

PIXEL-BASED AND OBJECT-BASED TERRACE EXTRACTION USING FEED-FORWARD DEEP NEURAL NETWORK

H. T. Do^{1*}, V. Raghavan², G. Yonezawa²

¹ Graduate School for Creative Cities, Osaka City University, Japan – dohanghumg@gmail.com

² Graduate School of Engineering, Osaka City University, Japan – (raghavan, yonezawa@media.osaka-cu.ac.jp)

Commission III, WG III/10

KEY WORDS: Terrace field, feed forward, deep learning, remote sensing, object-based

ABSTRACT:

In this paper, we present the identification of terrace field by using Feed-forward back propagation deep neural network in pixel-based and several cases of object-based approaches. Terrace field of Lao Cai area in Vietnam is identified from 5-meter RapidEye image. The image includes 5 bands: red, green, blue, reledge and nir-infrared. Reference data are set of terrace points and non-terrace points, which are generated by randomly selected from reference map. The reference data is separated into three sets: training set for training processing, validation set for generating optimal parameters of deep neural network model, and test set for assessing the accuracy of classification. Six optimal thresholds (T): 0.06, 0.09, 0.12, 0.14, 0.2 and 0.22 are chosen from Rate of Change graph, and then used to generate six cases of object-based classification. Deep neural network (DNN) model is built with 8 hidden layers, input units are 5 bands of RapidEye, and output is terrace and non-terrace classes. Each hidden layer includes 256 units – a large number, to avoid under-fitting. Activation function is Rectifier. Dropout and two regularization parameters are applied to avoid overfitting. Seven terrace maps are generated. The classification results show that the DNN is able to identify terrace field effectively in both pixel-based and object-based approaches. Pixel-based classification is the most accurate approach, achieves 90% accuracy. The values of object-based approaches are 88.5%, 87.3%, 86.7%, 86.6%, 85% and 85.3% correspond to the segmentation thresholds.

1. INTRODUCTION

Terrace is the specific type of landform that is constructed in the form of sections on sides of hills or mountains to grow crops into the slope. The condition of terraces is important for controlling soil and water loss, intercept runoff and sediment, increase soil moisture, maintain soil fertility, and significantly increase grain yield (Zhang et al., 2017). Also, it ensures the conversion of cropland to forest and grassland in current agriculture ecology and ecological agricultural construction (Ma et al., 2007).

Extracting terrace field area from remote sensing data is reasonable since field surfaces of terraces in the data have high-brightness because they are exposed to the ground. Based on the spatial resolution of remotely sensed images, terrace field is displayed in different levels of detail. High resolution RapidEye's imaging capabilities are able to capture terrace surface as separate objects to other Land Use/ Land Cover (LULC) classes.

There are two spatial levels of remote sensing image analysis: object-based and pixel-based. Pixel-based is the traditional analysis which operates directly on individual pixels. Object-based analysis has been gaining importance in the fields of remote sensing, especially for high spatial resolution image processing (Blaschke, 2010) which distinguishes on groups of contiguous pixels, allows exploiting spectral-spatial data (Benz

et al, 2004, Van der Werff et al., 2008, Benz, U.C. et al., 2004, Wuest et al., 2009, Gamanya et al., 2009). However, segmentation process may cause loss information in comparison of original data. Image objects are generated by image segmentation process in which scale parameter is a key parameter to partition the image into objects. Optimal scale of image segmentation can be taken from ROC-LV graph (Dragut et al., 2014).

Deep neural network (DNN) algorithms which learn the representative and discriminative features in a hierarchical manner from the data have been applied for remote sensing data analysis, including LULC (Zhang, Li, 2016). DNN has powerful computing capabilities base on the propagation of information between neurons. Due to the information transferring direction of neural, DNN can be divided into two categories: feed-forward DNN and feedback DNN. Feed-forward neural network model has been widely used in many fields due to its ability to estimated complex nonlinear mappings directly from the input data.

From all above, this paper adopts the feed-forward deep neural network to identify terrace field in Lao Cai - a mountainous area in Vietnam in both pixel-based and object-based approaches.

