IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

ISSN 2347–3657

International Journal of
Information Technology & Computer Engineering

Volume 13, Issue 1, 2025

TENSOR FLOW BASED FINE-TUNED CONVOLUTIONAL NEURAL NETWORK MODEL FOR SCENE CLASSIFICATION OF AERIAL IMAGES

[1]Dr. R.Bhavani, [2]Dr. B.Rajalingam, [3]Dr. K.Sampath, [4]N.Mahboob Subani

Professor and HOD[1,3], Associate Professor & HOD[2], Assistant Professor[4]

Department of CSE, Annamalai University, Chidambaram[1]

St.Martin's Engineering College, Dhulapally, Secunderabad – 500100[2,3,4]

bhavaniaucse@gmail.com[1], rajalingam35@gmail.vom[2]

Abstract

Scene classification of aerial images has received growing attention from the research community in recent years. Conventional classification algorithm such as k-means clustering, Support Vector Machine (SVM), Decision Tree, Expectation Maximization (EM algorithm), k-nearest neighbors (k-NN), Ada Boost, Navie Bayes, Artificial Neural Networks (ANN) etc., use spectrum, texture and profile based feature for classification. Extracting such features in high resolution remote sensing images creates challenges to object classification. So we have to introduce the convolutional neural network (CNN) for classifying the remote sensing images. The architecture consist of three convolution layers of 64 filters with kernel size of 5x5, pooling size of 2x2 and fully connected layer 1024 and 512 respectively. In the experiments, 8,000 remote sensing images has been collected from Pattern Net dataset for the performance assessment. The experimental results show the higher accuracy for the remote sensing image classification.

Keywords: scene classification, convolutional network, aerial images and convolutional layer.

I. INTRODUCTION

Scene Classification plays an important role in remote sensing images [1]. Its applications ranges from agriculture monitoring, environmental monitoring, land use/ land planning, urban planning, surveillance [2], geo-graphic mapping, disaster control [3], object detection, etc. A wide number of techniques have been developed for object classification [4]. In general the object classification methods are divided into three categories based on the features they use, namely handcraft feature learning method, unsupervised feature learning method and deep feature learning based method [5]. Earlier, scene classification was based on the handcraft feature learning based method. This method [6] [7] was mainly used for designing the engineering features, such as colour, shape, texture, spatial and spectral information. The unsupervised feature learning method [8] is an alternative for the handcrafted features method and learning the unlabeled data for remote sensing image classification. The aim of unsupervised feature learning method is used to identify the low-dimensional features that capture some underlying high-dimensional input data. When the feature learning is performed in unsupervised way, it enables a form of semi-supervised learning where features learned from an unlabeled dataset are then employed to improve performance in a supervised setting with labelled data. There are several unsupervised feature learning methods available such as k -means clustering, principal component analysis (PCA), sparse
coding and auto encoder. In real time applications, the unsupervised feature learning methods have achieved high performance for classification compared to handcraft feature learning methods[9]. However, the lack of semantic information provided by the category label cannot promise the best discrimination between the classes. So we need to improve the classification performance and to extract powerful discriminant features for improving classification performance.

