

A novel approach for scene classification from remote sensing images using deep learning methods

Xiaowei Xu , Yinrong Chen , Junfeng Zhang , Yu Chen , Prathik Anandhan & Adhiyaman Manickam

To cite this article: Xiaowei Xu , Yinrong Chen , Junfeng Zhang , Yu Chen , Prathik Anandhan & Adhiyaman Manickam (2020): A novel approach for scene classification from remote sensing images using deep learning methods, European Journal of Remote Sensing, DOI: [10.1080/22797254.2020.1790995](https://doi.org/10.1080/22797254.2020.1790995)

To link to this article: <https://doi.org/10.1080/22797254.2020.1790995>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 07 Oct 2020.



Submit your article to this journal



Article views: 48



View related articles



View Crossmark data

A novel approach for scene classification from remote sensing images using deep learning methods

Xiaowei Xu^a, Yinrong Chen^a, Junfeng Zhang^b, Yu Chen^c, Prathik Anandhan^d and Adhiyaman Manickam^e

^aCollege of Public Administration, Huazhong Agricultural University, Wuhan, Hubei, China; ^bSchool of Public Administration, Zhongnan University of Economics and Law, Wuhan, Hubei, China; ^cDepartment of Economics, Zhengzhou Institute of Light Industry School of Economics and Management, Zhengzhou, Henan, China; ^dDepartment of Computer Science, VIT University, Vellore, India; ^eCollege of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China

ABSTRACT

Deep learning plays a major task in classification of unsupervised data, which utilises network enhanced learning. This technique proves to be powerful in remote sensing (RS) field for spatial data classification. In the existing environment, huge amounts of data from the earth observation satellites known as satellite images time series (SITS) are gathered, which can be utilised for observing the areas related to geography over through time. In this proposed model the time series model utilised is based on geography. There exists a challenge on how these types of information can be analysed in the field of remote sensing. Notable, techniques related to deep learning substantiated in dealing with remote sensing usually for classification of scene. In this paper, we propose an enhanced classification method involving Recurrent Neural Network (RNN) along with Random forest (RF) for land classification using satellite images, which are publicly available for various research purposes. We utilised spatial data gathered from the satellite images (i.e. time series). Our experimental classification is based on pixel and object-based classification. The attained analysis illustrates that the proposed model outperforms the other present day remote sensing classification techniques by producing 87% target accuracy of classification scene from satellite images.

ARTICLE HISTORY

Received 8 May 2020
Revised 22 June 2020
Accepted 29 June 2020

KEYWORDS

Deep learning; satellite images; time series; recurrent neural network; remote sensing

Introduction

Natural Resource administrators, policy creators and scientists request information of progressive changes over huge spatial and temporal degrees for managing many problems that are begging to be addressed, for example, worldwide environmental change, carbon spending plans, and biodiversity (DeFries et al., 1999). Distinguishing and describing changeovers is the characteristic initial move toward recognising the driver of the change and understanding the mechanisms in change. Satellite remote sensing has been utilised over a long time as a method for distinguishing and ordering changes in the state of the land surface over time (Lu et al., 2004). Satellite sensors are appropriate to this undertaking since they give reliable and repeatable estimations at a spatial scale, which is proper for catching the impacts of numerous procedures that cause change, including usual (for example, forest fires, insect assaults) and anthropogenic (for example, farming, urbanisation) disturbances (Jin & Sader, 2005).

In spite of the fact that the estimation of remotely sensed datasets of long term indexes for change detection has been established firmly (De Beurs & Henebry, 2005), just a few number of time series change identification techniques have been created. Two significant

difficulties exist. To begin with, strategies must take into considering the discovery of changes inside complete long term data collections while representing seasonal variation. Evaluating change from remotely sensed information is not very simple, since the time series images usually contain a sequence of seasonal, continuous and unexpected changes, additionally errors from noises that begin from leftover geometric mistakes, scatterings in the atmosphere and cloud impacts (Roy et al., 2002). Survey of existing change identification techniques by Coppin et al. (2004) and Lu et al. (2004) have appeared, in any case, that most strategies center around short picture time arrangement (just 2–5 pictures). The danger of variability with change is high with rare pictures, since influences of disturbances can happen in the middle of picture acquisitions (De Beurs & Henebry, 2005).

Figure 1 shows the various components involved in remote sensing. From image acquisition to classification, various components are used to perform the tasks. Once the image is acquired using drone or satellite, various computer vision methodologies are applied to gather patterns from the images. Completely automated land-spread classification is a challenging issue that is continually pushing the envelope of AI and computer vision. Classifying land use from remote sensing is crucial to monitor and