2. STUDY AREA AND DATA

* Corresponding author

The remotely sensed data used is RapidEye image acquired on 9th September 2014 covering 524 km² of Lao Cai province - a northern mountainous region of Vietnam. The 5- meter resolution imagery is optical multispectral imagery with five distinct bands: red, green, blue, nir-infrared and RedEdge. Terrace field of the study area mostly locates in the southeast. A part of terrace field of the study area is continuously covers a wide area (Figure 1a), other small terrace fields scatter along some main roads (Figure 1b). Reference data includes 22682 terrace field sample points and 77755 other layers sample points, are randomly collected from forest map in scale 1:10000 (Ministry of Agriculture & Rural Development, Vietnam) and Google map. The reference data is separated into three sets: training set (75%); validation set (12.5%) and test set (12.5%) for deep learning classification.

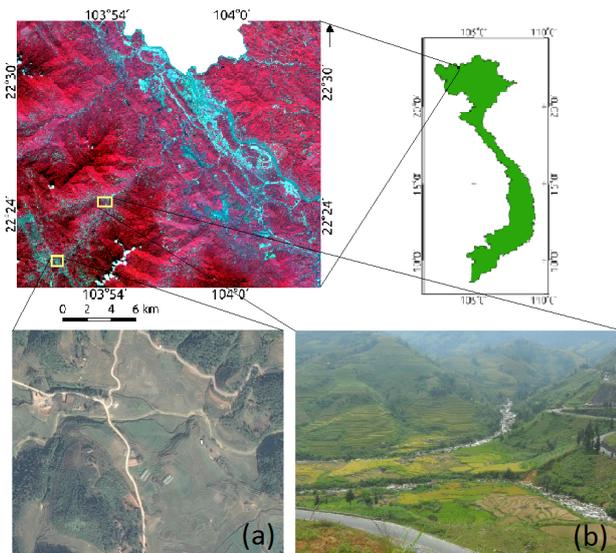


Figure 1. (a) a wide (Fa) and (b) a small (Fb) terrace field in Lao Cai area

3. METHODOLOGY

In this study, both pixel-based and object-based are approached for DNN classification to identify terrace area and other areas (includes other LULC classes, cloud and cloud shadow). Different levels of segment are distinguished using six thresholds which are selected from ROC-LV graph. Feed-forward back propagation neural network is employed with inputs are five bands of RapidEye image, eight hidden layers, and two output layers: terrace and non-terrace. Other parameters of DNN are also defined. Terrace field in pixel-based approach and six levels of object-based approaches are mapped.

3.1 Segmentation

Image segmentation which possesses intrinsic size, shape and geographic relationship with the imprecise nature of image data is a frequently used technique in remote sensing processing (Hay et al., 2001). "Region growing and merging" segmentation algorithm contains two processes: region growing which groups neighboring similar pixels into regions, and region merging which merges similar neighboring regions due to segmentation threshold value (Haralick et al., 1981). The threshold value

which must be smaller than 1 and larger than 0, defines the scale of segmentation.

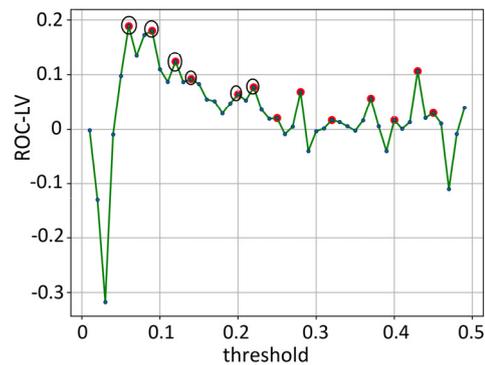


Figure 2. ROC-LV graph. Threshold values: 0.06, 0.09, 0.12, 0.14, 0.20, 0.22 are selected as segmentation thresholds.

Local Variance (LV) is the mean of the value of standard deviation (SD) in a small neighborhood over the entire image (Woodcock et al., 1987). The value shows the relationship between images spatial, the size of the objects in the real world and pixel resolution: if the spatial resolution is significantly finer than the objects in the scene, most of the SD in the image will be correlated highly with their neighbors and LV value will be low; if the objects approximate the size of the resolution cells, then the likelihood of neighbors being similar decreases and the LV value rises. In the case of object-based analysis, LV is defined by SD value of pixels inside segment (Kim et al., 2008).

To assess the dynamics of LV value of different segmentation levels, Rate of Change of Local Variance (ROC-LV) (Bauer et al., 1998) measurement is used:

$$ROC - LV = \frac{100 * (LV_i - LV_{i-1})}{LV_{i-1}} \quad (1)$$

Where: i is value of segmentation threshold, LV_i, LV_{i-1} are LV value at a given level and previous level, correspondingly.

