TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 19, No. 4, August 2021, pp. 1242~1250
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v19i4.18359

Classification using semantic feature and machine learning: Land-use case application

Hela Elmannai[1], Abeer Dhafer Al-Garni[2]
[1,2]Department of Information Technology, College of Computer and Information Sciences,
Princess Nourah bint Abdulrahman University, Saudi Arabia
[1]Laboratoire de Teledetection et Remote Sensing LTSIRS, ENIT, Tunisia

Article Info

Article history:

Received Aug 15, 2020
Revised Feb 12, 2021
Accepted Feb 26, 2021

Keywords:

Convolutional neural networks
Feature extraction
Land-use classification
Machine learning

ABSTRACT

Land cover classification has interested recent works especially for deforestation, urban are monitoring and agricultural land use. Traditional classification approaches have limited accuracy especially for non-heterogeneous land cover. Thus, using machine may improve the classification accuracy. The presented paper deals with the land-use scene recognition on very high-resolution remote sensing imagery. We proposed a new framework based on semantic features, handcrafted features and machine learning classifiers decisions. The method starts by semantic feature extraction using a convolutional neural network. Handcraft features are also extracted based on color and multi-resolution characteristics. Then, the classification stage is processed by three learning machine algorithms. The final classification result performed by majority vote algorithm. The idea behind is to take advantages from semantic features and handcrafted features. The second scope is to use the decision fusion to enhance the classification result. Experimentation results show that the proposed method provides good accuracy and trustable tool for land use image identification.

Corresponding Author:

Hela Elmannai
Department of Information Technology, College of Computer and Information Sciences
Princess Nourah bint Abdulrahman University
168 University Road, Airport Road, King Khalid International Airport, Riyadh, Saudi Arabia
Email: hselmannai@pnu.edu.sa, adalqarni@pnu.edu.sa

1. INTRODUCTION

Classifying the earth surface is a relevant application of high-resolution imagery. The goal is to find a semantic interpretation for further studies and analysis. Reaching good classification accuracy and an automatic interpretation is within the research goals in the remote sensing field [1], [2]. Land use and land cover data (LULC) have major interest for earth observation applications. Available data requires efficient classification tasks for urban planning, environment management, climate change monitoring and other socioeconomic challenges. Recent high resolution remotely sensed imagery and constantly land cover is challenging and have motivated many researches. Significant efforts have been made for accurate classification using very high resolution and very fine spatial resolution. Traditional approaches were based on pixelwise classification which produced limited accuracy due to the speckle noise effect. Feature extraction-based approaches have contributed to improve the classification results by including contextual information and texture as spatial pattern [3]-[5]. Object oriented approaches have arisen to allow more classes characterization by including specific features [6]. However, the classification performance depends on the feature extraction engineering. At the same time, the spatial pattern of the land cover is heterogenous and variable. Therefore, these approaches may fail for different land cover dataset and different resolutions.

To faster and enhance the automatic classification and to get advantages from the huge amount of data provided by high resolution technology, using machine learning and deep learning is inherently needed. In fact, the increase of resolution and scene complexity have increased the classification complexity and the curse of dimensionality problem [7], [8]. Machine learning and convolutional neural networks (CNN) have been widely used within the image processing frameworks in biometrics, remote sensing, object detection and similar pattern recognition problems. Machine learning theory have provided the ability of analyzing large amounts of data and high dimensionality. In [9], authors proposed a classification method for LULC datasets based on random forest approach with specific preprocessing approaches. Neural networks, support vector machine (SVM) and decision tree are also widely used for remotely sensed image classification [3], [4].

CNN have been used either for feature extraction of for classification problems. The extracted features provided a higher level of data description called as semantic features. Using these features for classifiers have proven to be efficient more than using handcrafted features. The CNN architecture can effectively reduce the number of training parameters and the network complexity. The obtained model is invariant to image transformations and has strong tolerance which encourages the transfer learning operating: training on a set of data and migrate the knowledge to another dataset. CNN have achieved successful results in land use image classification [10], [11]. Deep learning tends to overcome the feature engineering limit representation. The capability to learn more representative and discriminative features have attracted the pattern recognition community and related disciplines. In [12] and [13], deep learning performs high level feature extraction and classification by fully connected layers. Specifically, in [13] an object oriented deep convolutional neural networks (DCNN) is presented. Using transfer learning to empower the learning process have allowed to utilize features extracted form large dataset in a source domain to classify different and related datasets in a target domain. For instance, for urban land use classification, authors proposed in [14] a classification method based on high-level features obtained by transfer learning form ImageNet dataset. Three fully connected layers were applied for the classification task. Main existing approaches for land use classification are resumed in the following Table 1.