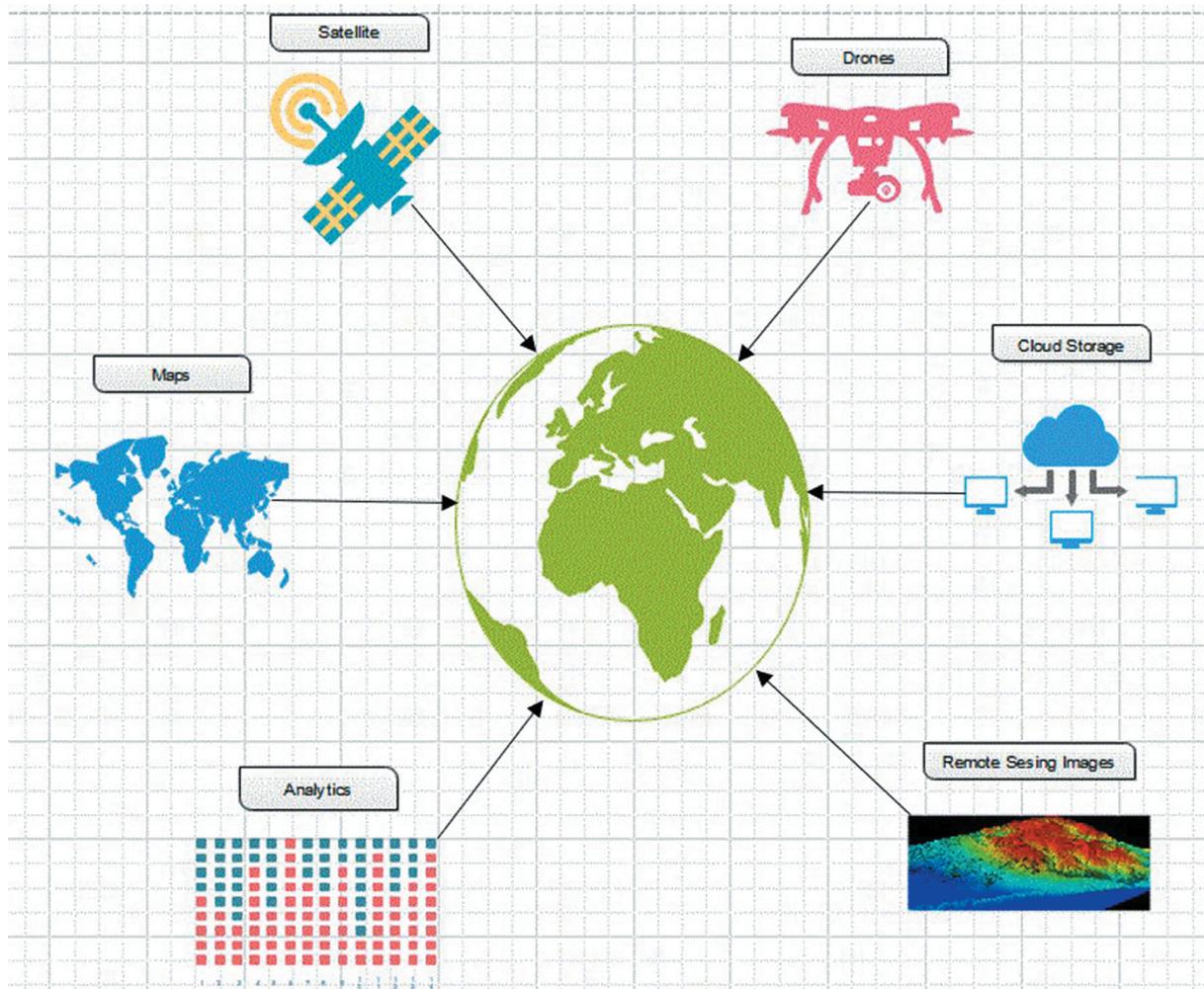


Figure 1. Components of remote sensing.

administer human advancement such that it is unimaginable in degree and scale from a ground point of view with human-driven techniques. Besides, mechanised grouping by means of AI is fundamentally significant for understanding the continually changing surface of the earth, particularly anthropogenic changes. Automated land-spread classification of high resolution remote sensing techniques is an exceptionally attractive capacity yet presents numerous difficulties because of the sheer volume and types of satellite and pictures gathered every day. This is additionally complicated by the wide types of human-related attributes and scales present across topographically various territories on the earth surface.

Advancements in computer vision includes finding specific features, which are proficient in computational time and memory utilisation. The classifiers, then again, are required to have great capacity of generalisation, yet accomplish superior performance rates. Remote sensing image characterisation has become a significant research field. The feature-based approach, which is an additional step from data mining strategies has indicated great degrees of execution on remote sensing image analysis. Image classification is a significant application in the Computer

vision problem domain. Our research objective is to improve techniques for remote sensing image classification using Machine learning-based approaches. Satellite images are classified and the features in the images such as building, landscape, desert, buildings are studied with time related changes.

As a subset of Artificial Intelligence, machine learning has made tremendous achievements, and is currently taking off in the remote detecting (L. Zhang et al., 2016). Benefiting by the use of deep convolutional features, the strategies dependent on deep learning have accomplished high efficiency in image classification and the precision is constantly being advanced with the improvement of new methods (Z. Li et al., 2018). Deep learning has likewise been acquainted with cloud discovery in satellite-based imagery in recent ongoing examinations. In the investigation of Mateo-García et al. (2017), a straightforward convolutional neural system (CNN) engineering was intended for the cloud-based Proba-V multispectral pictures. CNNs have guarantees for taking care of cloud masking issues, contrasted with the traditional Artificial Intelligence approach. In case of deep learning, the pictures are first fragmented, and then the features in the images are extracted pixel

wise and the features in the images are classified. Zhan et al. (2017) likewise applied a deep convolutional system to recognise cloud and snow from satellite pictures at the pixel level.

The successful use of deep learning in satellite image classification has made numerous new and promising roads of examination. Recurrent Neural Networks (RNN) are a typical deep neural design that was developed to mirror the cerebrum's capacity to learn and create various levelled highlight portrayals inside a multi-level feature extraction phase. These scholarly visual features give the characterisation stage to separate images into their individual classes. Deep learning-based models frequently beat ordinary feature extraction by conventional methods and so used in this study for satellite image classification. Technology involving remote sensing methods is utilised to observe the patterns in weather that consists the drought segments over a particular region. The details obtained from this information is further processed to predict the rainfall pattern of a significant region and also utilised for finding the association amongst the present and the nest climatic scenario, which will help the farmers in knowing the weather condition confidently.

Related works

In recent days, Deep neural networks (DNNs) have emerged as a worldview paradigm for Artificial Intelligence in numerous areas. Nonetheless, finding an appropriately huge informational dataset for preparing DNN is a huge test. This is a significant issue in the remote sensing space, where the enormous amount of satellite and aerial imagery is present, however this brings a new shortcoming of absence of index data that is required for promptly accessing other picture modalities. To overcome this usually analytics on remote sensing utilises two systems related to DNN: Transfer learning (TL) with optimising and data augmentation tailored explicitly for imagery in remote sensing. TL permits to bootstrap a DNN while protecting the deep visual element extraction in a picture corpus from an alternate image space (Scott et al., 2017). Data growth causes problems in different parts of remote sensing imagery to drastically extend image preparing strategies, informational collections and improve DNN strength for remote detecting image information. Lv et al. (2015) proposed a classification approach dependent on the Deep Belief Network model for pixel by pixel urban mapping utilising polarimetric synthetic aperture radar (PolSAR) data. Through the DBN model, successful relevant mapping features can be extracted automatically from the PolSAR information to improve the classification accuracy.

Lv et al. (2014) proposed an approach for land cover classification dependent on Deep Belief Network(DBN) for extensive land spread use and urban classification. By applying the DBN model, successful spatio-temporal mapping features can be consequently detected to improve the accuracy of the classification. Six-date RADARSAT-2 Polarimetric SAR (PolSAR) information over the Great Toronto Area was utilised for assessment. W. Li et al. (2016) applied Stacked Autoencoder (SAE), and developed a grouping system for large scale remote-sensing image preparing and advanced the model parameters dependent on test tests. The superiority of the approach is proved by focussing on the efficiency of the SAE-based methodology with conventional classification methodologies including Random Forest, Support vector machine, and Artificial Neural Networks with various execution investigations.

Li et al. (2018) investigated the utilisation of deep convolutional neural networks to classification of land use from extremely high spatial resolution, orthorectified, visible band multispectral imagery. Ongoing mechanical and business applications have driven the collection of an enormous measure of high spatial resolution images in the detectable red, green, blue (RGB) spectral band groups and investigated the potential of deep learning performance to exploit this imagery for programmed land use/land cover (LULC) characterisation. C. Zhang et al. (2019) proposed Joint Deep Learning (JDL) model that combines a multilayer perceptron (MLP) and convolutional neural system (CNN), and is executed through a Markov procedure including iterative refreshing. In the JDL, Land use classification is led by the convoluted neural networks, which are made contingent upon the Land Cover probabilities anticipated by the MLP. Finally, those Land Usage probabilities together with the first symbolism are re-utilised as contributions to the MLP to reinforce. Prathik et al. (2018, 2019) proposed a novel algorithm for soil image segmentation based on color and region and considered various spatial database for correlation analysis. Three entryways are used to control the information, yield and update of the LSTM model for optimisation. Additionally, the learned principle from the developed model can be applied to identify changes and move the change rule starting with one trained image then onto the next new objective multi-level picture the spatial and unearthly element portrayals. This procedure of refreshing the MLP and CNN, structures a joint conveyance, where both Land Cover and Land Usage are classified by simultaneous iterations.