Peaks in the ROC-LV graph show the object levels at which the image can be segmented in the most appropriate manner and the segments match the types of objects characterized by equal degrees of homogeneity (Dragut et al., 2010).

In this study, ROC-LV graph (Figure 2) is generated from object-based of RGB image of the study area. The graph produced in PyGRASS - an object-oriented Python application programming interface for GRASS GIS by the authors. Values of threshold i range from 0.01 to 0.5, step is 0.01. The peaks (red points) of the graph should be considered as optimal threshold. Since value of threshold increases, more number of neighbour pixels is grouped in a segment, leads to under-segmentation phenomenon. In this study, since the phenomenon is clearly shown at T = 0.22 (Figure 3), we choose segmentation at six thresholds: 0.06, 0.09, 0.12, 0.14, 0.20, 0.22, and pixel-based to introduce seven DNN classification cases. Five bands of RapidEye data in object-based which are calculated by averaging value of pixels inside each segment are exploited as input of DNN classification.

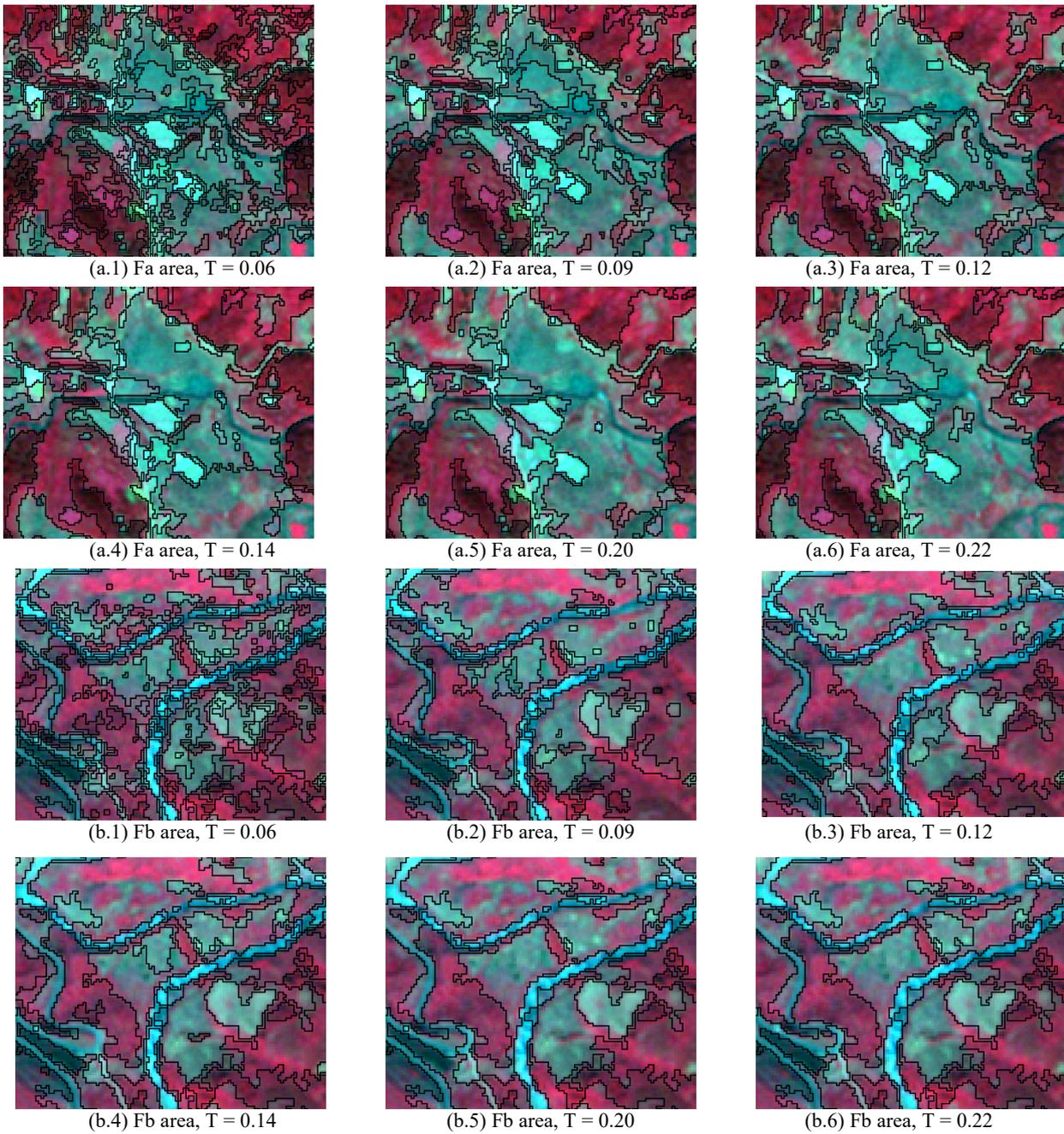


Figure 3. Segmentation boundaries of Fa and Fb areas in six selected thresholds

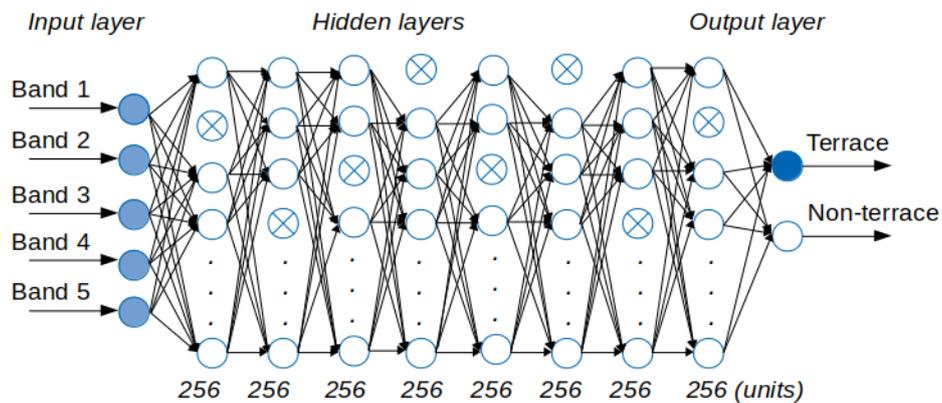


Figure 4. DNN model. In each hidden layer, 20% of units temporal randomly dropped

3.2 Feed-forward DNN classification

Feed-forward back propagation neural network, also known as multilayer perceptron (MLP) is one of the most popular neural network models. The structure of a feed-forward neural network includes three types of layers: input layer, hidden layer, and output layer. In the network, information moves in only one direction, forward, from the input nodes, through the hidden nodes, and to the output