Scene Classification of Remotely Sensed Images using Optimized RSISC-16 Net Deep Convolutional Neural Network Model

P. Deepan[1,*], L.R. Sudha[2], K. Kalaivani[3] and J. Ganesh[4]

[1]Assistant Professor, Department of CSE (AI&ML), St. Martin's Engineering College, Telangana, India. deepanp87@gmail.com
[2]Associate Professor, Department of CSE, Annamalai University, Tamilnadu, India. sudhaselvin@ymail.com
[3]Associate Professor, Department of CSE, E.G.S Pillay, Nagapattinam, Tamilnadu, India, kalai4best@gmail.com
[4]Assistant Professor, Department of CSE, Anjalai Ammal Mahalingam Engineering College,Tamilnadu, India, jaygan85@gmail.com

Abstract

Remote Sensing Image (RSI) analysis has seen a massive increase in popularity over the last few decades, due to the advancement of deep learning models. A wide variety of deep learning models have emerged for the task of scene classification in remote sensing image analysis. The majority of these models have shown significant success. However, we found that there is significant variability, in order to improve the system efficiency in characterizing complex patterns in remote sensing imagery. We achieved this goal by expanding the architecture of VGG-16 Net and fine-tuning hyperparameters such as batch size, dropout probabilities, and activation functions to create the optimized Remote Sensing Image Scene Classification (RSISC-16 Net) deep learning model for scene classification. Using the Talos optimization tool, the results are carried out. This will increase efficiency and reduce the risk of over-fitting. Our proposed RSISC-16 Net model outperforms the VGG-16 Net model, according to experimental results.

Keywords: Optimized RSISC-16, Scene classification, remote sensing image, and convolutional neural network.

Received on 06 September 2021, accepted on 25 January 2022, published on 01 February 2022.

Copyright © 2022 P. Deepan *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license, which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/_____

*Corresponding author. Email: deepanp87@gmail.com

1. Introduction

The evolution of remote sensing image classification from pixel and object level to scene level starts in 2010, after the emergence of land use/land cover area. Here the term 'scene' represents a local area cropped from a large scale satellite image. The Remote Sensing Image Scene Classification (RSISC) study, which has received a lot of attention, intends to classify remote sensing images with a set of semantic categories by analyzing differences in the spatial arrangement and structural pattern of ground objects[1].

The aim of RSISC is to correctly label the remote sensing images with equivalent semantic classes. Figure 1 shows how to categorize an urban area image into commercial building, residential area, and industrial area. In general, different types of ground objects can be found in remote sensing images [2]. An industrial scene may contain roads, trees, buildings and so on. Scene classification is a challenging task when compared to object-based classification since the scenes contain a variety of

complicated spatial ground objects that do not have a consistent shape and structure.
The study of Simonyan and Zisserman [3] in Neural Information Processing Systems (NIPS) 2012 and success in the International Large Scale Visual Recognition Challenge (ILSVRC) 2012 ImageNet competition popularized the use of deep neural networks for image classification and recognition. This improvement in performance motivated many other researchers to focus on deep neural networks in their own specific problems, and deep learning is now a hot topic in vision research [4].

Figure 1. Sample classes of urban area in remote sensing images (a) commercial building (b) residential area and (c) industrial area

Almost every day, a new scientific paper is published to improve deep learning solutions to vision problems. On the other hand, the availability of high-quality sensors, which is combined with an improved aerospace and satellite industry, allows researchers to collect larger amounts of remote sensing data with higher spectral and spatial resolution.
Thus, increasing the quality and number of remote sensing images allows researchers to attack this problem and makes deep learning for remote sensing possible. Deep learning in remote sensing is used to create a fully automated system that can classify geospatial objects and land cover into distinct classes such as airplane, barren land, building, cultivated field, forest, roadway, runway, ship, storage tank, water, and so on. It is critical to be able to classify land use/land cover in order to monitor the Earth's constant changes and manage urban development [5-7].
Utilization of machine learning techniques for this purpose is quite critical and challenging due to the small amount of remote sensing imagery that is available with ground truth labels. Therefore, many computer vision scientists have proposed a number of different deep learning algorithms to extract information from remote sensing images and significantly contributed to the literature of the computer vision and the remote sensing field.
Though several studies of RSISC have been made in the past few decades, no algorithm has yet been developed to accurately classify RSI scenes. The following are some of the problems faced by researchers while performing remote sensing image scene classification.

