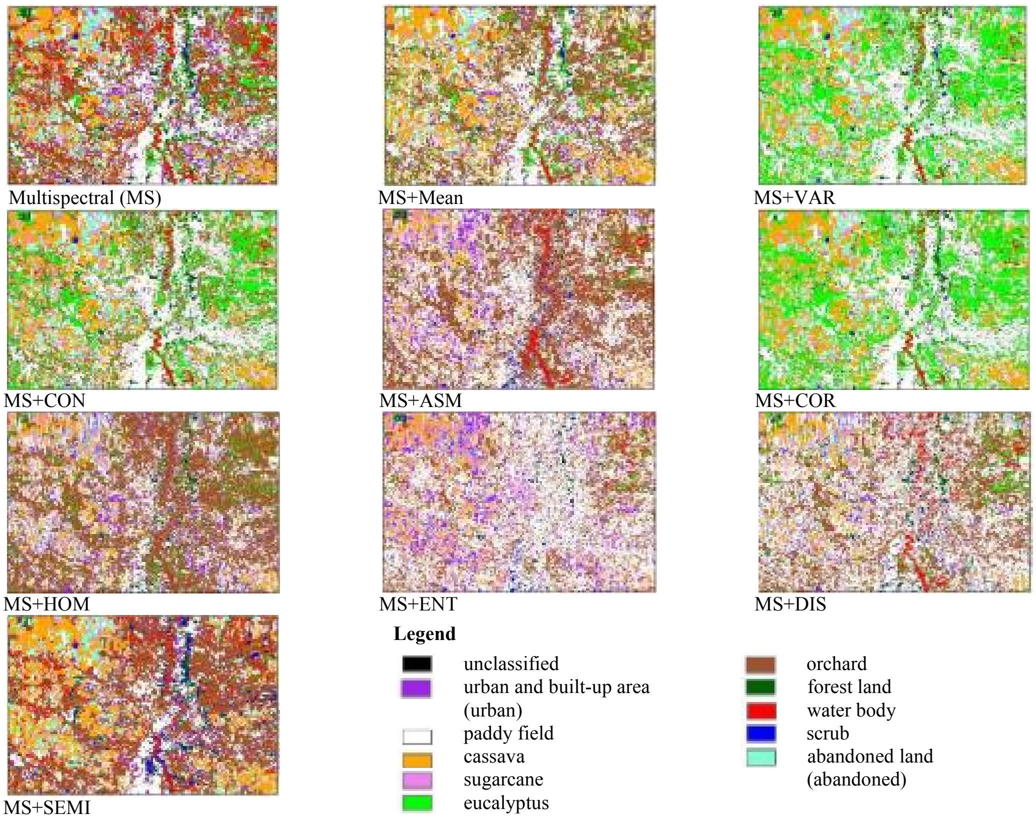

Figure 5: LULC classification from ten datasets with MLC algorithm

Table 1: The result of LULC classification of ten datasets with MLC algorithm

no.	LULC types (Rai)	Dataset MS	MS+Mean	MS+VAR	MS+CON	MS+ASM	MS+COR	MS+HOM	MS+ENT	MS+DIS	MS+SEMI
0	unclassified	73.69	65.11	78.89	133.31	124.03	99.14	68.48	66.23	75.80	51.05
1	urban	1,931.06	3,441.38	10,606.36	9,362.95	2,352.52	3,242.81	2,398.64	2,274.19	5,769.28	4,374.42
2	paddy field	27,834.47	23,543.16	16,826.20	19,156.08	28,021.50	14,887.41	26,818.45	27,591.75	21,581.86	27,976.78
3	cassava	12,499.45	12,860.30	18,502.73	15,276.66	11,246.77	19,875.23	12,389.20	11,279.95	15,932.25	8,216.44
4	sugarcane	2,321.58	4,102.17	2,780.58	3,755.67	1,965.94	0	2,617.59	2,193.47	2,874.23	1,236.66
5	eucalyptus	3,192.19	3,334.08	1,165.36	1,406.25	1,725.75	1,723.50	1,770.61	1,675.97	1,604.11	3,099.09
6	orchard	33,364.41	35,033.77	28,154.67	30,620.67	34,801.17	39,274.31	34,420.22	35,940.09	31,858.59	39,668.91
7	forest land	1,931.48	1,069.03	1,863.42	2,025.00	2,267.72	2,400.47	2,045.53	1,989.42	1,757.25	715.64
8	water body	4,692.38	5,442.33	8,760.38	7,612.03	5,171.34	4,575.66	5,660.72	4,869.00	7,373.53	5,174.86
9	scrub	284.91	265.78	603.56	653.63	286.31	189.56	303.05	278.02	494.72	217.69
10	abandoned	5,343.33	4,311.84	4,126.78	3,466.69	5,505.89	7,200.84	4,976.44	5,310.84	4,147.31	2,737.41
	total	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94