Table 1. Main features and classifiers for land use and land cover classification

Feature extraction method	Classification method	Application
Texture and spatial metrics [3]	Fisher linear discriminant	Urban land use classification
geometrical, textural, and contextual information [4]	Decision tree, Neural network, Majority rule-based naive model,	Urban land use classification
Spectral indices (NDVI, MNDWI) [5]	random forests, SVM, Extreme gradient boosting, Deep learning	complex mixed-use landscape classification
Object properties: shape, texture, color [6]	mean-shift-based multi-scale segmentation	Multi-scale image segmentation
Textual features and spectral indices [9]	Random Forest	Land Use and Coverage Area frame Survey
Semantic features extracted from Deep convolutional neural networks [12]	contextual-based convolutional neural network with deep architecture and pixel-based multi-layer perceptron neural network (MLP) rule-based decision fusion approach	Very fine spatial resolution (VFSR) remotely sensed imagery classification
Semantic features extracted form deep convolutional neural networks [13]	different CNN-based models	remote sensing land use classification
Transfer learning using ImageNet dataset [14]	Fully connected layers classifiers	street view images classification

In this paper, we propose a land-use classification method based on both handcrafted feature and high-level semantic feature adopting a deep transfer network. The classification is performed by multi-classifiers and a decision fusion classification. Using deep convolutional neural networks (DCNN) in remote sensing and getting advantages from both handcrafted and high-level features simultaneously seems attractive to investigate whether the two kinds of features should improve the classification results for land cover classification. The contributions of our work are:

– Using high level features extracted deep learning
– Using decision fusion to enhance the classification result by majority vote.

The methodology is detailed in section 2. Section 3 details the feature extraction and classification principles. We present the experimentation setup and results for real land use dataset in section 4. Finally, we draw our conclusion and future works.

2. RESEARCH METHOD

The presented work aims to provide reliable pattern recognition method for high resolution overhead imagery. The method is based on convolutional neural networks for semantic feature extraction. AlexNet have been widely used int the pattern recognition and classification problem since the ImageNet large scale visual recognition challenge (ILSVRC) [15]. The DCNN have proven to efficient in transfer learning for feature extraction as shown is [16]-[19]. The CNN presents 8 layers and performs a 1000 objects classification. The features are extracted from the fully connected layer with 4096 neurons.

The second category of features are extracted with the Gabor filters. The reason behind using the basic Gabor filters is their abilities for texture and multiresolution image classification. The classification is performed by three machine learning algorithms for image classification and majority vote for classification fusion. Thus, the method is named classifiers-majority vote (C-MV).

The flowchart of the approach is detailed bellow in Figure 1. It mainly consists of three steps:

– Feature extraction based high-level semantic features and handcrafted features. Deep features extraction is performed by AlexNet networks. A deep transfer network is used to fine-tune the AlexNet Networks so that the complexity of the model is reduced.
– The learning classification using: support vector machine, multilayer perceptron and K-nearest neighbors. The majority vote classifier C-MV combines the classifiers results to generate the final classification.
– Models comparison and analysis based on the overall accuracy and the land categories accuracies to highlight the feature fusion and the multi-classifiers contributions in improving the classification performance.

Machine learning has experienced two stages: surface learning and deep learning. Using surface learning approaches relies on the extracted features. In order to have more reliable and precise features, deep learning approaches provide low-levels that emphases the high-level features through the hidden layers. In this framework, convolutional neural networks (CNNs) are widely used for image recognition where the extracted high level features are the pattern extracted form low level features. Each higher layer extracted more prominent features from previous layers. Obtained final features are called high-level semantic features. Meaning transfer learning empowers the feature extraction process by learning from large datasets which reduce the model complexity and fine-tune the model.