Kampffmeyer et al. (2018) proposed a convolutional neural network (CNN) design for urban land spread classification, which can insert all accessible training modalities in the hallucination network. The system will replace missing information

modalities in the test stage, empowering combination capacities in any event, when information modalities are missing in testing. The efficiency of the missing information replacement is exhibited by utilising two datasets consisting of optical and Digital surface model (DSM) pictures. Kussul et al. (2016) developed a model dependent on a various leveled model that incorporates self-organising maps (SOM) for information preprocessing and division (clustering), fusion of multi-layer perceptrons (MLP) for classification and heterogeneous information combination and geospatial examination for post-processing.

Piramanayagam et al. (2016) utilised arbitrary methods for structured labels and convolutional neural systems for the land spread classification of multi-sensor images from remote sensing. In random forest classification, singular choice trees are prepared on features acquired from image patches and their corresponding labels in the patch. Basic data present in the image patches improves the classification execution when contrasted with simply using pixel features. Interdonato et al. (2019) proposed the initial deep learning engineering for the investigation of SITS information, to be specific DuPLO (DUal viewpoint profound Learning design for time arrangement classificatiOn), that joins Convolutional and Recurrent neural systems to exploit their complementarity. The speculation is that, since CNNs and RNNs catch various parts of the information, a mix of the two models would deliver an increasingly assorted and complete portrayal of the data for the hidden land spread classification task. This shows the efficiency of using ensemble models in land cover classification and so a random forest and recurrent neural network-based model is used in this study. Senthil Murugan and Usha Devi (Murugan & Devi, 2018a; Murugan & Devi, 2019; Murugan & Devi, 2018b) have proposed and analysed the machine learning techniques and optimisation model for large amounts of data.

Wang et al. (2019) utilised a time series classification information extraction model, which is a long term utilising a bidirectional long-term and short-term memory network (Bi-LSTM). In the model, quantitative remote sensing items joined with the Digital Elevation Model, evening time lighting information, and longitude/latitude rise information were utilised. and applied to this model in China's 1982–2017 0.05° land spread classification problem domain. Lyu et al. (2016) proposed an improved Long Short-Term Memory (LSTM) model to acquire and record the change data of long term succession of remote sensing data. Specifically, a core memory cell is used to take in the change rule from the data concerning twofold changes or multi-class changes. Three entryways are used to control the information, yield and update of the LSTM model for optimisation. Additionally, the learned principle from

the developed model can be applied to identify changes and move the change rule starting with one trained image then onto the next new objective multi-level picture. The above studies summarise the studies carried out in the field of remote sensing land cover classification. From the above studies, it is evident that deep learning could be the most effective tool for analyzing the remote sensed images, So in this study, a deep learning-based image classification using RNN-RF is proposed.

Proposed methodology

Dataset description

Large numbers of researchers have utilised the well-established UC Merced (UCM) dataset (Yang & Newsam, 2010) to scrutinise the classification of land use procedures for imagery for remote sensing with high resolution images. The dataset taken here contains 21 various classes of labelled groups of high resolution images. The weakness of utilising these dataset is that it has a smaller dataset, which uses deep learning methods. Figure 2 describes the sample images from the dataset, which is used for classification.

Pre-processing

Once the data is gathered, it is essential for pre-processing. This basically consists of radiometric and geometric corrections followed by image clipping. In this work, the data gathered already processed by geometric as well as radiometric corrections. Images are in the format of RGB, then the image is processed by converting into grayscale images and an unwanted portion of the image, which is found to be unnecessary is taken off. Image is filtered using digital filters to eradicate the noise and discrepancy.

Normalisation

In order for making the task of feature extraction and classification easier, this technique is utilised, the a/b is the size of dimension of the plane image. In feature relationship adaptive normalisation, therefore, standard planes of the dimension are not essentially filled. Based on the feature ratio, the image, which is normalized is centred with one dimension covered. Assuming the plane is square and the length of each side is referred to as X . width is denoted by U_2 and height of image is represented by V_2 . The feature ratio is represented by

$$F_2 = \left\{ \begin{array}{l} \frac{U_2}{V_2}, \text{ if } U_2 < V_2, \\ \frac{V_2}{U_2}, \text{ otherwise.} \end{array} \right. \quad (1)$$

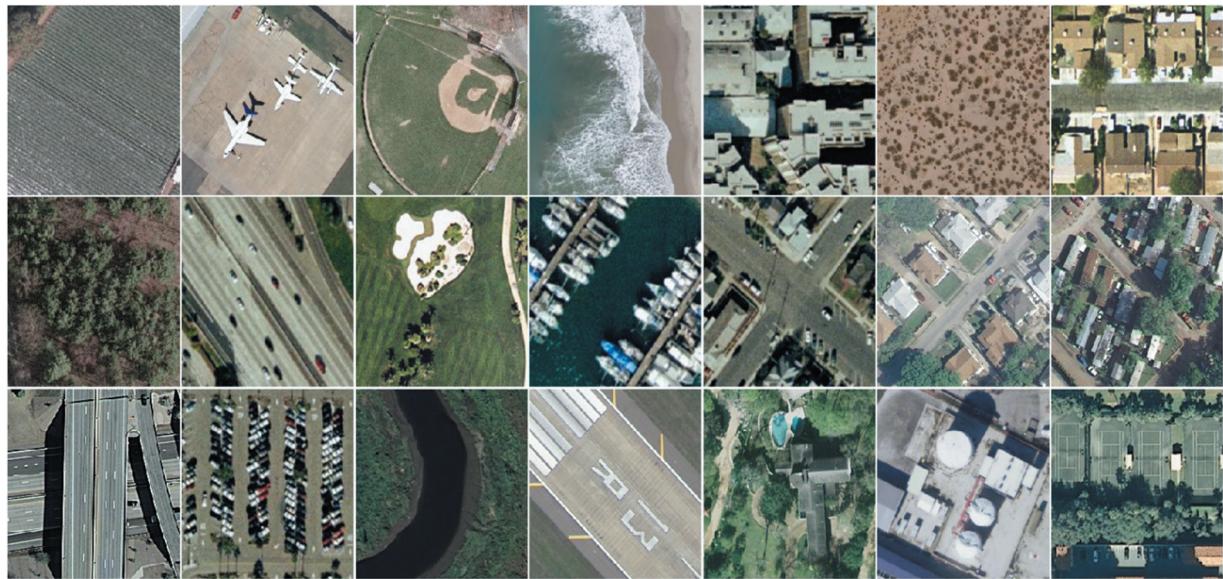


Figure 2. Illustration of sample images from UCM dataset consisting of 21 various land classes.