nodes. Back-propagation (BP) algorithms involves two phases (Werbos 1974; Rumelhart et al. 1986): forward phase in which free parameter of the network is fixed, the input signal is propagated through the network and computes error signal, and backward phase in which error signal is propagated through the network and the fixed parameters are adjusted to minimize the error signal. The backpropagation algorithm minimises error function in weight space using stochastic gradient descent (SGD).

SGD is a stochastic approximation of the gradient descent optimization and interactive method for minimizing error function (Botou, 1998). In a simple supervised learning setup, each example z is a pair (x, y) where x is arbitrary input and y is scalar output. The empirical risk $E_n(f)$ measures the performance of training set is calculated by following:

$$E(f) = \int \ell(f(x), y) dP(z) \quad E_n(f) = \frac{1}{n} \sum_{i=1}^n \ell(f(x_i), y_i) \quad (3)$$

Where:

ℓ : loss function

$f(x)$: function parameterized by a weight vector w

f : function which seeks to minimize ℓ

$dP(z)$: unknown distribution that embodies the Laws of Nature.

The expected risk $E(f)$ determines the generalization concert. SGD algorithm is the drastic simplification which is not computing the gradient of $E_n(f)$ precisely, each iteration estimates this gradient based on a single randomly selected example z :

$$w_{t+1} = w_t - \gamma_t \nabla_w Q(z_t, w_t) \quad (4)$$

The stochastic process $\{w_t, t = 1 \dots\}$ depends on the random examples at each interaction. Since the stochastic algorithm does not need to remember which examples were visited through the previous iterations, it optimizes the expected risk directly, since the examples are randomly drawn from the reference distribution (Bottou, L., 2010).