- There is high intra-class diversity in remote sensing images.
- There is a lot of inter class similarity between scenes.
- Aerial images have much larger scale variation than conventional images.
- Several ground objects are presenting in the same scene with complex background.

In addition to these, the quantity and quality of the images create a high computational cost which makes it difficult for near-real time applications.
The main motivation of our proposed work is to develop remote sensing scene image classification model using deep learning [21, 22] to extract features automatically and to classify the scenes accurately which in turn handles the problem of intra class diversity, inter class similarity, large scale variation present in aerial images and complex background scenes. Our proposed deep learning RSISC -16 Net Model is also optimized with Talos tool using hyperparameters of batch size, drop out probabilities and activation function that will reduce computational cost.
The remainder of the paper is structured as follows: Section "Related works" contains the literature survey of CNN classification for remote sensing images. Section "Proposed work" presents the newly developed optimized RSISC-16 Net model, Section "Experimental result and analysis" discusses how the proposed model performance is improved from VGG-16 Net model; and in Section "Conclusion" we reiterate the focus of the paper and summarize the work presented.

2. Related Works

Convolutional Neural Networks (CNN) is a broad idea that can be used to apply scene classification methods. LeCun et al. [8, 9] created the first CNN model, which is similar to a standard Artificial Neural Network (ANN) and serves as the basis for modern CNN. The neurons in animal and human brains provide inspiration for the CNN model's structure. In recent days, researchers have developed many models related to image classification problems.
For example, Liu et al. [10] developed Siamese networks for scene classification using remote sensing images. The results demonstrated that the performance of the Siamese CNN model is efficient and superior to the VGG-16 (Visual Geometry Group) model. The research in [11] suggested a CNN model for a road recognition system based on remote sensing images. Cheng et al. [12] presented a discriminative CNN model to improve RSI scene classification performance, which addresses both within-class diversity and between-class similarity issues. By using CNN-based sparse coding learning techniques, Qayyum et al. [13] established an efficient method for scene classification of aerial images. A capsule network for RSI-scene classification was introduced by Zhang et al. [14]. To improve the classification accuracy, this model first extracts features using CNN and then feeds the extracted features into a capsule network. The individual scene classification models do not efficiently classify the scene of remote sensing images. So, in order to improve the scene classification, two or three individual CNN models are combined. It is sometimes called as ensemble classification.
The ensemble classification model, also known as the fusion model, is widely used for image classification from remote

sensing images. The goal is to combine the results of two or more individual models. Fusion or combining the features of two or three models is seen to be the best solution to overcome the limitations of individual classification models. Many researchers developed feature level and decision level scene classification of remote sensing images.

Chaib et al. [15] combined VGG-16 and Inception model features to create a feature fusion model for high resolution remote sensing image scene classification. In [16], a deep learning decision level fusion was introduced for improving the classification accuracy of remote sensing images. This method combined the decision level features of three state-art-of the models, namely traditional CNN, VGG-16, and ResInception, to achieve greater accuracy than the individual models. Dong et al. [17] developed a combined deep learning model for High Resolution-RSI scene classification. This model combines the representation of CNN features with the LSTM model to improve scene classification accuracy. For land cover classification of HRI, Scott et al. [18] introduced a fusion technique in which multiple deep CNN models such as CaffeNet, GoogLeNet, and ResNet50 features were extracted. Travis et al. [19] introduced ensemble based image classification by using wavelet transform. This model converts the data into wavelet domain to achieve better accuracy and efficiency for image classification.

All the above mentioned individual models are not efficiently classify the scenes and also the fusion models require more computational time to train and validate the data. Taking the above disadvantages into consideration, we have proposed an optimized RSISC-16 Net model based on VGG-16 Net for scene classification of remote sensing images. When compared with VGG-16 Net model, our propsed RSISC-16 Net model requires only less number of parameters.