II. RELATED WORKS

Aerial images are more valuable tool for monitoring the earth surface and mainly used for various applications such as surveillance, agricultural monitoring, metrology, mineralogy and environmental science. A lot of classification methods have been proposed to deal with the remote sensing image classification. The traditional object classification based on either supervised or unsupervised learning methods. Some researchers have proposed object classification using supervised learning methods such as support vector machine, artificial neural networks, random forest, k-nearest neighbours, decision tree and sparse representation classifier [4]. The support vector machine is a supervised non-parametric learning technique, therefore no assumption made on data distribution. Yakoub Bazi et al. [13] have introduced optimal SVM classification system for hyper spectral remote sensing images. The system classify scene based on genetic optimization framework formulated in discriminative features and SVM parameters. Rick Archibald et al. [14] have proposed embedded feature-selection (EFS) algorithm that is tailored to operate with support vector machines (SVMs) to perform band selection and classification. The artificial neural network (ANN) is non-parametric learning technique, which is more sophisticated and robust method for image classification. The ANN produces higher accuracy from few training data. T. Kavzoglu et al. [15] have proposed back propagating artificial neural networks in land cover classification. The Random forest is an ensemble classifier that produces multiple decision trees using randomly selected training samples. Mariana Belgiu et al. [16] have done a survey the random forest in remote sensing and their applications. Matthew M. Hayes et al. have introduced High-resolution land cover classification using Random Forest [17]. Anil M. Cheriyadat et al. have proposed an unsupervised feature learning approach for scene classification [18]. Unsupervised feature learning approach is to generate feature representation for various high-resolution aerial scenes. Chaib et al. [19] have introduced an informative feature selection method for high resolution satellite image classification by using PCA classifier. Zheng et al. have developed a hybrid model satellite image classification using multi feature joint sparse coding [20]. Sheng et al. have proposed a sparse coding based multi feature for image classification [21]. But, the sparse coding is more expensive when dealing with big data. Therefore, the sparse feature coding can be used only for small scale problems.

In recent years, CNN has been used in various remote sensing applications, such as object classification, land use scene classification and object detection. Maarten et al. [22] have proposed an object recognition using deep convolutional neural networks with complete transfer and partial frozen layers. Qin Zou et al. have developed a deep learning based feature selection for remote sensing scene classification. The popular deep-learning technique, i.e., the deep belief network (DBN), has achieved feature abstraction by minimizing the reconstruction error over the whole feature set, and features with smaller reconstruction errors would hold more feature intrinsic for image representation [23]. The Deepan et al. have introduced very deep convolutional network for scene classification of remote sensing images data [24].

III. PROPOSED WORK

This section details the proposed method for scene classification of aerial images. There are many deep learning methods exist, but the CNN is common method for classifying the scene and obtain the excellent result for different kind of input. The LeCun et al. [25] introduced the first CNN model. The architecture of this model is shown in Figure 1. In this figure C1, C2, C3 represents convolutional layers, S1, S2, S3 represents sub sampling layers, FC1 and FC2 represents fully connected layers and SM represents soft max function. These architectures is mainly used to classify the object scenes like bridge, beach and forest.