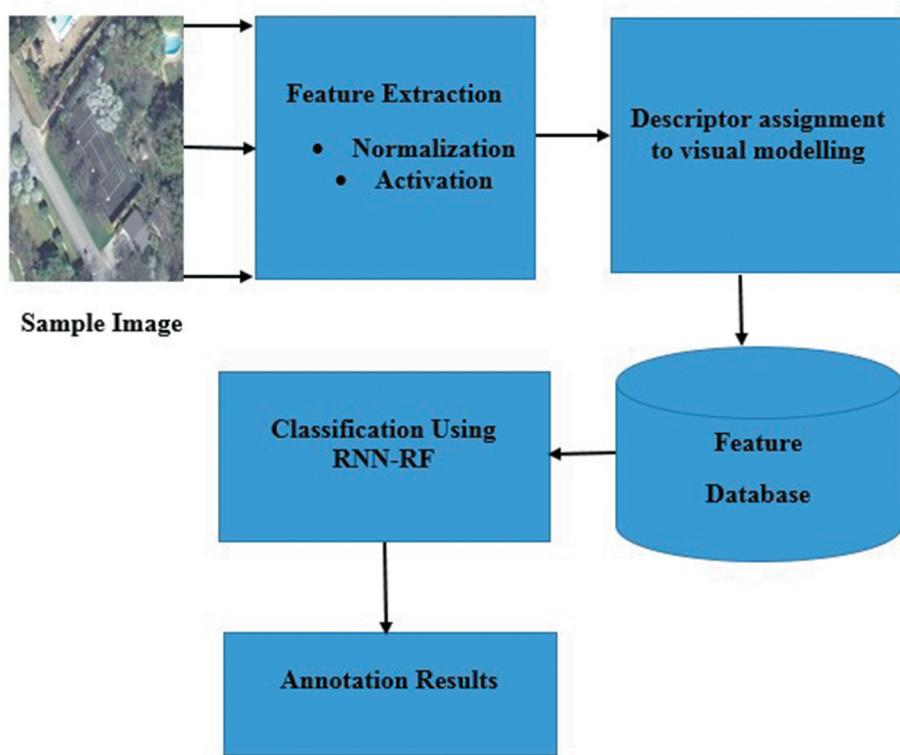


Figure 3. Proposed architecture for land image classification.

When one dimension is covered by the normalised images, then $(U_2, V_2) = X$. In case of normalisation based on the moment with the centroid of the image united to the middle of the image plane, moreover, the image, which is normalized does not essentially cover the dimension of the image and proceed beside the plane of image. In this portion of the image, U_2 and V_2 are not categorical by X and a portion of the image outside of the plane is cut off. The image size of $U_1 \times V_1$ transformed into size consisting of $U_2 \times V_2$. Image

can be accomplished by transforming by backward or forward mapping. Original image before transformation can be represented as $l(n, m)$ and $p(n', m')$, individually, image normalised is engendered by $p(n', m') = l(n, m)$ based on mapping of coordinate. The backward and forward mapping can be given as

$$n' = n'(n, m), m' = m'(n, m),$$

and

$$n = n(n', m'), m = m(n', m'),$$

Individually, In this case, we define the mapping coordinates of different normalisation procedures, and then gather the pixels interpolation.

The linear normalisations backward and forward mapping, normalisation of moment and slant along with nonlinear normalization. Transformations ratio is denoted by σ and γ , which is given by

$$\sigma = \frac{U_2}{U_1},$$

$$\gamma = \frac{V_2}{V_1},$$

In this methodology, normalisation of moment refers to the transformation, which is linear without any rotation, images normalised size is defined by moments. (n_r, m_r) represents the gravity's centre of the initial image before segmentation,

$$n_r = \frac{h_{10}}{h_{00}},$$

$$m_r = \frac{h_{10}}{h_{00}},$$

Here h_{xy} illustrates the moments in geometry:

$$h_{xy} = \sum_n \sum_m n^x m^y l(n, m), \quad (2)$$

and (n'_r, m'_r) represents the centre's geometry of plane's normalisation, which is given by

$$n'_r = \frac{U_2}{2},$$

$$m'_r = \frac{V_2}{2},$$

In moment-based slant normalisation, the slant angle is determined from moment of second order:

$$\tan\theta = \frac{\sigma_{11}}{\sigma_{02}},$$

Where,

$$\sigma_{xy} = \sum_n \sum_m (n - n_r)^x (m - m_r)^y l(n, m), \quad (3)$$

In mapping the image, n and m are discrete in nature but $m'(n, m)$ and $n'(n, m)$ need not be essentially discrete; in case of mapping backward, n' and m' are discrete but $n'(n, m)$ and $m(n', m')$ are not essentially discrete.

This image classification can be investigated further by imagining the results of classification on the region. Figure 4 illustrates the segmentation carried out before classification where the left side of the image describes the seashore image as the original image and on the left side of the image the seashore image is segmented by smoothing. There exists a difference between the water in the seashore and the sand accumulated along with the seashore. Many various demonstration in the area of remote sensing retain the identification of robust features and distinctive points of features are tracked by camera module. Significant instances on one such is 3D modelling of scene followed by the movement from structural component or localisation, which takes place simultaneously and mapping for remote image applications. The normal process of tracking point features can be classified as two phases. (1) Detection- initial phase is to identify the group of feature distinctive points. (2) Re- identification- the initial process of feature tracking is the prominent re identification of features. Processing includes image normalisation that tend to change the values of intensity pixels. Applications consists of images with lesser contrast because of glare for instances. The every

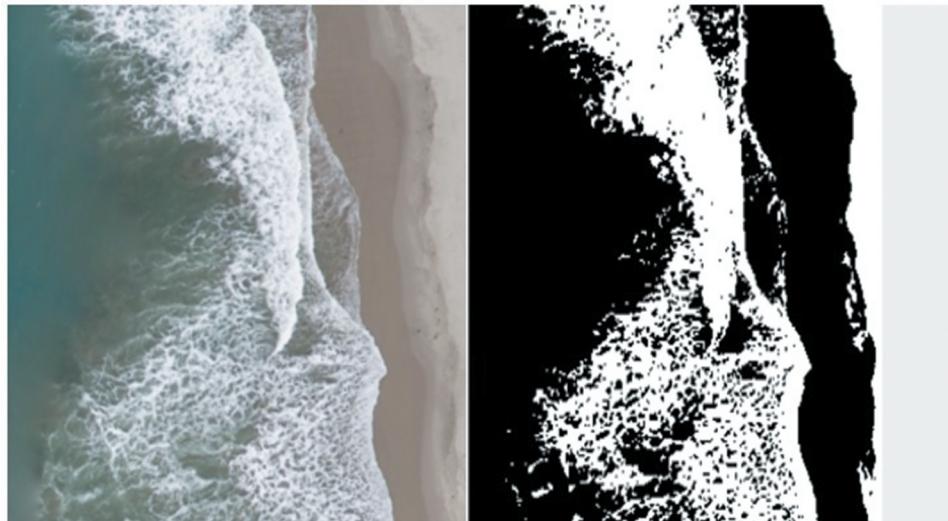


Figure 4. Left: sample image from dataset taken for classification. Right: classification of image after smoothing done by segmentation.

intensity of pixel is a multiplicative by 255/130 and making it as range 0 to 255.