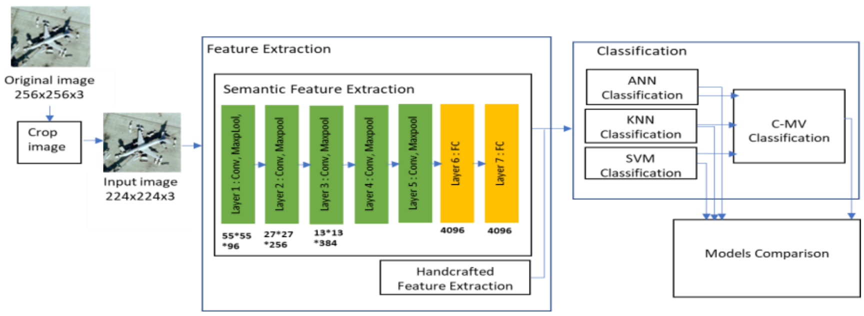


Figure 1. The architecture of the proposed method

2.1. Semantic features

CNNs have the advantages of less training parameters while providing robustness and high performance. That the features obtained by learning can transcend handcrafted feature as demonstrated in several researches and datasets [12]-[14]. AlexNet architecture includes multiple hidden layers: an input layer, five convolutional layers, first, second and fifth of which are followed by pooling layers (3 layers), three fully connected layers, and an output layer. The rectified linear units (ReLU) non linearity, data augmentation and dropout have allowed a fast training and an overfitting reduction. Hence, AlexNet has been considered as a deep convolutional architecture baseline. To allow a deep feature extraction meaning AlexNet, we remove the final fully connected layer (softmax layer) for the pretrained model and considerate the remaining architecture as deep feature extractor. Two groups of features are mixed in the first and second fully connected layers coming from two channels. The obtained feature vector has 4096 components. The 4096 vector obtained from the second fully connected layer (seventh layer) is considered as high-level semantic feature vector. For the input, the original image is cropped to obtained an input image 224*224*3. The cropping operation will reduce the original image to fit the AlexNet input without loosing information nor changing the image category or degrading the spatial resolution. We apply also a ReLU transform on the feature vector to ovoid negative features.

2.2. Handcrafted features

In this work, we considerate macro patterns like buildings, harbor, forest and chaparral. Large variety of descriptors were used for pattern recognition [20], [21]. An appropriate descriptor selection will improve the classification precision by giving a more fitted description for the identified categories [22]-[24]. In land-use classification of high-resolution overhead imagery, several descriptors were based on global image descriptors for instance color histograms descriptors and homogeneous texture descriptors [24]. More recent works use second-order visual features namely [25] fisher vectors and vectors of locally aggregated descriptors. Certainly, the spatial texture has major importance for land use data, therefore, we propose a texture-color descriptor based on Gabor transform applied to the red, green, blue (RBG) color space.

The concept of this descriptor is to extract three chromatic descriptors corresponding to red, green and blue channels and concatenate the three sub-features into one feature vector. Thus, we combine in one descriptor the texture and color information's. Considering the Gabor transformation [26], The transform acts as a multichannel filter offering different scale frequencies and directions. Gabor function is defined by (1) where u and v are respectively orientation and scale for the Gabor kernel and $k_{u,v}$=$K_v$exp(iφu). k_v=k_{max}/Ω where k_{ma}x is the maximum frequency. The usually used values in literature are five scales and eight orientations. Therefore v ∈ {0,1,2,3,4}, u∈{0,..,7} with k_{max}=π/2, σ=2 π and Ω =√2, Ω measures the spacing between kernels and z is the sampling data which is the pixel position $z=(x,y)$, φu=$\pi u/8$. σ measures the gaussian envelope width [27], [28].

$$\Psi_{u,v}(z) = \frac{\|k_{u,v}\|}{\sigma^2}\exp\left(-\frac{\|k_{u,v}\|\|z^2\|}{\sigma^2}\right)\left(\exp(ik_{u,v}z) - \exp(-\frac{\sigma^2}{2})\right) \tag{1}$$

2.3. Multi-classifiers and decision fusion

The classification task aims to predict the labels of test data based on training data [29]. The learning classifiers are based on training stage that produces the separability model for a set of data. And then, the given classifier model is used for the data test classification. We will process the classification by three approaches that have distinguished performances.

- Support vectors machine (SVM): The SVM is a learning approach [30] based on minimizing both an empirical error and the complexity of the learner [31]. The classifier defines an optimal hyperplane with the largest margin to separates classes categories from a training set. Then, the test data are classified regarding their position to the hyperplane.
- K-nearest neighbors (KNN): KNN is based on memorizing in the learning step the samples values and classifying test images classified by comparison to the training samples [29]. The majority rule concept in the classification process attributes a class to a test sample based on the major class of its k most similar samples in the training set where k is a hyperparameter.
- Artificial neural networks (ANN): The artificial neural networks model is a multilayer perception networks multi-layer perceptron neural network (MLP) with a back-propagation algorithm. The nodes are sigmoid [32]. The MLP are determined by their connection's weights and biases. The learning stage aims to find optimum values for the MLP parameters.