Table 2: The result of LULC classification of ten datasets with ANN algorithm

no.	LULC types (Rai)	Dataset MS	MS+Mean	MS+VAR	MS+CON	MS+ASM	MS+COR	MS+HOM	MS+ENT	MS+DIS	MS+SEMI
0	unclassified	153.28	90.14	109.41	114.19	91.55	96.61	107.16	134.72	97.17	84.52
1	urban	6,565.64	4,026.09	1,852.59	1,677.94	5,132.95	2,197.27	6,758.16	8,661.38	3,595.22	3,223.83
2	paddy field	24,456.94	33,391.13	33,235.88	32,980.36	36,869.06	32,505.75	26,470.13	50,386.92	44,648.30	21,255.89
3	cassava	9,619.88	13,025.53	13,544.02	12,469.50	5,275.41	16,384.50	5,587.59	6,300.14	4,879.69	13,802.06
4	sugarcane	1,822.08	1,368.98	2,633.91	2,504.11	3,137.20	1,626.05	2,395.41	5,038.73	2,545.59	1,552.36
5	eucalyptus	6,221.53	6,747.89	26,847.56	16,055.72	1,106.86	27,225.56	3,569.48	1,231.17	2,657.95	1,101.94
6	orchard	25,825.92	27,839.53	3,360.66	15,628.36	35,719.59	3,098.11	42,833.11	16,365.94	26,232.33	34,550.30
7	forest land	3,806.58	1,113.19	2,246.06	2,616.89	717.75	2,261.67	2,360.39	566.02	1,932.47	3,412.27
8	water body	8,126.44	1,987.17	2,625.05	2,857.64	2,729.39	2,076.33	1,250.44	2,576.53	3,092.20	6,961.22
9	scrub	568.55	220.08	358.59	277.45	1,630.55	253.83	239.20	1,160.30	683.02	2,477.25
10	abandoned	6,302.11	3,659.20	6,655.22	6,286.78	1,058.63	5,743.27	1,897.88	1,047.09	3,105.00	5,047.31
	total	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94	93,468.94

5.3 Ground Verification and Accuracy Assessment

LULC classification of ten datasets was checked by ground reference. The number of sampling points was based on multinomial distribution theory with 90% of level confident and 10% of precision by using stratified random sampling. This provided sampling points of 160. Then assess accuracy with overall accuracy (OA) and kappa hat coefficient of agreement ($\widehat{K}$).The results of accuracy assessment were presented in Table 3 and Table 4, respectively.

Table 3: the results of accuracy assessment of ten datasets with MLC algorithm

No.	Dataset	OA (%)	$\widehat{K}$ (%)	Order*
1	MS	79.38	75.96	6
2	MS+Mean	83.75	81.17	1
3	MS+VAR	75.00	71.00	8
4	MS+CON	76.88	73.22	7
5	MS+ASM	80.00	76.71	5
6	MS+COR	71.25	66.53	10
7	MS+HOM	80.63	77.41	4
8	MS+ENT	81.88	78.91	2
9	MS+DIS	80.63	77.44	3
10	MS+SEMI	73.75	69.22	9

Table 4: the results of accuracy assessment of ten datasets with ANN algorithm

No.	Dataset	OA (%)	$\widehat{K}$ (%)	Order*
1	MS	25	84.10	1
2	MS+Mean	13	62.57	3
3	MS+VAR	25	60.99	4
4	MS+CON	38	67.42	2
5	MS+ASM	13	45.32	8
6	MS+COR	25	60.96	5
7	MS+HOM	38	40.95	9
8	MS+ENT	50	38.44	10
9	MS+DIS	25	48.53	7
10	MS+SEMI	13	57.06	6

Note. *The order of accuracy of datasets that evaluate by kappa hat coefficient of agreement (Jensen, 2005)

From Table 3, accuracy assessment of ten datasets with MLC algorithm showed the overall accuracy between 71.25%-83.75% and kappa hat coefficient between 66.53%-81.17%, respectively. Herein, it was found that MS and mean combination dataset provided the highest accuracy with 83.75% of overall accuracy and 81.17% of kappa hat coefficient. These accuracy values were higher than that of accuracy from MS dataset alone at 4.37% and 5.21%, respectively. Additionally, kappa hat coefficient of agreement for each LULC types showed high accuracy in LULC classification of MS and texture measures combination datasets. This resulted in increasing accuracy of urban and built-up area, paddy field, cassava, eucalyptus, orchard and water body as shown in Table 5. This was agreed by researches from Emran, Hakdaoui, and Chorowicz, 1996; Zhang and Wang, 2001; Berberoglu, Curran, Lloyd and Atkinson, 2007; and Murray, Lucieer and Williams, 2010.