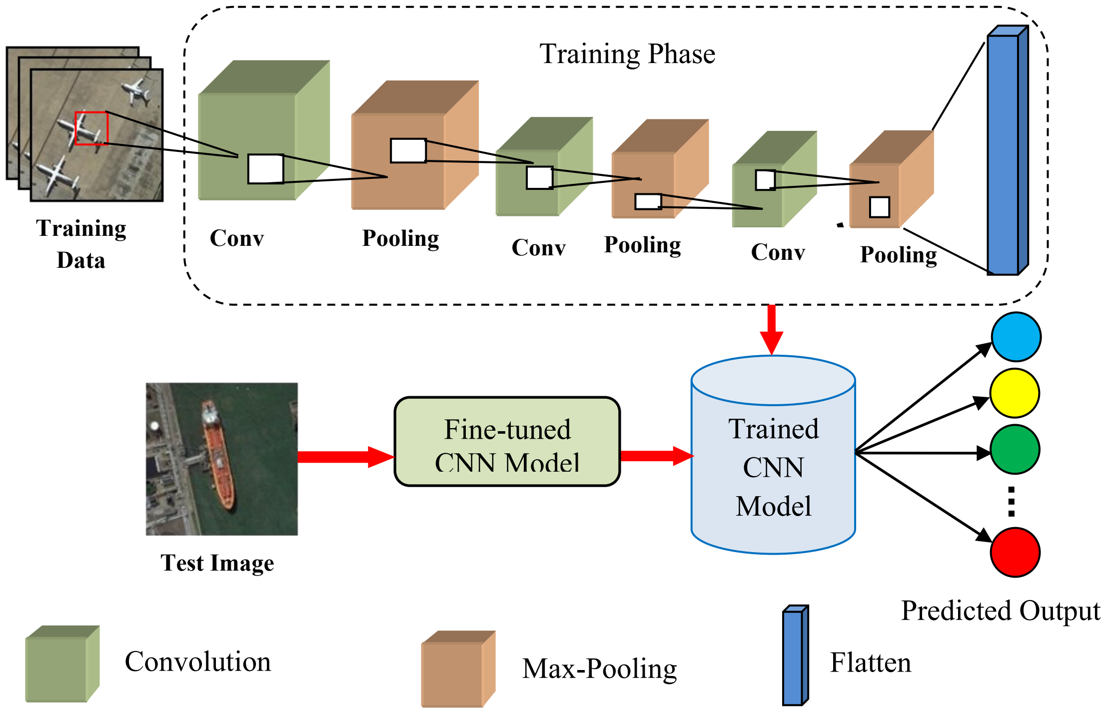

Figure 1. Architecture of Proposed fine-tuned Convolutional Neural Networks

The CNN is composed of three convolutional layers, three max-pooling layers and two fully connected layers. The basic functionality of CNN models is divided into three common categories [26].

3.1 Convolutional Layer

The convolutional layers extract different features for the input image. There are three convolutional layers in the model.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 1, 2025

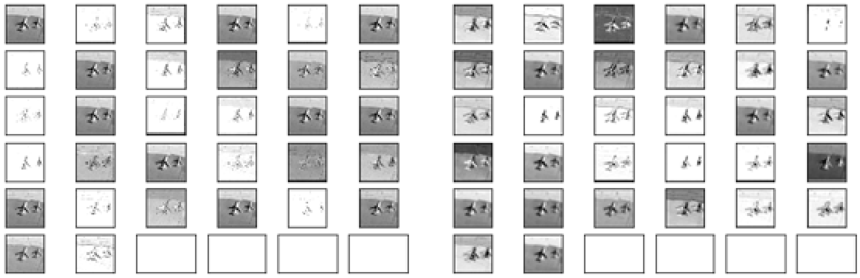

Figure 2. Airplane Visualization of conv1 and conv2 filter from CNN Model.

The first convolutional layer extracts low level features such as lines and edges. Higher level layer extracts high level features. The size of input and kernel is 256x256 and 5x5. Starting from top left corner, kernel is moved from left to right an element. Once the corner is reached kernel moves one element in down word direction and again move back to same place. The process repeated until it reaches the bottom right corner. For each case when N=256, k=5 there are 252 unique position from left to right and top to bottom. Corresponding to these positions, each feature in the output will contain 256x256 (i.e., (N-k+1) x (N-k+1)) elements. The Figure 2 visualizes the conv1 and conv2 filter for the airplane image using the CNN model.

3.2 Pooling Layer

The sub sampling layer is used to reduce the feature resolution. This layer reduces the number of connection between the convolutional layers, so it will lower the computational time also. There are three types of pooling: max pooling, min pooling and average pooling. In each case, the input image is divided into non-overlapping two dimensional spaces. For example, if the input size is 4x4 and sub sampling size is 2x2, a 4x4 image is divided into four non-overlapping of matrices 2x2. For max pooling, the maximum value of the four values is selected. In the case of min pooling, the minimum value of the four values is selected. The figure 2 shows the operation of max pooling and average pooling process.

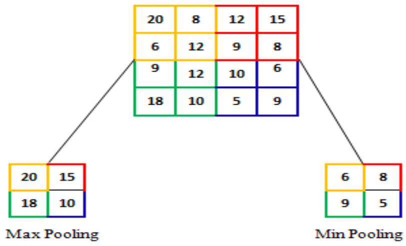

Figure 3. Pictorial representation of Max pooling and Average pooling process.