2.4. Majority vote fusion

Data decision fusion is a higher level of data fusion [33]. It aims to improve the classification results and compensate the individual classifier weakness. Many fusion methods exist namely decision tree and Dempster-Shafer evidence theory. In this work, the classifiers fusion is performed by the majority vote principle [34]. Considering that $C_j(P)=i$ is the classifier decision for the patch P, the indicator function X_{ij} is defined by (2). The majority vote principle determines for each label k the next expression $X_k^C(P)$ as the sum of the values of $X_{ij}(P)$ as presented by (3). The majority vote decision is presented in (3).

$$X_{ij}(P)=1 \text{ if } C_j(P)=i \text{ else } X_{ij}(P)=0 \tag{2}$$

$$X_k^C(P)=\text{sum}_{j=1..M}(X_k^j(P)) \tag{3}$$

$$S(P)=k \text{ if } M_k^C(P)>M/2 \tag{4}$$

3. EXPERIMENTS AND ANALYSIS

3.1. Data set

The described method was evaluated on the common land-use dataset which is UC Merced land use dataset (UCMerced) [24]. UCMerced contains 21 classes selected from many US regions namely Boston, Columbia, Houston, Las Vegas, and Miami. The images are from aerial orthoimagery and the spatial resolution is one foot. The land use categories are: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts [35]. Figure 2 shows one example per each class. Each image is 256x256 pixels. The inter-class diversity is very small among some categories such as (harbor, parking lot) and (medium residential, mobile home park). The resolution, scale and illumination are variable.

Figure 2. UCMercedes: classes samples

3.2. Materials

The AlexNet model was obtained from the Matlab deep learning toolbox. The classifiers experimentations were conducted using the open-source software Weka 3.8.2 [36], [37]. WEKA refers to The Waikato Environment for Knowledge Analysis. The environment provides machine learning techniques in various programming languages and platforms. The tool provides also a complete framework for data mining researches implementation, evaluation and benchmarking.

3.3. Land-use identification results and analysis

For the learning classifiers, we perform the cross validation based on the 10-fold cross-validation. Firstly, the data is portioned randomly to 10 subsets. Then we select one subset for testing while the remaining subsets are used for training. The process is repeated 10 times. The result is obtained by averaging all the experimentations results. The training rate is 20% which mean that for each class, 80 images are used for training and 20 images for testing. The classification accuracy denotes the rate of test samples correctly identified by the approach. We evaluate the final accuracy as an average over the 21 labels. For the training and the testing, all the UCMercedes classes had the same number of samples to ovoid having biased overall accuracy. Balanced data choice could have reduced the overall accuracy by including certain not-well-presented classes and having important inner-class variability such as golf-course, airplane and free-way.

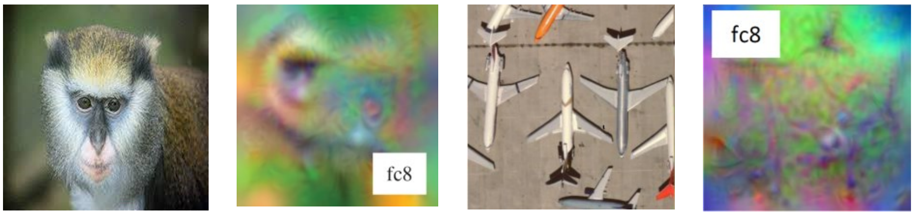
The accuracy results for the three classifiers and for the proposed fusion C-MV is presented are Table 2. The obtained accuracy for the SVM, ANN, KNN and C-MV are respectively 91.82%, 89.2%, 79.30% and 95.10 %. The high accuracies reveal the ability of deep features and texture handcrafted features to distinguish the land use categories. The deep features obtained from the pretrained baseline deep networks AlexNet on ImageNet dataset provide high level features. These features overcome the dataset inter-class similarities and in-class variabilities. The collected features from fully connected high layers in a deep architecture represent an abstract representation of lower level layers and provide more generic pattern description. In [38], the classification results using different deep pretrained architecture for feature extraction and SVM classifier reveals that the VGG-S reaches the best accuracy and outperforms the VGG-M, VGG-F, VGG-VD16, VGG-VD16, PlacesNet, CaffeNet and the AlexNet. Some deep models with considerably more layers may achieve lower result then the baseline AlexNet such as the VGG-VD. Similarly, PlacesNet performs worse than the AlexNet in land use scene but perform better than AlexNet in natural scene. Thus, the deep feature efficiency depends on the classification context. In [39], the accuracy result was based on performing ten data augmentation to the data test to improve the accuracy. The global feature of an image test is an average of the dense feature vector over the ten augmented images. That which explain that the SVM classification deep features extracted with AlexNet performs 94.37. In our case, without any augmentation for the testing set, the performance in 91.82 % when using deep features and texture features. In [40], the presented classification framework is based on deep feature extraction and SVM classification. The kernel is obtained by a different image features and are combined using weights. The authors concluded that deep features are more expressive than texture and spectral features.