3. Proposed Work

In this section, we have proposed RSISC-16 Net model by extending the architecture of VGG-16 Net. The RSISC-16 Net model consists of a total of 13 convolutional layers, 5 pooling layers, two fully connected layers and one soft-max classifier in 5 different blocks as specified below:

- First two blocks have 2 convolutional layers each
- Rest of three blocks have 3 convolutional layers each
- 5 pooling layers one in each block

The aim of convolution layer is to extract the low, mid and high level features from the given training datasets. The proposed baseline architecture of RSISC-16 Net model is shown in Figure 2.

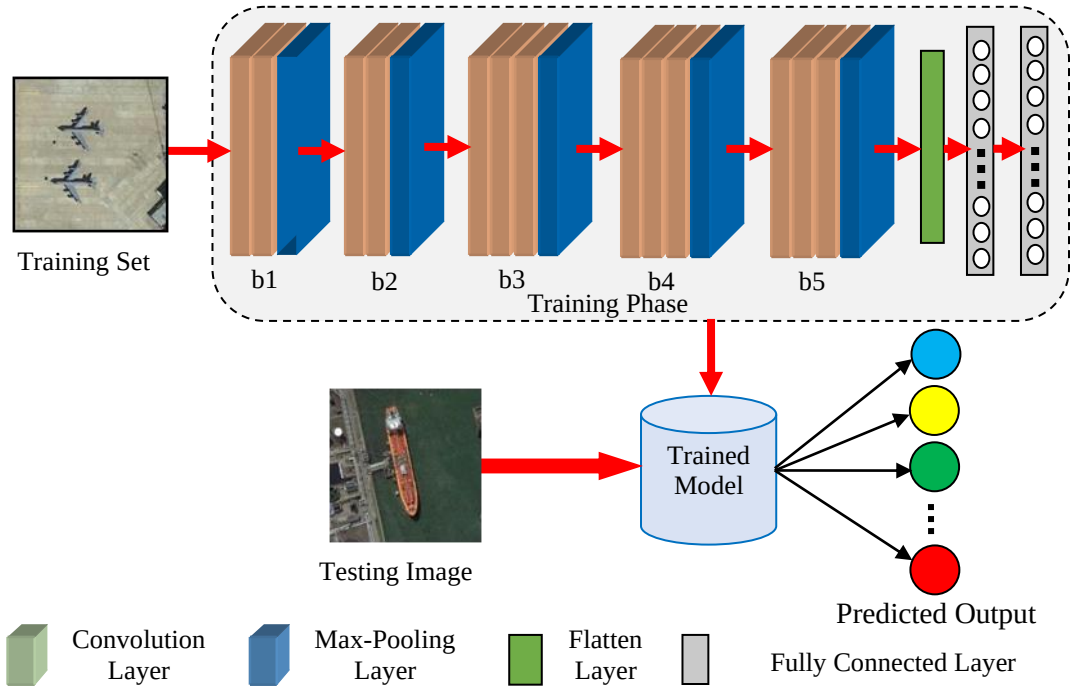

Figure 2. Proposed architecture of RSISC-16 Net Model

The input images are processed using a convolutional and pooling layer sequence. The first two blocks of convolutional layers extract the low-level features like lines, edges and shapes, which the next three blocks extract high level feature like internal shape of scenes. Every convolutional filter has a kernel size of 3 × 3 and a stride of 1. The depth of the convolutional layer is gradually increased from 32 to 64.

The summary of the proposed optimized RSISC-16 Net model is shown in Table 2 and visualization of each feature maps is shown in Figure 4. The parameter of each convolutional layer is calculated by using the following equation.

$$P_C=((F\times F)\times D+1)\times L) \quad (1)$$

where Pc represents a parameter calculation in convolutional layer, F represents the filter size, D represents the dimension of features and L represents number of layers. Similarly, parameter of fully connected layers is calculated by the equation (4.6):

$$P_{FC}=(D+1)\times L) \quad (2)$$

where PFC represents a parameter calculation in fully connected layer, D represents the dimension of features and L represents a number of layers.

4. Experimental Results and Analysis

Using the NWPU-RESISC 45 class dataset, the performance of the VGG-16 Net model and the proposed optimized RSISC-16 Net model was tested. Experiments are conducted in Jupyter Notebook and Anaconda Prompt IDE with different deep learning libraries such as Numpy, Matplotlib, open CV and sklearn. Both models were trained and tested using Keras and TensorFlow in a corei7 CPU 2.6GHz, 1TB hard disk drive, and 8GB of RAM with 7000 remote sensing images.