Recurrent neural network

It's a type of artificial neural network, which ranges the predictable feedforward network with connections. As like a feedforward network, this type of network can propagate a number of inputs in the sequential form consisting of hidden layers that are recurrent whose function of activation is based on the earlier step. In this case, network exhibition can be dynamic in behaviour.

Given a group of data $a = (a_1, a_2, \dots, a_X)$, here a_i is i the data step, hidden layer in RNN of state l_x given by

$$l_x = \{0, \text{if } x = 0 \omega(l_{x-1}, a_x), \text{otherwise} \quad (4)$$

Here, ω is a function, which is nonlinear, which consists of logistic function or function, which is hyperbolic function. Occasionally, input in RNN will have output $b = (b_1, b_2, \dots, b_X)$. For instance, classification done in some of hyperspectral images, we only need one output such as b_X .

In RNN traditional model, rule updated in the hidden recurrent state in [Equation 2](#) is actually implemented as given below:

$$l_x = \omega(T_{a_x} + Sl_{x-1}) \quad (5)$$

Here T and S are the matrices coefficient for present steps input and activation of units in recurrent hidden at the last step, individually.

Modelling of probability distribution of the RNN over the element, which is next in the sequence data, considering its present state l_x , by absorbing the spreading over the data, which are the sequence of variable length. Consider $h(a_1, a_2, \dots, a_X)$ be probability with sequence, which has is formed into

$$h(a_1, a_2, \dots, a_X) = h(a_1) \dots h(a_1, \dots, a_{X-1}) \quad (6)$$

Recurrent network can be utilised to distribute the conditional probability as

$$h(a_1, \dots, a_{X-1}) = \omega(l_x) \quad (7)$$

where as l_x is obtained using [Equations \(3\)](#) and [\(6\)](#), RNNs networks shows predominant results in various computer tasks and machine learning. In case of long term data, RNNs find it difficult to train as gradients tend to disappear. In order to handle this type of situation, designing of delicate recurrent units can be used.

Long short-term memory (LSTM) known as hidden recurrent unit, which can handle learning of sequential data with dependencies of long term. LSTM layer based on recurrent creates a storage cell a_x at step x . Unit in LSTM has a activation that can be calculated by

$$s_x = g_x \tanh(a_x) \quad (8)$$

Here $\tanh(.)$ is the function of hyperbolic tangent and g_x is the gate output, which calculates the memory part content, which is exposed. The gate output is upgraded as

$$g_x = \sigma(U_{gj}C_x + U_{gs}s_{x-1} + U_{ga}a_x) \quad (9)$$

In [Equation 9](#), $\sigma(.)$ is the function that represents the logistic sigmoid and U refers to the matrices weight; for instance, U_{gj} is defined as matrix weights input and output and U_{ga} defines the weight matrix memory output.

a_x is the memory cell, which is upgraded by adding on content of cells memory given as $\underline{a_x}$ and content of present memory is discarded.

$$a_x = j_x \emptyset \underline{a_x} + k_x \emptyset a_{x-1} \quad (10)$$

where \emptyset is multiplication carried element wise, and the memory cells content $\underline{a_x}$ is gathered by

$$\underline{a_x} = \tanh(\tanh(U_{aj}C_x + U_{as}s_{x-1})) \quad (11)$$

j_x is the input gate that controls the level to, which unique memory details are added to the cell's memory. k_x represents the forget gate, which is used to make decisions of existing memory cells contents. In order to evaluate the two gates, the following equation is represented.

$$j_x = \sigma(U_x j_x + U_{xs} s_{x-1} + U_{ja} a_{x-1}) \quad (12)$$

$$k_x = \sigma(U_k j_x + U_{fs} s_{x-1} + U_{fa} a_{x-1}) \quad (13)$$

Random forest

Classification of multi way in image is carried by random forest using various trees, the growth of trees are carried using randomisation. Labelling of each tree's leaf node is calculated by posterior distribution among various classes of image. Every node internally placed consists of a test that best splits the portion of data, which needs to be classified. While classification is carried on a given image, it is forwarded down the tree and calculating the aggregation of leaf distribution. Inclusion of randomness during training is done at two points: data used in training is subsampled such that various subset is used to grow trees and identifying the node tests.

Growing trees, here the trees are considered to be binary and are built in top-down method. The test of each binary node can be taken in two strategies: (i) aimlessly, that is data liberated; or (ii) using the greedy algorithm to pick a test that separates best in given example training. Information gain is calculated by using [Equation 13](#).

$$\Delta F = - \sum_j \frac{|V_j|}{|V|} F(V_j) \quad (14)$$

Which caused by dividing the V set of instances into two different subsets V_j conferring the given test. Here

$F(v)$ is considered as entropy - $\sum_{i=1}^M l_i(l_i)$ where l_i define the proportion of v included to class i . For every single non terminal node the procedure of selection if iterated with examples of training deteriorating in node. Stopping of recursion takes place whenever a node receives fewer instances or after reaching depth given.

Posterior learning phase, considering that R is every trees set, D is all classes set and P is all leaves set for a considered tree. The posterior probability in the training stage is $(l_{t,j}(X(J) = d))$ for every class $d \in D$ at every leaf node $j \in D$, are initiate for each tree $t \in R$. The estimation of probability is done by ratio of total of images J of class d that attain j to the total images that attain j . $X(J)$ is considered as class label d for image J .

Proposed RNN-RF-based image classification algorithm

The primary objective of the using RNN-RF is to effective visual analysis of remote sensing images for various features and development of enhanced classification models for identification of objects. The classification algorithm consists of a remote sensing image as input which follows certain preprocessing and segmentation procedures after which the classification of image is given using neural networks. Creation of dictionary is for holding the data based on the size and pixel range of image utilised for classification to extract the scenes.

Input: x : group of satellite images
 y : image, which is required to classify

Output:

Group y total of images that are classified with enhanced accuracy

RNN-RF model ()

begin

image x is preprocessed

dictionary created for (x)

feature extraction (x)

descriptor assignment

RNN- training (image samples)

Feature database (m)

end

dictionary creation (x)

Extraction of features (x)

apply logistic function using Equation (3)

evaluate the filter component

for

every row in image

apply descriptor

end

change- key (n)

feature vector group

relate PCA

Random Forest (data,

m

)

For every image sample, Generate classification tree
 Select split amongst all forecasters
 Sampling

m

randomly amongst predictors

Select the appropriate split among variables
 Calculate new data using aggregation from the trees
 RNN-training (illustrations, labels)
 implementing RNN classifier
 (image dictionary
 Image classification (

m

)

get input for classification processing

match (m)

//get descriptor locally for image used for training
 change- key (n)

modelling visually using descriptor

image comparison with change-key

end

end

Results and discussion

The typical benchmark situation in UC Merced image dataset initially stipulated in (DeFries et al., 1999) is surveyed to calculate accuracy in classification. Cross validation stratified, which is utilised for various experiments, here four fold is utilised for training and selection of model, and rest of fold, which are unseen is classified for accuracy measurement. Evaluating empirically states that the size of input image is vital for performance accuracy of the model, where convolutional size of filter and phase of filter makes the performance better. Selection ranges from hyper parameters are resulted in higher accuracy in classification. Different architectures are initiated for evaluation in order to enhance the hyper parameters.