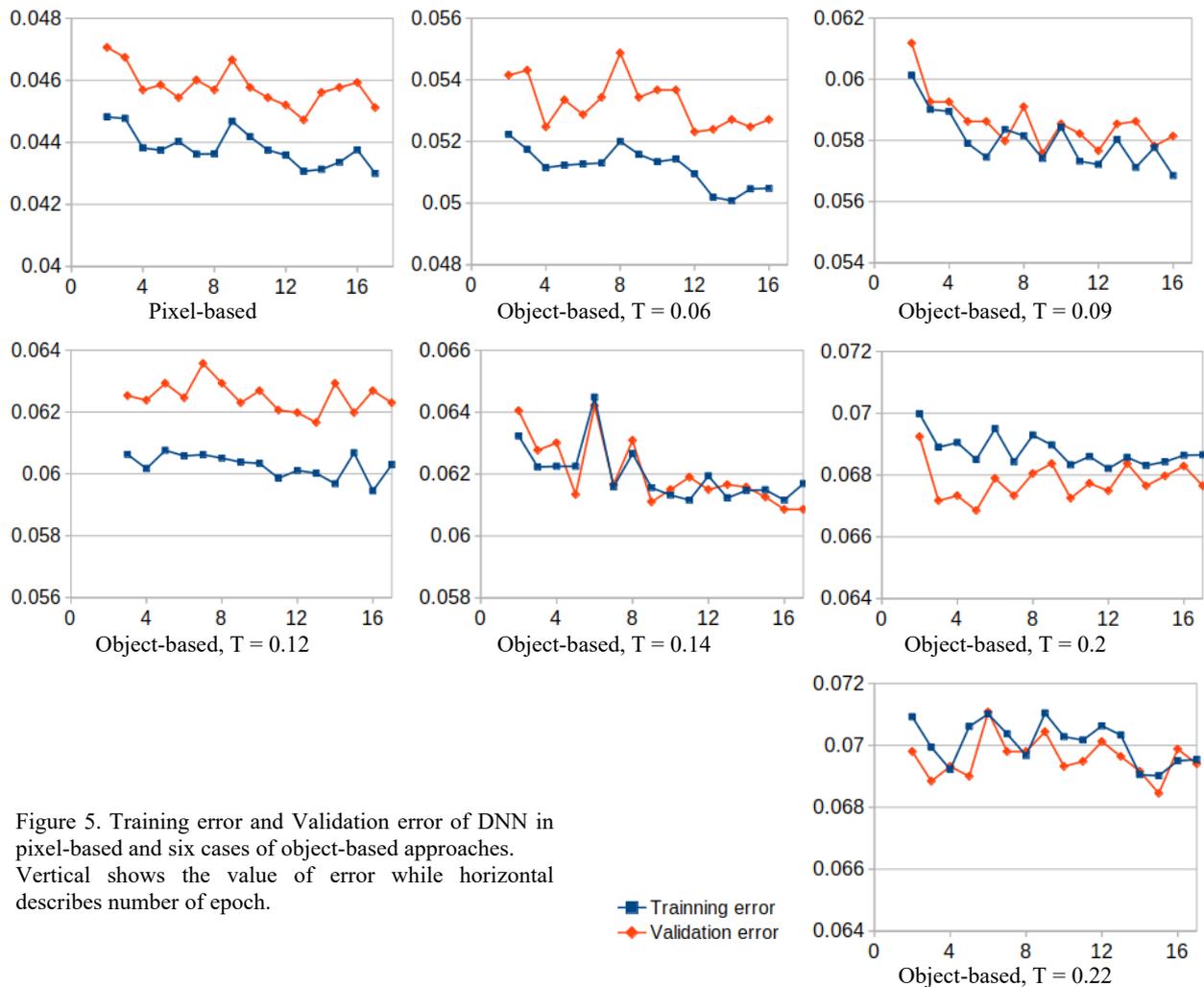


Figure 5. Training error and Validation error of DNN in pixel-based and six cases of object-based approaches. Vertical shows the value of error while horizontal describes number of epoch.