4.1. Performance Metrics

To demonstrate the effectiveness of the proposed optimized RSISC-16 Net model, performance metrics such as Precision, Recall, F1-score, and Accuracy are calculated using the confusion matrix of the model.

Precision

Precision is one of the most effective ways to demonstrate how the model is accurate. The ratio of accurately predicted positive observations to total predicted positive observations can be used to measure it. Precision value can be calculated using the equation (3).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

Recall

Recall is the ratio of correctly predicted positive observations to all observations in the actual class. It is used to calculate how many of the actual positives the model catches by labelling it as positive. Recall value can be calculated using the equation (4).

$$\textbf{Recall} = \frac{\textbf{TP}}{\textbf{TP+FN}} \quad (4)$$

Accuracy

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 (5).

$$\textbf{Accuracy} = \frac{\textbf{TP+FP}}{\textbf{TP+FP+TN+FN}} \quad (5)$$

F1-Score

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

$$\mathbf{F} = \mathbf{2} * \frac{\textbf{Precision}*\textbf{Recall}}{(\textbf{Precision}+\textbf{Recall})} \quad (6)$$

4.2. Dataset Descriptions

Dataset plays a crucial role in developing and evaluating various scene classification models. We have used NWPU 45-class publicly available datasets for scene classification of remotely sensed images. The dataset which was extracted from Google Earth and covers the high resolution remotely sensed images in more than 100 countries. It is released by the North Western Polytechnical University and is currently the largest scene classification dataset.

This dataset consists of 31,500 remote sensing images which are categorized into 45 classes. For each class, there are 700 images with the size of 256×256 resolution in the Red-Green-Blue (RGB) color space. The spatial resolution may vary from the 30m to 0.2m for each pixel. To evaluate the effectiveness of the proposed approach, we have chosen ten classes, namely airplane, beach, commercial area, desert, forest, lake, overpass, river, tennis court and wetland from the above mentioned benchmark datasets for scene classification.