Accuracy comparison

Stratified five-fold validation analysis of various standard techniques is deployed. The enhanced accuracy is obtained using unsupervised feature learning (UFL) utilising UC Merced and aggregating vectors locally, which is considered to be an addition of visual glossary methods such as bag-of-words. Stranded DCNN is enhanced by using these techniques, which shows 90.26% accuracy of UFL. The greatest mentioned class confusion was amongst residential, which is accumulated in medium density regions, also regions between storage tanks and buildings. Various classes with lower correctness predictions are made on

buildings, storage tanks, residential places at medium density regions and tennis court fields.

The rate of convergence of training in RNN is depicted in [Figure 6](#), analysing the progression of both testing and training accuracies taking total epochs into considerations. The accuracy of the test in single-view is shown in [Figure 5](#), where the training and testing accuracy of the neural network along with the total number of epochs are illustrated. Core features are obtained by convolutional filters that are cultured by RNN and it seems the network lowers the convolutional reaction dimensions to absolute single dimensional reaction suitable for utilisation of activation function.

The association amongst the image size and the accuracy of labelling is illustrated in [Figure 7](#) against the number of samples of bands. The precision percentage illustrated in the figure attained only after fine tuning for every various features obtained from the

image. Wherever the size of the image used for classification increases then the learning rate of the network also increases drastically and inclining was faster initially and then leading to gradual stabilisation as the size of network increases. The neurons in the network were large enough and future enhancement of the size of the network was not initially essential and the size of the map was used for classification.

As the feature size increases with improvement in labelling accuracy of the samples, which has classes assigned non-uniformly. For classes, which are easily separable, there is marginal enhancement. For example, forest cover does not display much enhancement and it is definitely separated even though the size of map is minimal. Classes, which overlap the distribution of data (e.g., beach and forest), the details are more obvious. This indicates that huge sized feature maps may be utilised when handling data that are overlapping strongly.

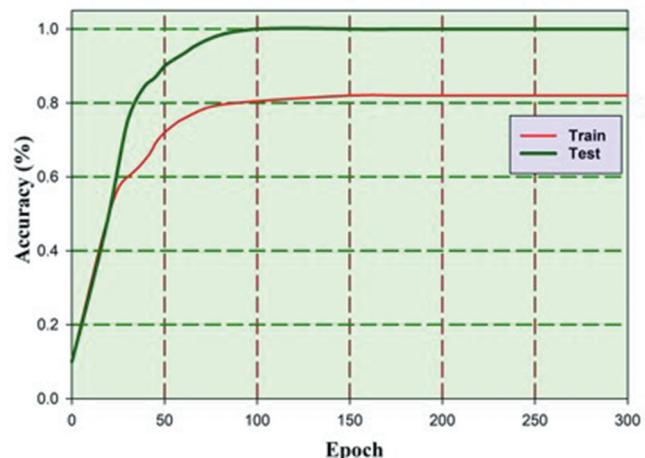


Figure 5. Averaged accuracy during cross validation of five-fold.

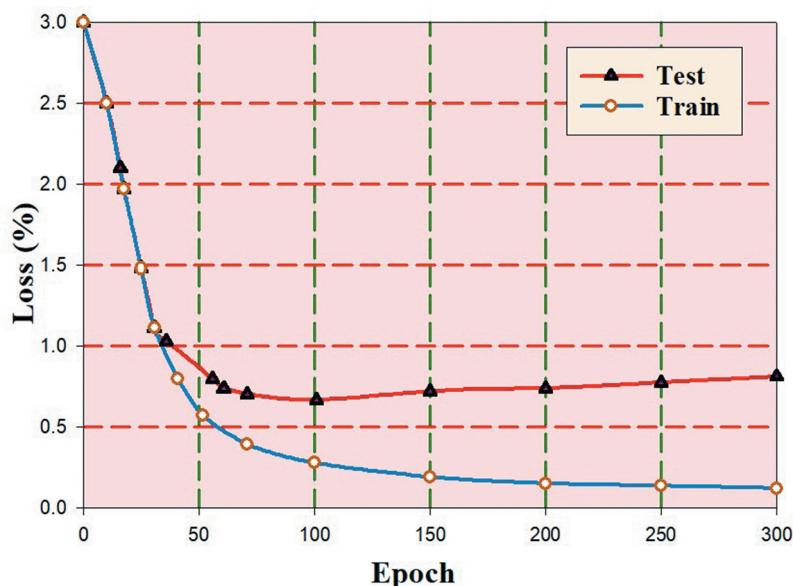


Figure 6. Average loss in cross validation of five-fold.

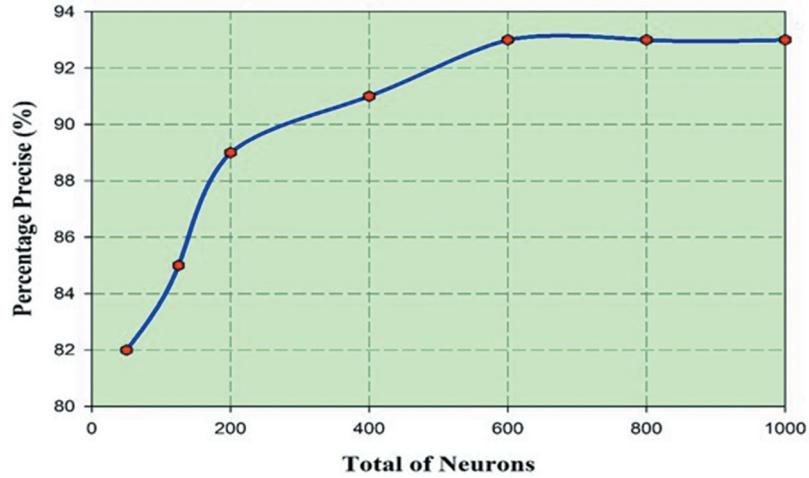


Figure 7. Association among size of image and labelling accuracy.

