

## A ROUGH SET DECISION TREE BASED MLP-CNN FOR VERY HIGH RESOLUTION REMOTELY SENSED IMAGE CLASSIFICATION

C. Zhang<sup>a,\*</sup>, X. Pan<sup>b</sup>, S. Q. Zhang<sup>c</sup>, H. P. Li<sup>c</sup>, P. M. Atkinson<sup>a</sup>

<sup>a</sup> Lancaster Environment Centre, Lancaster University, Lancaster, Lancashire, LA1 4YQ UK - (c.zhang9, pma)@lancaster.ac.uk

<sup>b</sup> School of Computer Technology and Engineering, Changchun Institute of Technology, 130021 Changchun, China - 101103991@qq.com

<sup>c</sup> Northeast Institute of Geography and Agroecology, Chinese Academic of Science, 130102 Changchun, China - (zhangshuqing, lihuapeng)@neigae.ac.cn

Commission IV, WG IV/3

**KEY WORDS:** Convolutional Neural Network, Multilayer Perceptron, VHR remotely sensed imagery, Rough Set Theory, Decision Tree, Fusion Decision

### ABSTRACT:

Recent advances in remote sensing have witnessed a great amount of very high resolution (VHR) images acquired at sub-metre spatial resolution. These VHR remotely sensed data has posed enormous challenges in processing, analysing and classifying them effectively due to the high spatial complexity and heterogeneity. Although many computer-aid classification methods that based on machine learning approaches have been developed over the past decades, most of them are developed toward pixel level spectral differentiation, e.g. Multi-Layer Perceptron (MLP), which are unable to exploit abundant spatial details within VHR images. This paper introduced a rough set model as a general framework to objectively characterize the uncertainty in CNN classification results, and further partition them into correctness and incorrectness on the map. The correct classification regions of CNN were trusted and maintained, whereas the misclassification areas were reclassified using a decision tree with both CNN and MLP. The effectiveness of the proposed rough set decision tree based MLP-CNN was tested using an urban area at Bournemouth, United Kingdom. The MLP-CNN, well capturing the complementarity between CNN and MLP through the rough set based decision tree, achieved the best classification performance both visually and numerically. Therefore, this research paves the way to achieve fully automatic and effective VHR image classification.

### 1. INTRODUCTION

Over the last decade, ground-based, airborne and satellite sensors and platforms have evolved dramatically, of which a major achievement is the acquisition of very high resolution (VHR) remotely sensed imagery. These VHR images provide sub-metre ground resolution at an increasing level of spatial details, facilitating a wide range of geospatial applications such as urban land use retrieval, precision agriculture, and tree delineation etc. However, the increase of spatial resolution does not signify the increase of accuracy for processing, classifying and labelling such kinds of data, mainly because of spectral variations within class and similarity between class occurred within VHR imagery. Therefore, it is of necessity and urgent to develop robust and accurate image classification methods that are able to address the challenges from VHR imagery.

Researchers and practitioners have developed numerous computer based classification methods, ranging from unsupervised K-mean clustering, supervised parametric methods such as Maximum Likelihood, and non-parametric machine learning algorithms e.g. Multi-layer Perceptron (MLP), Support Vector Machine and Random Forest etc. MLP, as a classical non-parametric machine learning approach, has been widely used in remote sensing domain including VFSR-based land cover classification (e.g. Del Frate et al., (2007), Pacifici et al. (2009)). It was invented to mimic the human brain through layer-wised information processing (Atkinson 1997) with nonlinearity to handle the spectral features irrespective to its

statistical characteristics. However, the MLP model was difficult to go deep with typically shallow model structure due to its full connection properties that involve a large amount of parameters, and its intractable “black-box” machine learning characteristics. The MLP is essentially a pixel-based classifier with shallow architectures that predicts the membership association of each pixel to a particular land cover type.