In the meantime, the results of overall accuracy and kappa hat coefficient of agreement of all datasets that using ANN algorithm were between 47.50% - 86.25% and 38.44% - 84.10%, respectively. Herein, MS dataset provided the highest overall accuracy and kappa hat coefficient with 86.25% and 84.10%, respectively (Table 4).

From Table 6, the results showed that the accuracy of paddy field, sugarcane, eucalyptus, forest land and water body were increased.

Table 5: Comparison of kappa hat coefficient of agreement for each LULC type of datasets with MLC algorithm

Land use and land cover classes (Rai)	Datasets MS	MS+Mean	MS+VAR	MS+CON	MS+ASM	MS+COR	MS+HOM	MS+ENT	MS+DIS	MS+SEMI
Unclassified	.00%	.00%	6%	7%	.00%	.00%	.00%	.00%	4%	0%
Urban	6%	8%	3%	1%	7%	2%	4%	.6%	5%	5%
Paddy field	2%	7%	0%	1%	1%	0%	2%	1%	2%	.00%
Cassava	0%	6%	1%	6%	0%	%	0%	0%	6%	7%
Sugarcane	.00%	.00%	.00%	.00%	5%	.00%	.00%	.00%	.00%	.00%
Eucalyptus	2%	4%	4%	8%	5%	5%	3%	8%	.9%	8%
Orchard	7%	6%	7%	.4%	8%	1%	6%	8%	6%	5%
Forest land	4%	7%	7%	7%	4%	4%	4%	4%	4%	4%
Scrub	.00%	.00%	.00%	.00%	.00%	.00%	.00%	.00%	.00%	.00%
Water body	3%	.00%	5%	.00%	0%	4%	0%	0%	.00%	.00%
Abandoned land	.00%	.00%	6%	7%	.00%	.00%	.00%	.00%	4%	0%
Total	6%	8%	3%	1%	7%	2%	4%	.6%	5%	5%

Table 6: Comparison of kappa hat coefficient of agreement for each LULC type of datasets with ANN algorithm

Land use and land cover classes (Rai)	Datasets MS	MS+Mean	MS+VAR	MS+CON	MS+ASM	MS+COR	MS+HOM	MS+ENT	MS+DIS	MS+SEMI
Unclassified	.00%	6%	.00%	.00%	3%	.00%	0%	.0%	1%	1%
Urban	0%	6%	4%	0%	1%	7%	8%	.1%	0%	0%
Paddy field	4%	0%	.0%	1%	7%	9%	.0%	.0%	1%	5%
Cassava	.00%	.00%	.00%	7%	7%	.00%	0%	9%	4%	.00%
Sugarcane	1%	6%	5%	9%	9%	5%	1%	.7%	1%	.00%
Eucalyptus	3%	5%	%	7%	.8%	30%	.8%	9%	3%	5%
Orchard	.00%	5%	.4%	2%	1%	1%	7%	.4%	1%	4%
Forest land	5%	2%	7%	7%	3%	7%	8%	2%	7%	4%
Scrub	.00%	.00%	.00%	%	5%	.00%	.00%	4%	.00%	6%
Water body	0%	5%	.6%	5%	1%	0%	.00%	.00%	5%	1%
Abandoned land	.00%	6%	.00%	.00%	3%	.00%	0%	.0%	1%	1%
Total	0%	6%	4%	0%	1%	7%	8%	.1%	0%	0%

Table 7: Comparison of accuracy assessment for LULC classification of ten datasets with MLC and ANN algorithm

No.	Dataset	MLC		ANN	
		OA %	$\hat{K}$ %	OA %	$\hat{K}$%
1	MS	79.38	75.96	5	84.10
2	MS+Mean	83.75	81.17	3	62.57
3	MS+VAR	75.00	71.00	5	60.99
4	MS+CON	76.88	73.22	8	67.42
5	MS+ASM	80.00	76.71	3	45.32
6	MS+COR	71.25	66.53	5	60.96
7	MS+HOM	80.63	77.41	8	40.95
8	MS+ENT	81.88	78.91	0	38.44
9	MS+DIS	80.63	77.44	5	48.53
10	MS+SEMI	73.75	69.22	3	57.06

6. DISCUSSION

For evaluation of optimum algorithm and dataset for LULC classification based on THEOS data. It was found that MLC algorithm was more suitable than ANN algorithm for MS and texture measures combination datasets and provided higher accuracy than ANN algorithm in all 9 combined datasets. However, if using only MS dataset, ANN algorithm was more suitable than MLC algorithm since it provided higher accuracy than MLC algorithm. Whereas, three best optimum dataset of LULC classifications with MLC algorithm were the combinations of MS and mean, MS and entropy, and MS and dissimilarity. These provided kappa hat coefficient at 81.17, 78.91 and 77.44 %, respectively. At the same time, three best optimum dataset of LULC classifications with ANN were MS, MS and contrast, and MS and mean. These provided kappa hat coefficient of 84.10, 67.42 and 62.57%, respectively

(Table 7). In addition, if considering of conditional kappa hat coefficient of each LULC class with MLC algorithm and ANN algorithm, it was found that the combination of texture data with MS can increase accuracy of each class. This was agreed by research from Ge, Carruthers, Gong, and Herrera, 2006.