3.3 Activation Function

The Activation Function improves the CNN performance. There are many activation function exist namely, tanh, sigmoid and ReLU. In this paper, ReLU activation function has been used. The Rectified linear unit (ReLU) is one of the standard and popular activation functions in the last few years. The ReLU activation function is defined as:

$$b_{i,j,k} = \max\,(a_{i,j,k}, 0) \qquad (1)$$

where, $a_{i,j,k}$ is the input of the activation function at location (i, j) on the k-th channel. In this layer we remove every negative values from the filtered images and replaces it with zeros. The Figure 4 elaborates the process of activation function.

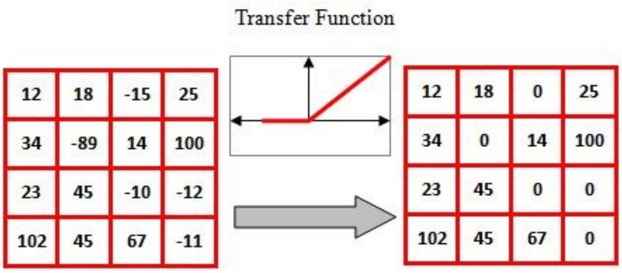

Figure 4.Pictorial representation of Activation Function

3.4 Fully Connected Layer

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 1, 2025

After the several convolutional and pooling layer processes, the two-dimensional data is converted into one-dimensional vector. The one dimensional data will be the input for fully connected layers. There may be one or more hidden layers which perform high level reasoning. Each neuron uses the data from the previous layers to multiplies the connection weights and add a bias value. The output of final fully connected layers fed into the classifier ie. softmax function. The softmax function is used to classify the object. The general form of softmax is defined as in Eq. (2)

$$\text{class}_j = \frac{\exp(sf_j)}{\sum_q (sf_q)} \qquad (2)$$

where, exp (sf_j) is the probabilities of each target class where as sf_q is possible of all the target classes.

3.5 Training a CNN Model

The implementation of Convolutional Neural Networks based scene classification of aerial images using tensorflow is done by Python jupyter notebook. The input image 256x256x3 is gives to first convolutional layer with kernel size of 3x3x64 and stride of 2. The output of first convolutional layer is the input to the second convolutional layer with kernel size of 3x3x64 and strides 2 pixel. Similarly the third convolutional layer takes input to the third layer with kernel size of 3x3x64 and stride 2. The fully connected layer has 1024 and 512 neurons respectively. Finally, soft max function classifies the ten class remote sensing images. The same operations was performed after fine tuned the hyperparamter of filer size 3x3 and 5x5 respectively.

IV. DATASET DESCRIPTION

For evaluating our proposed approach, we randomly choose ten classes such as airplane, baseball court, beach, bridge, forest, harbor, parking lot, river, runway and storage tank from Pattern Net dataset. The Pattern Net dataset is a large scale high-resolution remote sensing image that has been extracted from Google Earth imagery [27]. The dataset contains 30,400 images which consist of 38 classes. Each class contains 800 images of size of 256x256 pixels. The Figure 5 shows image examples of different classes for scene classification. In the CNN model, the Pattern Net dataset has been split into training, validation and testing datasets separately. The training and testing dataset description are shown in Table 1. and validation sample image is randomly chosen from training sample based on the validation size. In each image, three spectral bands were used including red, green and blue.

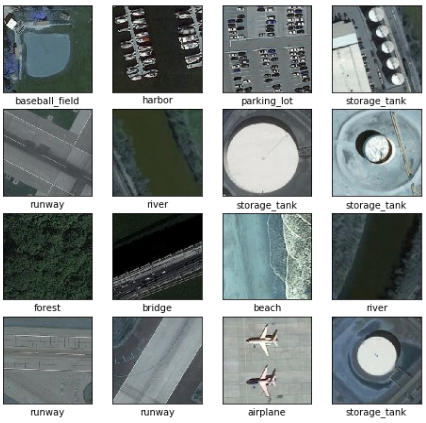

Figure 5. Image examples of different classes from the Pattern Net dataset

S. No.	Class Name	No. of Training Samples	No. of Validation Samples	No. of Testing Samples
1.	Airplane	640	80	80
2.	Baseball field	640	80	80
3.	Beach	640	80	80
4.	Bridge	640	80	80
5.	Forest	640	80	80
6.	Harbor	640	80	80
7.	Parking lot	640	80	80
8.	River	640	80	80
9.	Runway	640	80	80