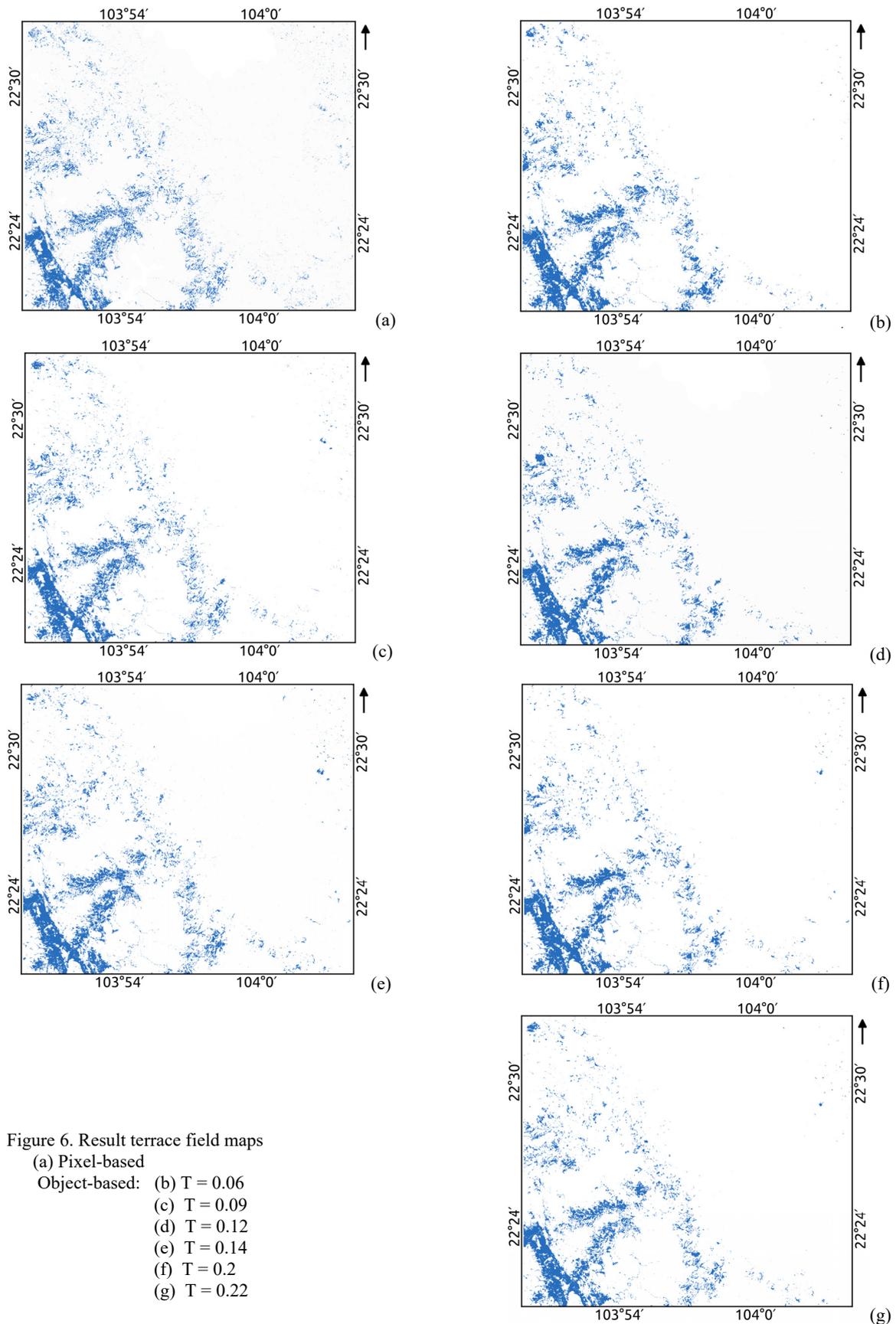


Figure 6. Result terrace field maps

(a) Pixel-based

Object-based: (b) $T = 0.06$
(c) $T = 0.09$
(d) $T = 0.12$
(e) $T = 0.14$
(f) $T = 0.2$
(g) $T = 0.22$

In DNN, output behaviour of each node is set by activation function which introduces non-linear properties to the network (deepai.org) and allows the DNN to be able to learn from complicated, non-linear mappings between inputs and consisting variables. Some widely used activation functions are: tanh, logistic, softsign, maxout, rectifier, etc. Rectifier, the most popular activation function for DNN in 2017 (Ramachandran et al., 2017), is defined as:

$$f(x) = x^+ = \max(0, x)$$

Where: x: input to a neuron.

To prevent overfitting, dropout and regularization are two popular ways. The term "dropout" refers to dropping out units (hidden and visible) in a DNN. Dropout decreases the number of neurons by temporarily removing the units during a particular forward or backward pass. These units are chosen randomly, number of them is defined through dropout ratio parameter (Hinton et al., 2012, Srivastava et al., 2014). Regularization reduces over-fitting by adding a penalty to the error function that adds stability and improves generalization. Two regularizations are Laplacian (L1) and Gaussian (L2).