However, the reconstructed image from high level features obtained by AlexNet, as presented in [41] presented an abstract representation of the original image and is not restricted to the dataset context. The example of a natural image presented in [41] shows the original image and the approximation obtained from the AlexNet highest fully connected layer as shown in Figure 3 (a). An airplane image and the reconstructed image by the eight fully connected layer is presented in Figure 3 (b) [5].

Table 2. Classification accuracy for the classifiers (test set)

	SVM	ANN	KNN	C-MV
Accuracy	91.82 %	89.2%	79.30%	95.10%

The texture information in the original image are not apparent in the reconstructed images. Thus the texture information's needs specific representation for textured scenes which is the case of several land categories in the land use dataset: agricultural, beach, buildings, chaparral, dense residential, forest, freeway, and harbor. Moreover, the texture features obtained by Gabor filters from the high-resolution land use images reveal structural and textural patterns which are very are different form the texture features in natural scenes. Thus, using handcrafted features remain important when transfer learning deals with contextually different scenes such as natural scene and high-resolution land use scenes. Using transfer learning for high-level semantic feature extraction and three fully connected layers for street view images classification achieved 61.8% accuracy in the work presented in [14].



(a) (b)

Figure 3. Natural image and reconstruction from eight fully connected layer of the AlexNet: (a) land use image, (b) reconstruction from eight fully connected layer of the AlexNet

Related works using hybrid feature extraction in land use classification demonstrated that the accuracies have been improved when using the texture features. For instance, in [3] the comparison of the classifications results using the spatial metrics, the texture parameters, and both measures show that the spatial metrics and texture provides the highest classification that reaches an overall accuracy of 76.4%. The overall accuracies provided by a similar approach presented in [9] using textual features and spectral indices and based on different preprocessing is within the range from 78.1% to 93.6%. For further comparison, we reported the classification performance for each class as presented in Figure 4. We observed that all classes achieved high accuracies for all classifiers. The crop process applied to the the original image 256*256 to the input required size 224*224 (or 227*227) leads to a little information loss that will not impact the classification results.

The SVM classifier, with linear kernel, performs better than ANN and KNN. Previous works based on handcrafted feature extraction and machine learning applied to land use dataset reached 76% using SVM [4], [38]. Although the state-of-the art feature extraction method represented the land use information well, the inner-class variabilities require more empowered classification frameworks.

Using the decision fusion approach C-MV the classification accuracy is 95.10%. Classifiers fusion have improved the classification for all land cover classes. Specifically the following classes accuracies have been improved: buildings (+20%), mobile home park (+more than 6%) intersection (+5%) and agricultural and beach. These classes present wide inner-class variability and wide proportions of the land-cover. Taking these land proportion could increase the overall accuracy and favorize biased accuracy that focused on class distribution in land use scenes.

The majority vote allows a better classes recognition especially for complex classes. Highest accuracies belong to buildings, chaparral, forest, medium density residential, mobile home park, parking lots and runway. These lands thematic have specific color and texture which allows a good separability by learning algorithm. The C-MV improved the accuracy for some complex classes by using the texture feature which is important for some land use categories. Specifically, medium density residential, sparse residential and storage tanks are better recognized than results presented in [38]. These classes present wide inner-class diversity and have specific texture which favorites the good accuracy results. Lower accuracies belong to airplane, baseball diamond, freeway, golf course, harbor and river. These classes are visually identified by their shape rather than by texture features.

Classifiers fusion contribution is therefore needed to empower the classification task. Many existing works have investigated the decision fusion and ensemble classification abilities. Zhang *et al.* [12] compared four classifiers using for spectral feature classification: a pixel-based MLP, a spectral and textural MLP, a contextual based CNN and an ensemble classifier MLP-CNN. The ensemble classifier outperforms the remaining classifiers in both urban and rural study sites with respectively an overall accuracy of 89% and 87% and for all classes. However, the grey level co-occurrence matrix (GLCM) textures have enabled the process of spatial information but using the MLP classifier was unable to deal with the curse of dimensionality when considering the 96 components of the GLCM feature vector. Using MLP and CNN classifiers have improved the pattern recognition process by providing contextual features in addition to the shallow structures provided by MLP. In [13], using different CNN architecture for land use provide accuracies of up to 77.4% using the contribution of the land cover classification based on a pixelwise classification.