ISSN 2347–3657

Volume 13, Issue 1, 2025

10.	Storage tank	640	80	80
Total		**6400**	**800**	**800**

Table 1. Dataset description for scene classification

4.1 Performance Metrics

We have evaluated the performance of a proposed model by using various performance metrics such as Accuracy, Precision, Recall and F1-measure. The Accuracy can be calculated by the number of properly classified data in a dataset divided by the total number of samples, as shown in the equation (3).

$$Accuracy = {}^{t}/_{n} \qquad (3)$$

where t is a number of properly classified samples and n is a total number of samples in a dataset.

The precision can be measured by number of properly classified data in a datasets divided by total number of all samples in a class. Precision value of the class c, P_c can be shown in equation (4) where, t_c is a total number of properly classified samples in class c and n_c is a total number of samples in the class c.

$$P_c = t_c / n_c \qquad (4)$$

The recall can be measured by number of properly classified datas are divided by the number of all relevant samples in the corresponding class. Recall value of the class c, R_c can be shown in equation (5) where, t_c is a total number of properly classified samples in class c and k_c is number of samples classified as relevant to class c.

$$R_c = t_c/k_c \qquad (5)$$

The F1-measure (harmonic mean) is used to show the balance between the precision and recall measures. F1-score value can be calculated using equation (6):

$$F_1 = (2 * (P_c * R_c))/(P_c + R_c \qquad (6)$$

It is essential to find the confusion matrix while calculating the performance measures. Confusion matrix is a technique used to summarise results and used for validating classification methods. There are two common classes, which are usually deal with confusion matrix namely positive class and negative class. These two common classes can be further divided into four categories. True Positive is an outcome, where the model that has correct classification of positive example. False Negative is an outcome, where the model that has incorrect classification of positive examples. False Positive is an outcome, where the model that has incorrect classification of positive examples. True Negative is an outcome, where the model that has correct classification of negative examples. The concept of confusion matrix is shown in figure 6.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this experiment, we have focused on Convolutional Neural Network based scene classification for aerial images using tensorflow. The accuracy may vary based on the changing hyper parameter values in CNN model. We tested the proposed model in four different types of hyper parameter such as number of filter 32 and 64 with kernel size 3x3 and 5x5. Based on the performance of result, number of filter 64 with 5x5 kernel size gives more accuracy rather than other filter with kernel. The different accuracy results are shown in Table 2.

Table 2. Classification accuracy based on various hyper parameters

Kernel Size	Filter Size	FC1 and FC2	Activation Function	Accuracy
3x3	32	1024 & 512	ReLU	89.5
3x3	64			90.7
5x5	32			91.5
5x5	**64**			**95.2**

Figure 7. Some miss- prediction data items from Pattern Net Dataset

The Figure 7 describes some sample miss-prediction error for image scene classification. The correctly classified data items are placed in diagonal of matrix, remaining miss predicted data items are placed above and below the diagonal of matrix. We found that the error occurs when ‘bridge’ classified as ‘tennis court’ and the ‘tennis court’ classified as ‘bridge’. The confusion matrix for scene classification is shown in figure 8.

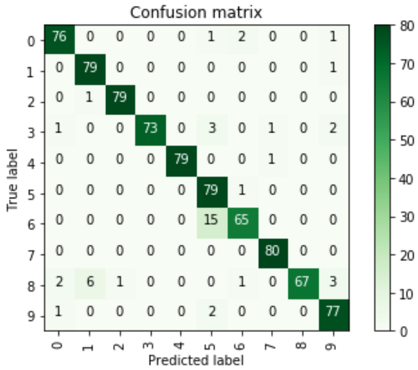

Figure 8. Confusion Matrix for Proposed model

Figure 9 shows the accuracy of the CNN model for 50 epochs on training and validation datasets. Under the framework of tensor flow backend, the entire process was run in Core i5 CPU 2.6GHz, 1 TB of Hard Disk and 4 GB of RAM.