Recent advances in machine learning and computer vision inspired that the deep feature representations can be learnt hierarchically at multiple levels through deep machine learning methods (LeCun et al. 2015). These deep learning methods represent the state-of-the-art in a variety of domains, including object detection, information retrieval, image recognition and robotics etc. The convolutional neural network (CNN), as a well-established deep learning approach, has widely recognized as one of the best deep neural networks in pattern recognition and computer vision. Its popularity is largely related to the success in the ImageNet Large Scale Visual Recognition Challenge at 2012 where 11% less error rate was achieved by CNN in comparison with several contestants. Since then, the CNN remained active in multiple domains, and introduced to the remote sensing domain. The considerable majority of researches in remote sensing were focused on object detection and scene classification. Recent researches also show the possibility of CNNs for the remote sensing image classification task. For example, Chen et al. (2016) introduced a 3D CNN to jointly extract spectral-spatial features, thus, making full use of the continuous hyperspectral and spatial spaces. Zhao and Du,

\* Corresponding author

(2016) used image pyramid to learn deep features through CNN at multiple scales. Långkvist et al., (2016) used CNN model with different contextual size to classify and segment VHR satellite images. All of these works proved the superiority of CNN in spatial feature representations. However, none of them investigated the merits and shortcomings of CNN as a classifier itself. For example, the object edges might be over-smoothed by the contextual filters used by CNN, whereas the MLP might do better job in such case through pixel-based differentiation. In fact, any classifiers, even the CNN classifier, have their inherent uncertainties. These uncertainties can be low and high at different spatial locations, which requires further analysed and provides corresponding solutions.

Rough set theory proposed by Pawlak (1982) is a mathematical tool to quantify the uncertainties within the data itself, and further divided them into positive (100% correctness), non-positive (contains uncertainty). It was successfully applied in diverse domains such as pattern recognition, machine learning, knowledge acquisition, and decision support system (Regniers et al., 2016). In the field of remote sensing, rough set model has been applied in rule-based feature reduction, knowledge discovery, land cover classification. To deal with the inconsistency in remotely sensed data, Pan et al., (2010) introduced a variable precision rough set (VPRS) to tolerate some errors within the positive region. The VPRS can be to quantify the uncertainty in the CNN classification. For those high uncertainties due to the lack of spectral differentiation, the MLP was used as an alternative, where a decision tree was built up in this paper to integrate MLP and CNN as an ensemble classifier MLP-CNN. The proposed MLP-CNN was compared with the benchmark MLP and CNN in an urban area to test its effectiveness.

## 2. METHODS

### 2.1 Multi-layer Perceptron (MLP) and Convolutional Neural Network (CNN)

A multi-layer perceptron (MLP) is a classical neural network that composed of an input layer, one or several hidden layers and an output layers. These layers are connected subsequently with nonlinear function to strengthen the nonlinearity. The input is spectral features and the output target is the land cover class. Each layer involves weight and bias parameters that are trained and learnt through backpropagation algorithm. Convolutional neural network (CNN) is a variant of neural network that specifically designed for classifying images or multiple arrays. Unlike MLP that utilized the spectral feature only, the CNN inputs the contextual image patch to learn the spatial features or patterns with context. Those input feature maps were fed into convolutional layer with convolution functions and pooling layer that subsamples the features alternatively until the higher level feature representations were acquired. Those learnt higher level features were classified by Logistic regression to predict and label the land cover types using the maximum membership associations.

### 2.2 Rough Set Decision Tree based MLP-CNN

Suppose the membership prediction of CNN at each pixel are  $n$ -dimensional vectors  $C$ , where  $n$  represents the number of classes while each dimension corresponds to the pixel's probability of a specific class ( $i$ -th class) with certain membership association. For a pixel, the confidence being determined as  $class(C)$  which is the maximum membership association, can be quantified as:

$$conf = Max(C) - Mean(C) \quad (1)$$

Where the  $Max(C)$  denotes the maximum value of vector  $C$ , and the  $Mean(C)$  represents the average value of vector  $C$ . The  $conf$  quantifies how much confidence, or conversely how much uncertainty, that the pixel was being predicted.

Image classification results of CNN can be regarded as partially correct and partially incorrect at geometric space. Taking an easily classifiable ground object for example, its central region is often accurately classified by CNN, but its boundary region is likely to be misclassified. The two regions (i.e. patch centre and patch boundary) can then be described theoretically by using rough set theory.

Standard formulation of rough set theory can be referred to Pan et al. (2010), where an indiscernible relation  $IND(P)$  between two objects  $x$  and  $y$ :

$$IND(P) = \{(x, y) \in U^2 \mid \forall a \in P, a(x) = a(y)\} \quad (2)$$

The  $U$  is a non-empty set of finite objects known as the universe of discourse,  $P$  represents an attribute set, and the  $a$  is the attribute values. The equation 2 means that the objects  $x$  and  $y$  within that set/region were inseparable or indiscernible.