In this study, DNN network is used is feed-forward back propagation neural network, input layer is the average of DN in a segment of object-based of five bands of RapidEye image, output layer is two classes: terrace and non-terrace, 8 hidden layers which include 256 units for each layer. Rectifier with dropout is chosen as activation function with dropout ratio of hidden layers equals 0.2 (20% random units of each hidden layer are dropped out of learning processing). L1 and L2 are set as 10^{-5} . Other parameters also are defined such as epoch which is the number of times to iterate the dataset is set as 40; number of folds is 50, fold assignment is Stratified. 75% of reference set is used as training data and other 12.5% is utilized as validation set. The model is trained 40 times (iteration) corresponds to 40 epochs. Logloss metric is exploited to decide the best model it can be reached. In most cases, the best models are achieved after 17 or 18 epochs. Figure 5 shows training and validation errors of these epochs.

4. RESULTS AND DISCUSSIONS

Terrace maps in seven classification approaches are shown in Figure 6. There are two types of misclassification: terrace samples of test set (real terrace) are not identified as terrace and non-terrace are categorized as terrace. User's accuracy and producer's accuracies correspond to these misclassification types respectively:

$$\text{User's accuracy} = \frac{\text{Correctly identified terrace}}{\text{Real terrace}} * 100\%$$

$$\text{Producer's accuracy} = \frac{\text{Correctly identified terrace}}{\text{Identified terrace}} * 100\%$$

Test set which is used for accuracy assessment is the random selection of 7% of reference data set. 2867 real terrace points are extracted from the test set. Otherwise, numbers of identified terrace points are different because they depend on classification maps.

In order to evaluate each classification approach, overall accuracy is generated by averaging user's and producer's accuracies, calculated as follows:

$$\text{Overall accuracy} = \frac{\text{User's accuracy} + \text{Producer's accuracy}}{2}$$

Pixel-based approach reaches 90% - the highest value of user's accuracy, followed by object-based threshold 0.06 (89.2%) and 0.14 (85.4%) (Table 1). Object-based threshold 0.2 produces the most inaccuracy (81.2%). Pixel-based classification also attains high producer's accuracy (89.9%), equal object-based threshold 0.12, and less than only object-based threshold 0.09 (90.5%). Object-based threshold 0.22 is the most inaccuracy case (87.5%).

By considering overall accuracy values, it performs that feed-forward back propagation DNN is able to identify terrace field accurately, even though there are some differences between pixel-based and cases of object-based approaches. Pixel-based classification is the most accurate approach, achieves 90% accuracy. The values of object-based approaches are 88.5%, 87.3%, 86.7%, 86.6%, 85% and 85.3% correspond to 0.06, 0.09, 0.12, 0.14, 0.2 and 0.22 of segmentation thresholds.

In object-based approach, by averaging value of pixels in each segment, original image information is lost in object-based classification (Bartesaghi et al., 2005). In addition, in this study, object-based classification do not distinguish any spatial detail such as shape, length, structure, etc., causes to failing to exploit the advantage of object-based approach.

The accuracies of object-based approach decrease when segmentation threshold values increase. The reason is over-segmented areas (neighbor segments are same class) are able to be merged into desired objects in the classification processing but under-segmented areas cannot (Neubert et al., 2008). At T = 0.06 (Figure 3a.1, 3b.1), mostly there is no under-segmented area. At T = 0.09 (Figure 3a.2, 3b.2), both under-segmented and over-segmented areas exist. When T increases, under-segmentation phenomenon happens critically, especially at T = 0.22 (Figure 3a.6, 3b.6). It suggests that the first peak of ROC-LV graph should be the optimal threshold of segmentation.

5. CONCLUSION

In this paper we present the identification of terrace area by using Feed-forward back propagation DNN in both pixel-based and object-based approach. Optimal thresholds are chosen from ROC-LV graph. The DNN model is built with 8 hidden layers which have 256 units for each layer, dropout rate equals 0.2, 5 bands of RapidEye are used as input units, and output is terrace and non-terrace classes. The classification results show that the DNN is able to identify terrace field effectively in both pixel-based and object-based approaches, overall accuracies of all study cases are higher than 85%, and pixel-based is the best accurate approach.