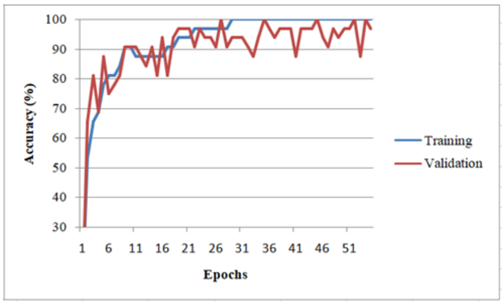

Figure 9. Training and Validation Accuracy of CNN Model

VI. CONCLUSION

In this paper, we have proposed an approach convolutional neural network based scene classification for aerial images using tensorflow. The CNN model act as a feature extractor and classifier for the given training images as well as validation images. The use of CNN outperforms other traditional method and allowed us to achieve 95.2% of accuracy for ten class scenes such as airplane, baseball court, beach, bridge, forest, harbor, parking lot, river, runway and storage tank. In future, we have planned to implement the proposed work in GPU configuration for reducing the computational time.

References

1. L. Janssen and H. Middelkoop, “Knowledge based crop classification of a Landsat thematic mapper image,” Int. J. Remote Sens., vol. 13, no. 15, pp. 2827–2837, 1992.
2. Blaschke, T., “Object based image analysis for remote sensing”, ISPRS J. Photogramm. Remote Sens. 65, 2010, pp. 2-16.
3. Draguţ, L., Blaschke, T., “Automated classification of landform elements using object-based image analysis”, Geomorphology 81, 2006. , pp. 330–344.
4. P.Deepan, L.R. Sudha, “Remote Sensing Image Scene Classification using Dilated Convolutional Neural Network”, International Journal of Emerging Trends in Engineering Research, Vol. 8, No.7, pp.3622-3630, 2020.
5. Gong Cheng, Junwei Han and Xiaoqiang Lu, “Remote Sensing Image Scene Classification Benchmark and State of the Art”, Proceedings of the IEEE, 2018, pp. 1-19.
6. L. Zhao, P. Tang, and L. Huo, “Feature significance- based multi bag-of-visual-words model for remote sensing image scene classification,” J. Appl. Remote Sens., vol. 10, no. 3, p. 035004, 2016.
7. K. Qi, H. Wu, C. Shen, and J. Gong, “Land use scene classification in high-resolution remote sensing images using improved correlatons,” IEEE Geosci. Remote Sens. Lett., vol. 12, no. 12, pp. 2403–2407, Dec. 2015.