While applying the rough set on the CNN classification confidence, the confidence value ( $conf$ ) of any two samples within this region should belong to the same indiscernible relation, of which they should be treated simultaneously. Therefore, the CNN classification confidence map can further be partitioned as a series of intervals, each of which represents a particular indiscernible relation:

$$[0, step], [step, step \times 2), \dots, [step \times prob / step, 1] \quad (3)$$

Where,  $step$  is the atomic granule representing the least unit of indiscernible relation. Within a specific interval, the equivalence class of the indiscernible relation based on  $IND(P)$  can be defined as:

$$[x]_p = \{y \in U \mid (x, y) \in IND(P)\} \quad (4)$$

Give a set  $X \in U$ , such equivalence class describes all the training samples are classified consistently within that interval.  $X$  can therefore be approximated using only the information contained within the region  $R$ , including a  $R$ -lower approximation:  $\underline{R}X = \{x \mid [x]_R \subseteq X\}$  and a  $R$ -upper approximation:  $\overline{R}X = \{x \mid [x]_R \cap X \neq \emptyset\}$ . If  $\underline{R}X \neq \overline{R}X$ , then the tuple  $(\underline{R}X, \overline{R}X)$  forms a rough set. The positive ( $POS_R(X)$ ), negative ( $NEG_R(X)$ ) and boundary ( $BND_R(X)$ ) of  $R$  regions can be defined as:

$$POS_R(X) = \underline{R}X \quad (5)$$

$$NEG_R(X) = 1 - \overline{R}X \quad (6)$$

$$BND_R(X) = NEG_R(X) - POS_R(X) \quad (7)$$

Fig. 1 shows the positive (Eq. 5), negative (Eq. 6) as well as boundary (Eq. 7) regions of a standard rough set to represent the correctness, incorrectness and uncertainty of image classification. Note that, according to the classification confidence that matches the training samples, the CNN classification results can be partitioned into a range of regions:

the positive region (the negative region, respectively) represents that the entirety of training samples lying in the region are correctly (incorrectly, respectively) classified, whereas only part of those in the boundary region are correctly classified.

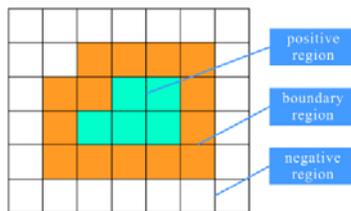


Figure 1. An illustration of the standard rough set with positive, boundary and negative regions.

Given the uncertainty description by rough set. For those positive regions, the CNN was directly trusted in consideration of its robustness in spatial representations. While the negative and boundary regions were re-classified by MLP and CNN jointly using decision trees. The decision tree provides a transparent and convenient manner to combine both classifiers with a set of rules being learnt from data themselves.

### 3. EXPERIMENT AND RESULTS

#### 3.1 Study Area and Data Material

The city of Bournemouth, UK and its surrounding environment, which lies on the south coast of England, was chosen as a case study area. It covers the urban and suburban areas with a mixture of anthropogenic urban surface (e.g. roof materials, asphalt, concrete) and semi-natural environment (e.g. vegetation, bare soil), thereby representing a good test for classification algorithms.

A scene of aerial imagery of Bournemouth was captured on 22 July 2012 using a Vexcel UltraCam Xp digital aerial camera with 50 cm spatial resolution and four multispectral bands (Red, Green, Blue and Near Infrared). Nine dominant land cover classes, including Clay roof, Concrete roof, Metal roof, Asphalt, Grassland, Trees, Railway, Bare soil and Shadow were carefully chosen, in consideration of study area characteristics and spatial details.

Sample points were collected using a stratified random scheme from ground data provided by local surveyors at Southampton, and split into 50% training samples and 50% testing samples for each class. Field land cover survey was conducted throughout the study area on July 2012 to further check the validity and precision of the selected samples. In addition, a highly detailed vector map from Ordnance Survey, namely the MasterMap Topographic Layer, was fully consulted and cross-referenced to gain a comprehensive appreciation of the land cover and land use within the study area.