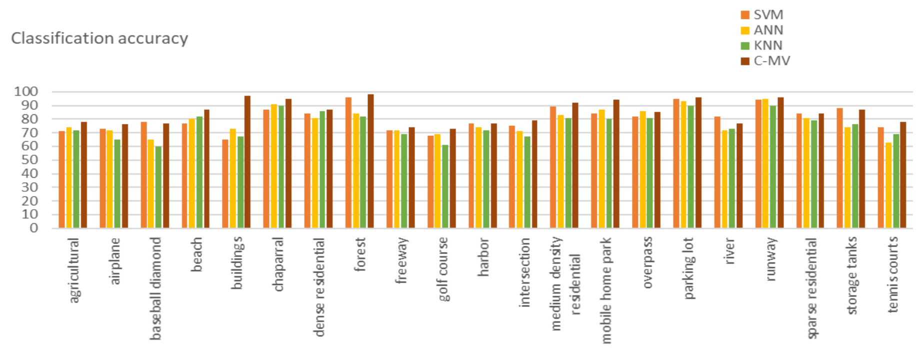


Figure 4. Per-Class accuracy for classifiers and the prposed C-MV

4. CONCLUSION

Land use data are crucial information for many environmental and socioeconomic application. Thus, finding a trustable classification framework is challenging especially for complex and changeable areas such as urban areas images. Traditional classification approaches based on handcrafted features have limited classification performance due to the diversity of the land use classes, the resolution impact on the pattern, the luminosity variability and the inner-class diversity. In this paper, se have investigated the ability of deep learning transferring in high resolution remoted sensed image recognition task. Taking into account the DCNN limitation due to the adoption of filters that cause the loss of spatial information and object boundaries, we use DCNN for contextual feature extraction and perform the ensemble classification by conventional learning classifiers. We utilized high-level feature extracted by deep learning and texture handcrafted features. The classification is performed by several machine learning algorithm including SVM, ANN and KNN. The final classification is a decision fusion by majority vote. Our findings indicate that specific handcrafted features are inherent for specific scene classifications when the available pretrained deep features are extracted from different scene context. Moreover, the remote sensed land use datasets have limited size which will not allow to train the model without getting advantages form transfer learning.

Our approach outperforms existing state-of-the-art method based on only handcrafted features and on machine learning classification. The deep features and transfer learning provide generic pattern description and accurate representation of the land use classes and allow high accuracies These features overwhelms the dataset inter-class similarities and in-class variabilities. Texture features have also contributed to overcome the texture information dissimilarity in natural images and land use images. Similar deep feature extraction approaches are based on further settings including coding approaches and averaging augmented testing images. Moreover, the classifiers fusion approach provides reliable results compared to single learning machine accuracy which prove that the fusion enhances classes separability. Complex region with high inner-class dissimilarities have taken advantages from the majority vote process. The overall accuracy achieved 95.10% whereas SVM, ANN and KNN provided respectively 91.82%, 89.2% and 79.30%. The ensemble classifier C-MV based on both contextual and low-level handcrafted features have allowed to compensate the classifiers weaknesses and have taken advantages from the hybrid features in highlighting specific object properties. This research paves the way to effectively select the suitable feature space and the more accurate classifier depending on the datasets and the specific relevant information. Future researches will focus on the object-based method to enhance the classification accuracy for complex areas. Deep learning approaches within this interest can be utilized to detect and recognize specific land use objects.

ACKNOWLEDGEMENTS

This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program

REFERENCES

[1] Gong P., *et al*., "2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data," *International Journal of Remote Sensing*, vol. 34, no. 7, pp. 2607-2654, Nov. 2012, doi: 10.1080/01431161.2012.748992.

[2] Chen G. F., Zeng G. W., and Chen H., "Study of RS image classification method based on texture features and neural network algorithm," *Journal of Chinese Agricultural Mechanization*, vol. 35, no. 1, pp. 270-274, 2014.

[3] D. Bhat, H. R. Acharya and S. H. R., "A Novel Solution to the Curse of Dimensionality in Using KNNs for Image Classification," *2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS)*, 2019, pp. 32-36, doi: 10.1109/ICoIAS.2019.00012.

[4] T. Kavzoglu, H. Tonbul, and Yildiz Erdemir, "Dimensionality Reduction and Classification of Hyperspectral Images Using Object-Based Image Analysis, "*J. Indian Soc. Remote Sens.*, vol. 46, pp. 1297-1306, Oct. 2018.