8. Yansheng Li, Chao Tao, Yihua Tan, Ke Shang, and Jinwen Tian, "Unsupervised Multilayer Feature Learning for Satellite Image Scene Classification" IEEE Geoscience and Remote Sensing Letters, Vol. 13, NO. 2, Feb. 2016.
9. T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" IEEE Trans. Image Process., vol. 24, no. 12, pp. 5017–5032, 2015.
10. L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state Thomas, A.K., Shekhar, H., Bhagyalakshmi, L., Suman, S.K., Sreelakshmy, R., Bhasha, A. (2025). IoT Enabled Smart Street Light and Air Quality Control. In: Kumar, A., Gunjan, V.K., Senatore, S., Hu, YC. (eds) Proceedings of the 5th International Conference on Data Science, Machine Learning and Applications; Volume 1. ICDSMLA 2023. Lecture Notes in Electrical Engineering, vol 1273. Springer, Singapore. https://doi.org/10.1007/978-981-97-8031-0_90
11. Shaik, M.A., Shekhar, H., Bhagyalakshmi, L., Patra, P.S.K., Suman, S.K., Prabakaran, G. (2025). Text Detection and Recognition of Characters in Medical Prescription. In: Kumar, A., Gunjan, V.K., Senatore, S., Hu, YC. (eds) Proceedings of the 5th International Conference on Data Science, Machine Learning and Applications; Volume 1. ICDSMLA 2023. Lecture Notes in Electrical Engineering, vol 1273. Springer, Singapore. https://doi.org/10.1007/978-981-97-8031-0_93
12. Bhagyalakshmi, L., Srivastava, R., Shekhar, H., Suman, S.K. (2023). A Vision for Industry 4.0 Utilising AI Techniques and Methods. In: Mishra, A., Lin, J.CW. (eds) Industry 4.0 and Healthcare . Advanced Technologies and Societal Change. Springer, Singapore. https://doi.org/10.1007/978-981-99-1949-9_11
13. Suman S. K., Shekhar H, Mahaton C. B, D. Gururaj, L. Bhagyalakshmi, P. S. K. Patra, "Sign Language Interpreter", Advances in cognitive science and communications, Springer, ICCCE 2022, pp. 1021-1031. eBook ISBN 978-981-19-8086-2. https://doi.org/10.1007/978-981-19-8086-2
14. Sanjay Kumar Suman, Rajiv Srivastava, L. Bhagyalakshmi and Himanshu Shekhar, "On Safety Evaluation of Artificial Intelligence Based Systems to Software Reliability", Multi-Criteria Decision Models in Software Reliability, 1e, CRC Press, Taylor & Francis Group, USA. 2022. eBook ISBN9780367816414. DOI: 10.1201/9780367816414Belgiu, M., and L. Drăguţ. 2016. "Random Forest in Remote Sensing: A Review of Applications and Future Directions." ISPRS Journal of Photogrammetry and Remote Sensing 114: 24–31.
15. Matthew M. Hayesa, Scott N. Millera and Melanie A. Murphya, "High-resolution landcover classification using Random Forest", Remote Sensing Letters, Vol. 5, No. 1, pp. 112–121, 2014.
16. A. M. Cheriyadat, "Unsupervised feature learning for aerial scene classification," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 1, pp. 439–451, Jan. 2014.
17. S. Chaib, Y. Gu, and H. Yao, "An informative feature selection method based on sparse PCA for VHR scene classification," IEEE Geosci. Remote Sens. Lett., vol. 13, no. 2, pp. 147–151, Feb. 2016.
18. P. Deepan, L.R Sudha, "Object Classification of Remote Sensing Image Using Deep Convolutional Neural Network", The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems, pp.107-120, 2020.
19. G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," Int. J. Remote Sens., vol. 33, no. 8, pp. 2395–2412, 2012.
20. Maarten C. Kruithof, Henri Bouma , Noëlle M. Fischer, Klamer Schutte, " Object recognition using deep convolutional neural networks with complete transfer and partial frozen layers ", Proc. Society of Photo-Optical Instrumentation Engineers, Vol. 9995, Mar.2016.
21. Qin Zou, Lihao Ni, Tong Zhang and Qian Wang, "Deep Learning Based Feature Selection for Remote Sensing Scene Classification", IEEE, 2015.
22. P. Deepan, L.R. Sudha, "Fusion of Deep Learning Models for Improving Classification Accuracy of Remote Sensing Images", Journal Of Mechanics of Continua and Mathematical Sciences, pp. 189-201, 2019.
23. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
24. Keiron O'Shea and Ryan Nash, "An Introduction to Convolutional Neural Networks", International Journal of Computer Vision and Pattern Recognition, Nov.2015, pp.2-11.
25. Hari Prasad Bhupathi, Srikiran Chinta, 2024. "AI-Powered Efficiency Machine Learning Techniques for EV Battery Charging" ESP International Journal of Advancements in Science & Technology (ESP-IJAST) Volume 2, Issue 3: 64-73.