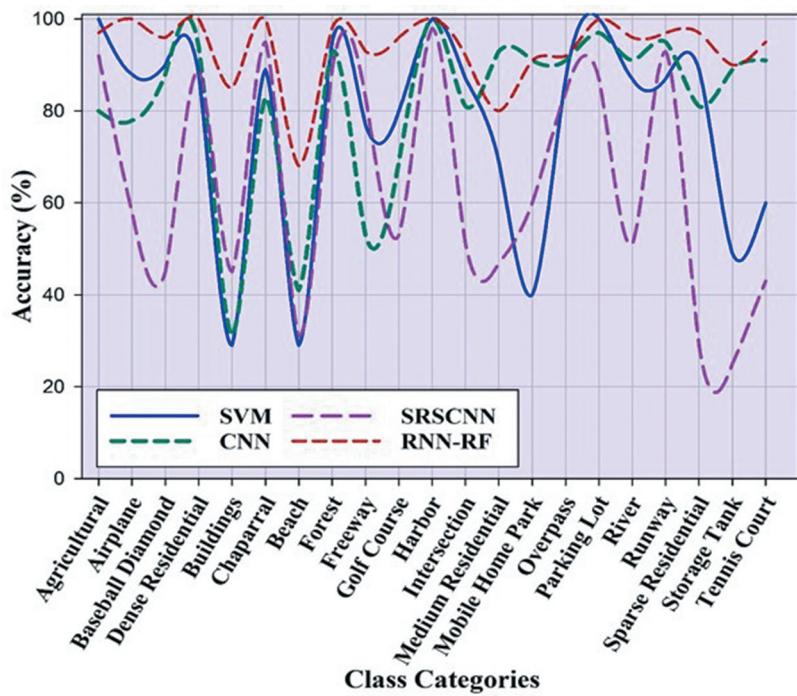


Figure 8. Accuracy comparison of various method for land use classification.

The overall performance of RNN-RF is illustrated in Figure 9 in the form of confusion matrix. There we can see some confusion amongst medium residential and harbor, chaparral and medium residential. This illustrates the fact where different types of classes have the same structural or spectral features such as airplane and storage tanks, tennis court and sparse residential. In case of some classes such as chaparral and medium residential has seen typical shape characteristics.

More concentration has been presented with regards to utilising shape characteristics. Confusion matrix produced from perfect classification outcome with 80% training ratio. As displayed in the confusion matrix, 21 classes attained enhanced accuracies than 90%, half of achieved has 95% accuracy. This attained

accuracy is from the detail that two different classes have same distributions in image, such as density and structures of buildings, which distinguished utilisation amongst each other.

The process of image retrieval whose quality is measured using this precision metric and recall illustrates the image retrieval amount. For each and every image in the query, it has two quantities recall and precision ranges were forecasted. The attained higher precision worth suggests that further relevant images are gathered in our proposed deep learning method.

Conclusion

Remote sensed photographs plays a significant role in monitoring environmental conditions, disaster

intersection	0.95	0.0	0.0	0.0	0.0	0.0	0.0	0.02	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.0	0.02	0.0	0.0	
baseballdiamond	0.0	0.96	0.0	0.0	0.0	0.0	0.0	0.03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.02	0.0	0.0	
overpass	0.0	0.0	0.92	0.0	0.0	0.01	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.02	0.02	0.02	
parkinglot	0.0	0.0	0.0	0.98	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
mediumresidential	0.0	0.03	0.0	0.0	0.76	0.0	0.0	0.0	0.0	0.0	0.0	0.06	0.01	0.0	0.0	0.02	0.0	0.0	0.05	0.0	0.0
agricultural	0.0	0.0	0.0	0.0	0.0	0.93	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
beach	0.0	0.0	0.0	0.0	0.05	0.0	0.85	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.02	0.0	0.02	0.0	0.0	0.0	
freeway	0.0	0.0	0.02	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
storage tanks	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.93	0.0	0.0	0.01	0.0	0.0	0.0	0.02	0.0	0.03	0.0	0.0	
sparseresidential	0.0	0.0	0.0	0.03	0.0	0.0	0.0	0.0	0.0	0.92	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
river	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.99	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
harbor	0.0	0.0	0.0	0.0	0.0	0.02	0.0	0.0	0.0	0.0	0.0	0.96	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
mobilehomepark	0.0	0.0	0.01	0.0	0.01	0.0	0.1	0.0	0.0	0.0	0.0	0.86	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
forest	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
airplane	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.07	0.0	0.02	0.0	0.0	0.97	0.0	0.01	0.0	0.0	0.0	0.0	
runway	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.91	0.0	0.0	0.0	0.0	0.0	0.0	
golfcourse	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.96	0.02	0.02	0.02	0.0	0.0	
denseresidential	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.88	0.0	0.0	0.0	0.0	0.0	
tenniscourt	0.0	0.0	0.02	0.0	0.0	0.03	0.0	0.0	0.05	0.0	0.0	0.02	0.0	0.0	0.02	0.0	0.87	0.0	0.0	0.0	
chaparral	0.0	0.02	0.03	0.0	0.2	0.0	0.0	0.0	0.02	0.0	0.0	0.03	0.01	0.0	0.0	0.03	0.0	0.03	0.8	0.0	
buildings	0.0	0.0	0.02	0.0	0.04	0.0	0.01	0.0	0.02	0.01	0.0	0.0	0.01	0.0	0.01	0.0	0.03	0.04	0.0	0.85	

Figure 9. Confusion matrix obtained for UC merced dataset.

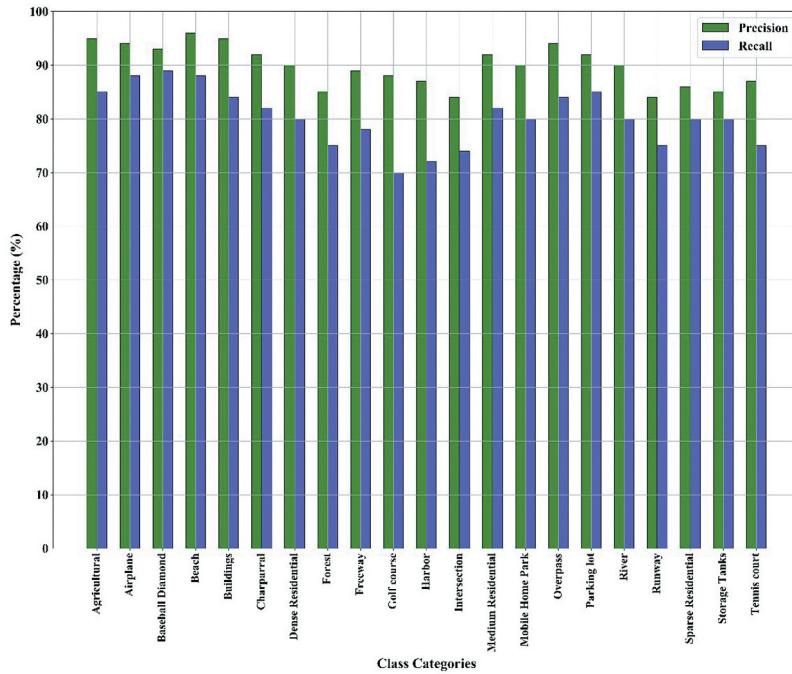


Figure 10. Percentage of precision and recall.

mitigation, surveying geologically and various other applications. With increasing demand for images obtained through satellites, which are enormously captured every day. This resulted in a dramatic increase in the amount of remote sensing images in the database. However, retrieving and classification of images accurately and meaningfully from a formless database develops as a challenge. Deep learning is

considered to the resolution for the above challenge. In this paper, deep learning based approach is proposed for remote sensing image classification. RNN along with Random forest were deployed for getting higher optimisation of cross validation on UC Merced dataset, and it was seen that training neural network can be processed using multiscale views and enhance land use classification accuracy. Using Random forest

with suitable node test lowers the testing and training cost dramatically and obtains optimal performance. The proposed method was compared with different deep learning technique for performance evaluation. RNN-RF produced 87% target accuracy of classification scene from satellite images. Future enhancement of the proposed model is it classification strategy can be considered for real time application of large complex image scene classification data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was sponsored in part by National Natural Science Foundation of China [71603288], Ministry of Education Humanities and Social Sciences Research Project [15YJC630011, 16YJC790133], National Social Science Foundation of China [14BGL203, 15CJY013].