To ignore under-segmentation phenomenon, the first peak of ROC-LV should be optimal threshold in case none of spatial information such as shape, length, structure is distinguished in classification processing. Since boundary of terrace area is quite clear, we suggest discriminating spatial information such as shape, length, structure, etc. to gain the advantage of object-based approach. Moreover, a scale level of object-based may not be suitable for all terrace areas, it would be an idea to divide

a study region into smaller areas and select optimal threshold of these areas.

Table 1. Accuracies of terrace classification (unit: sample point)

		Correctly identified terrace	Identified terrace	Real terrace	User's accuracy (%)	Producer's accuracy (%)	Overall accuracy (%)
Pixel-based		2580	2870	2867	90.0	89.9	90.0
Object-based T =	0.06	2557	2883		89.2	88.7	88.5
	0.09	2406	2658		84.1	90.5	87.3
	0.12	2386	2653		83.2	89.9	86.7
	0.14	2447	2789		85.4	87.7	86.6
	0.20	2328	2624		81.2	88.7	85.0
	0.22	2379	2718		83.0	87.5	85.3

REFERENCES

- Bartasaghi, A., Sapiro, G., 2005. An energy-based three-dimensional segmentation approach for the quantitative interpretation of electron tomograms. *IEEE Trans Image Process*, 14(9), 1314-1323.
- Bottou, L., 1998. *Online Algorithms and Stochastic Approximations*. Cambridge University Press, ISBN 978-0-521-65263-6.
- Bottou, L., 2010. Large-scale machine learning with Stochastic gradient descent. *Proceedings of COMPSTAT'2010*, Pyhisca-Verlag, 178-179p.
- Curtis, E.W., Alan, H. S. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3), 311-332.
- Deepai.org. Activation Function. <https://deepai.org/machine-learning-glossary-and-terms/activation-function>. Retrieved on 30th October 2018.
- Fu, K.S., Mui, J.K. (1981). A survey on image segmentation. *Pattern Recognition*, 13, 3-16.
- G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R.R. Salakhutdinov, 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv*.
- Haralick, R.M., Shapiro, L.G. (1985). Image segmentation techniques. *Computer vision, graphics, and image processing*, 29, 100-132.
- Hay, G.J., Marceau, D.J., Dube, P., and Bouchard, A. (2001). A multi-scale framework for landscape analysis: Object-specific analysis and up scaling. *Landscape Ecology*, 16(6), 471-490.
- Kim, M., Madden, & M., Warner, T. (2008). Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. In: T. Blaschke, S. Lang, and G.J. Hay, eds. *Object-based image analysis—spatial concepts for knowledge driven remote sensing applications*. Berlin: Springer, 291–307.
- Ma, B. H., Xue, X. X., & Liu, Y., 2007. Study on role of terraces and warping dam in ecological recovery of loess plateau. *Research of Soil and Water Conservation*, 14(4), pp. 27–30.
- Neubert, M., Herold, H., Meinel, G. (2008). Assessing image segmentation quality - concepts, methods and application. In: Blaschke T., Lang s., Hay G.J. (eds) *Object-Based Image Analysis. Lecture Notes in Geoinformation and Cartography*. Springer, Berlin, Heideberg.
- Srivastava, N., Hinton, H., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15, pp. 1929 – 1958.
- Rumelhart, D.e., G.E. Hinton, r.J. Williams, 1986. Learning internal representations by error propagation, in D.E. rumelhart and J. L. McClelland, eds, (Cambridge, MA: MIT Press), vol. 1, chapter 8.
- Ramachandran Prajit, Barret Zoph, Quoc, V. Le, 2017. Searching for Activation Functions. *arXiv:1710.05941*.
- Werbos, P.J., 1974. Beyond regression: New tools for prediction and analysis in the behavioural sciences. Ph.D Thesis, Harvard University, Cambridge, MA.
- Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3), 311-332.
- Zhang, Y., Shi, M., Zhao, X., Wang, X., Luo, Z., Zhao, Y., 2017. Methods for automatic identification and extraction of terraces from high spatial resolution satellite data (China-GF-1). *International Soil and Water Conservation Research*, 5, pp.17–25.
- Zhang, Li., Zhang, Le., Du, B. (2016), Deep Learning for remote sensing data. *IEEE Geosciences and remote sensing magazine*, 0274-6638/16.