#### 3.2 Rough Set Uncertainty and Decision Tree

The rough set uncertainty was derived from the membership association predicted by CNN using a softmax classifier. The uncertainty map, partly shown in Figure 2 (b), characterized how much uncertain the prediction was in terms of making decisions. The bright areas indicate the regions that are very much certain on the prediction, whereas the dark regions are with low confidence about the CNN classification.

The continuous measurements of uncertainty were further partitioned into binary (certain vs uncertain) using a threshold. The uncertainty threshold was set through a trail-and-error

approach to estimate the low confidence areas. Those low confidence (high uncertain or very likely incorrect) areas are reclassified by MLP in consideration of its pixel-level spectral differentiation. The threshold was sampled broadly from 0.1 to 0.9 with a small step of 0.05 to cover the entirety of the space. It was cross-validated using 10% of the training samples to approximate the global optimization for the decision fusion. The uncertainty threshold, eventually, was tuned as 0.75 with the binary result illustrated in Figure 2 (c).

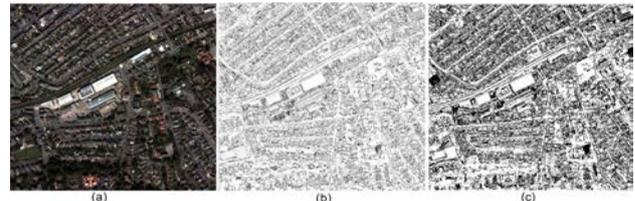


Figure 2. The rough set uncertainty of (a) original aerial image, with (b) uncertain confidence and (c) the binary threshold result that show the certain and uncertain regions

For those uncertain areas, a new sample set was employed to build a decision tree to re-classify them using MLP and CNN jointly. The variables of the decision tree are 18 membership predictions of the 9 land covers, including 9 membership associations derived from CNN and 9 memberships from MLP respectively. The decision rule clearly indicates how those land cover classes are differentiated through either MLP or CNN. It suggests that the Asphalt (MLP) is the most important variable. When the depth of the decision rules increases, the difficulties of differentiation increases simultaneously. The top level is the most tough land cover categories to be distinguished, i.e. the concrete roof and the asphalt.

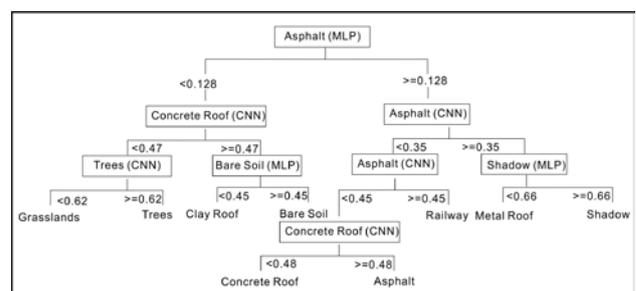


Figure 3. The decision tree of the MLP-CNN to derive the land cover classes at the uncertain areas

#### 3.3 Classification Results and Analysis

The land cover classification accuracy was validated through per-class mapping accuracy, overall accuracy (OA) and Kappa coefficient ( $\kappa$ ) shown in Table 1. The classification results show that the MLP-CNN achieved the best overall accuracy (OA) of 90.46% with kappa coefficients ( $\kappa$ ) of 0.89, higher than the CNN (86.37% OA with  $\kappa$  of 0.84), and the MLP (81.52% OA with the  $\kappa$  of 0.77).

In terms of individual classes, the MLP demonstrated high performance on characterizing the Clay roof and Shadow due to its spectral differentiation. It, however, failed to differentiate Trees and Grasslands because of huge similarity spectrally. The CNN managed to characterize the Trees owing to its spatial representations. This was also shown in the decision tree in Figure 3, where the Trees (CNN) was used for decision fusion.

However, the CNN still made some mistakes on some classes that are spatially complicated such as concrete roof and asphalt. Those classes without clear textures are hardly differentiable by pure CNN. At the same time, the edges and small features are easily omitted due to the usage of convolutional filters. The Figure 4 visually demonstrates a part (S1) of the classification results, where the MLP tends to produce salt-and-pepper classification results, and the CNN wrongly take the concrete roof as asphalt. The rough set decision tree based MLP-CNN successfully captured the complementary patterns of both MLP and CNN classifiers and made the best predictions. Therefore, the MLP-CNN could provide a good alternative for VHR remotely sensed image classification.