References

- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Review article digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, 25(9), 1565–1596. <https://doi.org/10.1080/0143116031000101675>
- De Beurs, K. M., & Henebry, G. M. (2005). A statistical framework for the analysis of long image time series. *International Journal of Remote Sensing*, 26(8), 1551–1573. <https://doi.org/10.1080/01431160512331326657>
- DeFries, R. S., Field, C. B., Fung, I., Collatz, G. J., & Bounoua, L. (1999). Combining satellite data and biogeochemical models to estimate global effects of human-induced land cover change on carbon emissions and primary productivity. *Global Biogeochemical Cycles*, 13(3), 803–815. <https://doi.org/10.1029/1999GB900037>
- Interdonato, R., Ienco, D., Gaetano, R., & Ose, K. (2019). DuPLO: A DUAL view point deep learning architecture for time series classificatiOn. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, 91–104. <https://doi.org/10.1016/j.isprsjprs.2019.01.011>
- Jin, S., & Sader, S. A. (2005). MODIS time-series imagery for forest disturbance detection and quantification of patch size effects. *Remote Sensing of Environment*, 99(4), 462–470. <https://doi.org/10.1016/j.rse.2005.09.017>
- Kampffmeyer, M., Salberg, A. B., & Jenssen, R. (2018). Urban land cover classification with missing data modalities using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(6), 1758–1768. <https://doi.org/10.1109/JSTARS.2018.2834961>
- Kussul, N., Shelestov, A., Lavreniuk, M., Butko, I., & Skakun, S. (2016, July). Deep learning approach for large scale land cover mapping based on remote sensing data fusion. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 198–201). IEEE.
- Li, W., Fu, H., Yu, L., Gong, P., Feng, D., Li, C., & Clinton, N. (2016). Stacked Autoencoder-based deep learning for remote-sensing image classification: A case study of African land-cover mapping. *International Journal of Remote Sensing*, 37(23), 5632–5646. <https://doi.org/10.1080/01431161.2016.1246775>
- Li, Z., Shen, H., Wei, Y., Cheng, Q., & Yuan, Q. (2018). Cloud detection by fusing multi-scale convolutional features. In *Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science, Beijing, China* (pp. 7–10).
- Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing*, 25(12), 2365–2401. <https://doi.org/10.1080/0143116031000139863>
- Lv, Q., Dou, Y., Niu, X., Xu, J., & Li, B. (2014). Classification of land cover based on deep belief networks using polarimetric RADARSAT-2 data. *2014 IEEE Geoscience and Remote Sensing Symposium*. doi: [10.1109/igarss.2014.6947537](https://doi.org/10.1109/igarss.2014.6947537)
- Lv, Q., Dou, Y., Niu, X., Xu, J., Xu, J., & Xia, F. (2015). Urban land use and land cover classification using remotely sensed SAR data through deep belief networks. *Journal of Sensors*, 2015, 1–10. <https://doi.org/10.1155/2015/538063>
- Lyu, H., Lu, H., & Mou, L. (2016). Learning a transferable change rule from a recurrent neural network for land cover change detection. *Remote Sensing*, 8(6), 506. <https://doi.org/10.3390/rs8060506>
- Mateo-García, G., Gómez-Chova, L., & Camps-Valls, G. (2017, July). Convolutional neural networks for multi-spectral image cloud masking. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 2255–2258). IEEE.
- Murugan, N. S., & Devi, G. U. (2018a). Detecting spams in social networks using ML algorithms-a review. *International Journal of Environment and Waste Management*, 21(1), 22–36. <https://doi.org/10.1504/IJEWM.2018.091308>
- Murugan, N. S., & Devi, G. U. (2018b). Detecting streaming of Twitter spam using hybrid method. *Wireless Personal Communications*, 103(2), 1353–1374. <https://doi.org/10.1007/s11277-018-5513-z>
- Murugan, N. S., & Devi, G. U. (2019). Feature extraction using LR-PCA hybridization on twitter data and classification accuracy using machine learning algorithms. *Cluster Computing*, 22(6), 13965–13974. <https://doi.org/10.1007/s10586-018-2158-3>
- Piramanayagam, S., Schwartzkopf, W., Koehler, F. W., & Saber, E. (2016, October). Classification of remote sensed images using random forests and deep learning framework. In *Image and signal processing for remote sensing XXII* (Vol. 10004, p. 100040L). International Society for Optics and Photonics.
- Prathik, A., Anuradha, J., & Uma, K. (2018). Survey on spatial data mining, challenges and its applications. *Journal of Computational and Theoretical Nanoscience*, 15(9–10), 2769–2776. <https://doi.org/10.1166/jctn.2018.7537>
- Prathik, A., Anuradha, J., & Uma, K. (2019). A novel algorithm for soil image segmentati on using colour and region based system. *International Journal of Innovative Technology and Exploring Engineering*, 8(10), 3544–3550.
- Roy, D. P., Borak, J. S., Devadiga, S., Wolfe, R. E., Zheng, M., & Descloitres, J. (2002). The MODIS land product quality assessment approach. *Remote Sensing of Environment*, 83

- (1–2), 62–76. [https://doi.org/10.1016/S0034-4257\(02\)00087-1](https://doi.org/10.1016/S0034-4257(02)00087-1)
- Scott, G. J., England, M. R., Starms, W. A., Marcum, R. A., & Davis, C. H. (2017). Training deep convolutional neural networks for land-cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(4), 549–553. <https://doi.org/10.1109/LGRS.2017.2657778>
- Wang, H., Zhao, X., Zhang, X., Wu, D., & Du, X. (2019). Long time series land cover classification in china from 1982 to 2015 based on bi-lstm deep learning. *Remote Sensing*, 11(14), 1639. <https://doi.org/10.3390/rs11141639>
- Xie, F., Shi, M., Shi, Z., Yin, J., & Zhao, D. (2017). Multilevel cloud detection in remote sensing images based on deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(8), 3631–3640. <https://doi.org/10.1109/JSTARS.2017.2686488>
- Yang, Y., & Newsam, S. (2010, November). Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems* (pp. 270–279).
- Zhan, Y., Wang, J., Shi, J., Cheng, G., Yao, L., & Sun, W. (2017). Distinguishing cloud and snow in satellite images via deep convolutional network. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1785–1789. <https://doi.org/10.1109/LGRS.2017.2735801>
- Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2019). Joint deep learning for land cover and land use classification. *Remote Sensing of Environment*, 221, 173–187. <https://doi.org/10.1016/j.rse.2018.11.014>
- Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40. <https://doi.org/10.1109/MGRS.2016.2540798>