7. CONCLUSIONS AND RECOMMENDATIONS

The best of three optimum datasets for LULC classification with MLC algorithm were the combinations of MS and mean, MS and entropy, and MS and dissimilarity. In the meantime, three best optimum dataset for LULC classification with ANN algorithm were MS, MS and contrast, and MS and mean. In addition, conditional kappa hat coefficient of each LULC class with MLC algorithm and ANN algorithm can be increased with using combination of texture data with multispectral data dataset.

In conclusion, applying texture measures with multispectral data of THEOS can increase the accuracy of LULC classification, especially using MLC algorithm.

REFERENCES:

Berberoglu, S., Curran, P.J., 2006. Merging Spectral and Textural Information for Classifying Remotely Sensed Images. Quoted in De Jong, S.M.; Van der Meer, F.D., (eds.). Remote Sensing Image Analysis: Including The Spatial Domain, Springer Netherlands, pp. 113-136.

Berberoglu, S., Curran, P.J.; Lloyd, C.D. and Atkinson, P.M., 2007. Texture Classification of Mediterranean Land Cover. International Journal of Applied Earth Observation and Geoinformation. 9, pp. 322-334.

Emran, A., Hakdaoui, M. and Chorowicz, J., 1996. Anomalies on Geologic Maps from Multispectral and Textural Classification: The Bleida Mining District (Morocco). Remote Sens. Environ. 57, pp. 13-21.

313, pp. 1-12.

Ge, S., Carruthers, R., Gong, P. and Herrera, A., 2006. Texture Analysis for Mapping Tamarix Parviflora Using Aerial Photographs Along The Cache Creek. Environmental Monitoring and Assessment. 114, pp. 65-83.

Jensen, J.R., 2005. Introductory Digital Image Processing: A Remote Sensing Perspective (2nd ed.), New Jersey: Prentice Hall. Remote Sens. Environ. 112, pp. 1895–1908.

Joshi, P.K., Roy, P.S., Singh, S., Agrawal, S. and Yadav, D., 2006. Vegetation Cover Mapping in India Using Multi-Temporal IRS Wide Field Sensor (WiFS) Data. Remote Sens. Environ. 103, pp. 190–202.

Landgrebe, D., 2002, Hyperspectral Image Data Analysis, IEEE Signal Process. Mag., 19, pp. 17-28.

Lobe, A., 1997. Image Segmentation and Discriminant Analysis for the Identification of Land Cover Units in Ecology, IEEE lkansactions on Geoscience and Remote Sensing, 35, pp. 1136-1145.

Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance, International Journal of Remote Sensing, 28(5), pp. 823-870.

Murray, H., Lucieer, A. and Williams, R., 2010. Texture-Based Classification of Sub-Antarctic Vegetation Communities on Heard Island. International Journal of Applied Earth Observation and Geoinformation.

Perumal, K. and Bhaskaran, R., 2010. Supervised Classification Performance of Multispectral Images, Journal of Computing, 2(2), pp. 6188-6202. Quoted in Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance, International Journal of Remote Sensing, 28(5), pp. 823-870,

Powell, S.L., Cohen, W.B., Yang, Z., Pierce, J.D. and Alberti, M., 2008. Quanti cation of Impervious Surface in the Snohomish Water Resources Inventory Area of Western Washington from 1972-2006, Remote Sens, Environ, 112, pp. 1895–1908.

Reis, S., 2008. Analyzing Land Use/Land Cover Changes Using Remote Sensing and GIS in Rize, North-East Turkey. Sensors. 8, pp. 6188-6202. Quoted in Rogana J.; Chen, D., 2004. Remote Sensing Technology for Mapping and Monitoring Land-Cover and Land-Use Change. Progress in Planning, 61, 301–325.

Samaniego, L. and Schulz, K., 2009. Supervised Classification of Agriculture Land Cover Using a Modified k-NN Technique (MNN) and Landsat Remote Sensing Imagery, Remote Sensing. 1, pp. 875-895.

Zhang, Q., and Wang, J., 2001. Texture Analysis for Urban Spatial Pattern Study Using SPOT Imagery. In Geoscience and Remote Sensing Symposium, 2001. IGARSS 01. IEEE 2001 International. (1-3) Sydney: NSW.