[5] Zhong Y. F., Fei F., and Zhang L. P., "Large patch convolutional neural networks for the scene classification of high spatial resolution imagery," *Journal of Applied Remote Sensing*, vol. 10, no. 2, pp. 1-20, Apr. 2016, doi: 10.1117/1.JRS.10.025006.

[6] Marmanis D., Datcu M., Esch T., and Stilla U., "Deep learning earth observation classification using ImageNet pretrained networks," *IEEE Geoscience and Remote Sensing Letter*, vol. 13, no. 1, pp. 105-109, Jan. 2016, doi: 10.1109/LGRS.2015.2499239.

[7] M. Herold, X. Liu, and K. C. Clarke, "Spatial metrics and image texture for mapping urban land use," *Photogramm. Eng. Remote. Sens*., vol. 69, no. 9, pp. 991-1001, Sep. 2003, doi: 10.14358/PERS.69.9.991.

[8] S. S. Wu, X. Qiu, E. L. Usery, and L. Wang, "Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use," *Annals of the Association of American Geographers*, vol. 99, no. 1, pp. 76-98, Jan. 2009, doi: 10.1080/00045600802459028.

[9] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIScience & Remote Sensing*, vol. 57, no. 4, pp. 1-20, Aug. 2019, doi: 10.1080/15481603.2019.1650447.

[10] D. Ming, J. Li, J. Wang, and M. Zhang, "Scale parameter selection by spatial statistics for GeOBIA: using mean-shift based multi-scale segmentation as an example," *ISPRS J. Photogramm. Remote Sens*., vol. 106, pp. 28-41, Aug. 2015, doi: 10.1016/j.isprsjprs.2015.04.010.

[11] M. Weigand, J. Staab, M. Wurm, and H. Taubenböck, "Spatial and semantic effects of LUCAS samples on fully automated land use/land cover classification in high-resolution Sentinel-2 data*," International Journal of Applied Earth Observation and Geoinformation*, vol. 88, pp. 102065, Jun. 2020, doi: 10.1016/j.jag.2020.102065.

[12] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson, "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS J. Photogramm. Remote Sens*., vol. 140, pp. 133-144, Jun 2018, doi: 10.1016/j.isprsjprs.2017.07.014.

[13] M. Li, L. Wang, J. Wang, X. Li, and J. She, "Comparison of land use classification based on convolutional neural network," Journal of Applied Remote Sensing, vol. 14, no. 1, pp. 016501, Jan. 2020, doi: 10.1117/1.JRS.14.016501.

[14] Yu Y., Fang F., Liu Y., Li S., and Luo Z., "Urban Land Use Classification Using Street View Images Based on Deep Transfer Network," *Urban Intelligence and Applications. Studies in Distributed Intelligence*, Jan. 2020, doi: 10.1007/978-3-030-45099-1_7.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25,no. 2, pp. 1106–1114, 2012, doi: 10.1145/3065386.

[16] Nawaz W., Ahmed S., Tahir A., and Khan H. A., "Classification Of Breast Cancer Histology Images Using ALEXNET," *Image Analysis and Recognition. ICIAR*, Jun. 2018, doi: 10.1007/978-3-319-93000-8_99.