Class	MLP	CNN	MLP-CNN
Clay roof	91.37%	88.56%	<b>92.64%</b>
Concrete roof	68.52%	74.37%	<b>78.21%</b>
Metal roof	89.75%	91.42%	<b>92.19%</b>
Asphalt	88.59%	85.98%	<b>91.26%</b>
Grassland	73.51%	88.63%	<b>93.52%</b>
Trees	65.68%	82.28%	<b>88.45%</b>
Bare soil	80.46%	85.23%	<b>90.34%</b>
Shadow	91.56%	90.14%	<b>92.15%</b>
Railway	82.14%	90.23%	<b>91.46%</b>
Overall Accuracy	81.52%	86.37%	<b>90.92%</b>
Kappa Coefficient ( $\kappa$ )	0.77	0.84	<b>0.89</b>

Table 1. Classification accuracy comparison amongst MLP, CNN and the proposed MLP-CNN approach for Bournemouth using the per-class mapping accuracy, overall accuracy (OA) and Kappa coefficient ( $\kappa$ ). The bold font highlights the greatest classification accuracy per row.

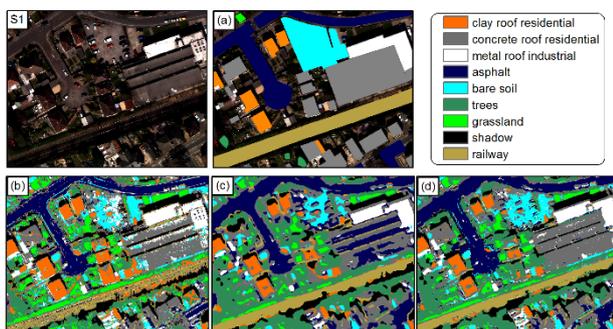


Figure 4. The partial region of the original image (S1), with (a) ground truths for validation, (b) the classification results of MLP, (c) the classification results of CNN, and (d) the classification results of MLP-CNN fused by the rough set decision tree

#### 4. CONCLUSION

Due to its high intra-class variability and low inter-class disparity, VHR image classification poses great challenges to any single machine learning algorithm, even for the powerful deep learning convolutional neural network (CNN). In this paper, we built a novel rough set-based decision tree to combine CNN and MLP in a transparent, concise manner. The results of the proposed MLP-CNN show very promising in both quantitative and visual aspects. Therefore, this research paves

the way to an effective solution to the complicated problem of automatic VHR image classification.

#### ACKNOWLEDGEMENTS

This research was supported by PhD studentship “Deep Learning in massive area, multi-scale resolution remotely sensed imagery” (NO. EAA7369), sponsored by Ordnance Survey and Lancaster University. The authors thank to the staff from the Ordnance Survey for the supply of aerial imagery and supporting ground data.

#### REFERENCES

- Atkinson, P.M., Tatnall, A.R.L., 1997. Introduction neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4), pp. 699-709.
- Chen, Y., Jiang, H., Li, C., Jia, X., Member, S., 2016. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), pp. 6232-6251.
- Del Frate, F., Pacifici, F., Schiavon, G., Solimini, C., 2007. Use of neural networks for automatic classification from high-resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4), pp. 800-809.
- Långkvist, M., Kiselev, A., Alirezaie, M., Loutfi, A., 2016. Classification and segmentation of satellite orthoimagery using convolutional neural networks. *Remote Sensing*, 8(329), pp. 1-21.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature*, 521(7553), pp. 436-444.
- Pacifici, F., Chini, M., Emery, W.J., 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment*, 113(6), pp. 1276-1292.
- Pan, X., Zhang, S., Zhang, H., Na, X., Li, X., 2010. A variable precision rough set approach to the remote sensing land use/cover classification. *Computers & Geosciences*, 36(12), pp. 1466-1473.
- Pawlak, Z., 1982. Rough sets. *International Journal of Computer & Information Sciences*, 11(5), pp. 341-356.
- Regniers, O., Bombrun, L., Lafon, V., Germain, C., 2016. Supervised classification of very high resolution optical images using wavelet-based textural features. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6), pp. 3722-3735.
- Zhao, W., Du, S., 2016. Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 113, pp. 155-165.