[17] Han X., Zhong Y., Cao L., and Zhang L., "Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification," *Remote Sensing*, vol. 9, no. 8, pp. 848, Aug. 2017, doi: 10.3390/rs9080848.
[18] Jing Sun, Xibiao Cai, Fuming Sun and J. Zhang, "Scene image classification method based on Alex-Net model," *2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS)*, Jinzhou, Aug. 2016, pp. 363-36, doi: 10.1109/ICCSS.2016.7586482.
[19] F. A. Jokhio and A. Jokhio, "Image Classification using AlexNet with SVM Classifier and Transfer Learning," *JICTRA*, vol. 10, no. 1, pp. 44-51, Jun. 2019.
[20] R. Soulard and P. Carré, "Quaternionic wavelets for texture classification," *Pattern Recognition Letters*, vol. 32, no. 13, pp. 1669-1678, Oct. 2011, doi: 10.1016/j.patrec.2011.06.028.
[21] H. Clemmensen, M. E. Hansen and B. K. Ersbøll, "A comparison of dimension reduction methods with application to multi-spectral images of sand used in concrete," *Machine Vision and Applications*, vol. 21, no. 6, pp. 959-968, Oct. 2010, doi: 10.1007/s00138-009-0193-z.
[22] C. Chakraborty, "Computational Approach for Chronic Wound Tissue Characterization," *Elsevier: Informatics in Medicine Unlocked*, vol. 17, pp. 100162, 2019, doi: 10.1016/j.imu.2019.100162.
[23] V. Yadav, V. Bharadi Dr, and S. K. Yadav, "Feature Vector Extraction based Texture Feature using Hybrid Wavelet Type I & II for Finger Knuckle Prints for Multi-Instance Feature Fusion," Procedia Comput-er Science, vol. 79, pp. 351-358, 2016, doi: 10.1016/j.procs.2016.03.046.
[24] Yi Yang and Shawn Newsam, "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification," *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS)*, Jan. 2010, doi: 10.1145/1869790.1869829.
[25] R. Negrel, D. Picard, and P-H. Gosselin. "Evaluation of Second-order Visual Features for Land-Use Classification". *12th International Workshop on Content-Based Multimedia Indexing*, Jun 2014, doi: 10.1109/CBMI.2014.6849835.
[26] G. Gilanie, U. I. Bajwa, M. M. Waraich, and M. Nasir, "Classification of normal and abnormal brain MRI slices using Gabor texture and support vector machines," *Signal, Image and Video Processing*, vol. 20, pp. 479-487, 2018.
[27] P. Maho, S. Barthelme, and P. Comon, "Non-linear source separation under the Langmuir model for chemical sensors," *IEEE Sensor Array and Multichannel Signal Processing Workshop*, Jul. 2018.
[28] Y. Deville, L. Tomazel, D. Leonardo, and T. Duarte, "An Overview of Blind Source Separation Methods for Linear-Quadratic and Post-nonlinear Mixtures," *International Conference on Latent Variable Analysis and Signal Separation*, Aug. 2015, doi: 10.1007/978-3-319-22482-4_18.
[29] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for kNN Classification," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 3, pp. 1-19, Jan. 2017, doi: 10.1145/2990508.
[30] V. N. Vapnik, "Statistical learning theory," *Wiley, New York*, 1998.
[31] Huang, X., Zhang, L., Wang, B. et al. Feature clustering based support vector machine recursive feature elimination for gene selection. Appl Intell 48, pp. 594-607, 2018, doi: https://doi.org/10.1007/s10489-017-0992-2.
[32] Mirjalili Seyedali, S. Mohammad Mirjalili, and Andrew Lewis, " Let a biogeography-based optimizer train your Multi-Layer Perceptron," *Information sciences*, vol. 269, pp. 188-209, Jun. 2014, doi: 10.1016/j.ins.2014.01.038.
[33] K. Tidriri, T. Tiplica, N. Chatti, and S. Verron, "A generic framework for decision fusion in Fault Detec-tion and Diagnosis," *Eng. Appl. of A,* vol.71, pp. 73-86, May 2018, doi: 10.1016/j.engappai.2018.02.014.
[34] A. J. H. Simons, "The Theory of Classification, Part 5: Axioms, Assertions and Subtyping," *Journal of Object Technology JOT*, vol. 2, no. 1, Jan. 2003, pp. 13-21, doi: 10.5381/jot.2003.2.1.c2.
[35] Shahriari M., Bergevin R., "Land-use scene classification: a comparative study on bag of visual word framework," *Multimed Tools Appl,* vol. 76, pp. 23059-23075, 2017, doi: 10.1007/s11042-016-4316-z.
[36] Frank E., Hall M. A. and Witten I. H., "The WEKA Workbench. Online Appendix for: Data Mining: Practical Machine Learning Tools and Techniques," Morgan Kaufmann, Fourth Edition, 2016. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf.
[37] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., and Witten I. H., "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10-18, Nov. 2008.
[38] V. Kumar Rana, and T. MaruthiVenkata Suryanarayana, "Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands," *Remote Sensing Applications: Society and Environment*, vol. 19, pp. 100351, Aug. 2020, doi: 10.1016/j.rsase.2020.100351.
[39] Fan H., Gui-Song X., Jingwen H. and Liangpei Z., "Classification of High-Resolution Remote Sensing Imagery," *Remote Sens*, vol. 9, pp. 1-5, 2017, doi: 10.3390/rs71114680.
[40] Ling D., Hongyi L., Changmiao H., Wei Z. and Shumin W., "Alexnet Feature Extraction And Multi-Kernel Learning For Object-Oriented Classification Classification," *Remote Sensing and Spatial Information Sciences*, Apr. 2018, doi: 10.5194/isprs-archives-XLII-3-277-2018.
[41] Mahendran A., and Vedaldi A., "Understanding Deep Image Representations by Inverting Them," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2015.