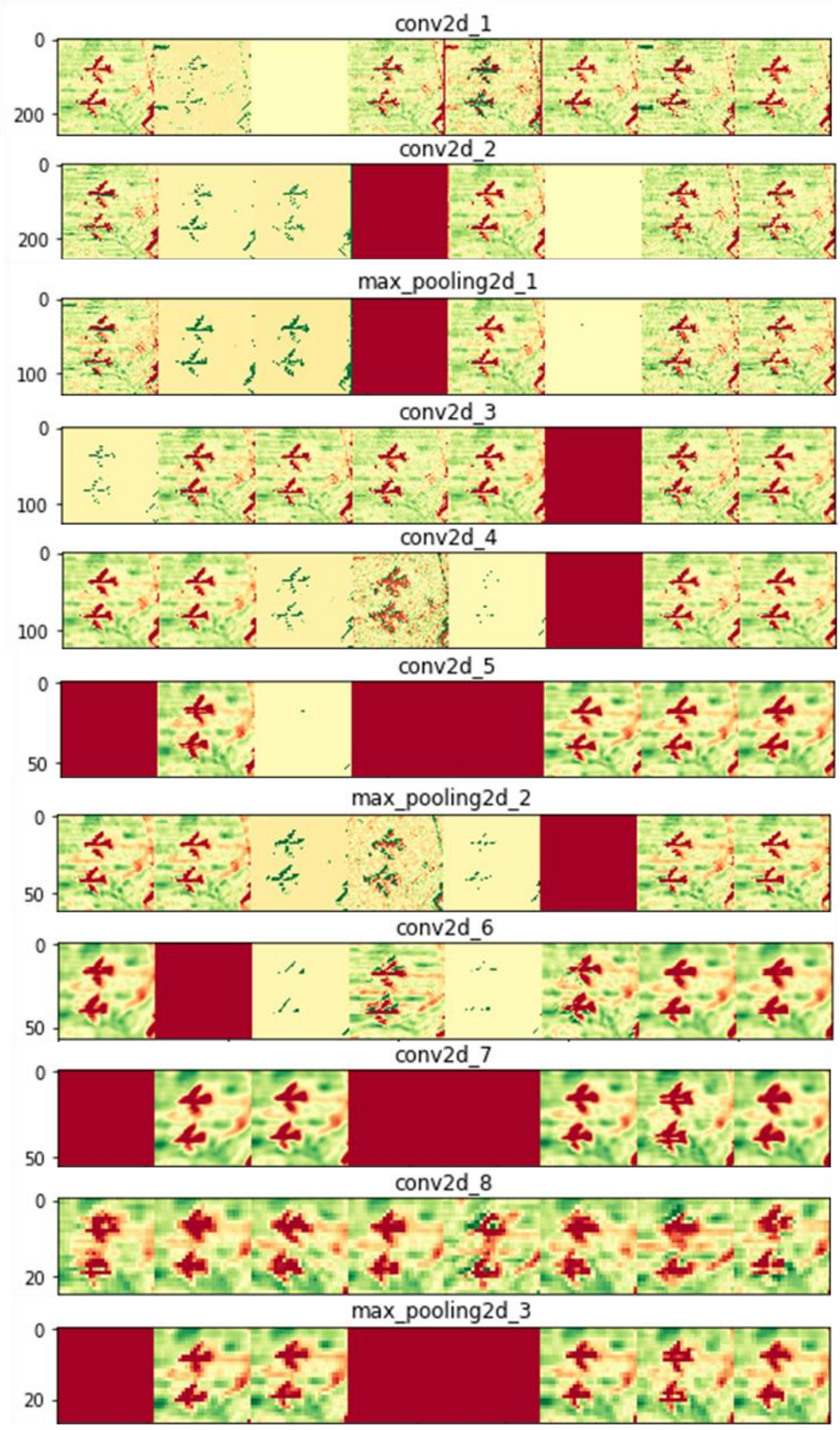

Figure 4. Feature visualization of optimized RSISC-16 Net Model

4.3. Experimental Results of VGG-16 Net Model

The confusion matrix of the VGG-16 Net model is shown in Table 3, where correct responses are represented in the diagonal of matrix, the airplane, lake and river were almost recognized well. From this tennis court was misclassified as commercial area vice versa. The experimental results of VGG-16 Net model with individual results and overall results are shown in Table 4.

Table 3. Confusion Matrix for VGG-16 Net Model in (%)

	AI	BE	CA	DE	FO	LK	OP	RI	TC	WL
AI	94.3	1.4	0.0	0.0	0.0	0.0	1.4	1.4	1.4	0.0
BE	0.0	91.4	0.0	0.0	0.0	2.8	0.0	2.8	1.4	1.4
CA	0.7	0.0	91.4	0.0	0.0	0.0	0.7	0.7	6.4	0.0
DE	0.0	0.0	2.1	96.4	0.0	0.0	0.0	0.0	1.4	0.0
FO	0.0	0.0	2.8	1.4	95.7	0.0	0.0	0.0	0.0	0.0
LK	0.0	1.4	0.0	1.4	0.0	95.7	0.0	1.4	0.0	0.0
OP	0.0	0.0	0.0	0.0	0.0	0.0	92.8	7.1	0.0	0.0
RI	0.0	0.0	0.7	0.7	0.0	0.0	0.7	97.8	0.0	0.0
TC	0.0	0.0	3.5	2.8	2.1	1.4	0.0	0.7	88.5	0.7
WL	1.4	1.4	0.0	0.0	0.0	0.0	0.0	1.4	1.4	94.3

Table 4. Performance Metrics of VGG-16 Net Model

Class	Accuracy	Precision	Recall	F1-Score
AI	94.3	97.77	94.28	96.0
BE	91.4	95.52	91.42	93.43
CA	91.4	90.78	91.42	91.10
DE	96.4	93.75	96.42	95.07
FO	95.7	97.81	95.71	96.75
LK	95.7	95.71	95.71	95.71
OP	92.8	97.01	92.85	94.89
RI	97.8	86.16	97.85	91.63
TC	88.5	88.02	89.28	88.65
WL	94.3	97.77	94.28	96.0
Total	93.8	94.04	93.92	93.98

4.4. Experimental Results of RSISC-16 Net Model

The proposed RSISC-16 Net model is optimized by varying the values of the following three parameters:

- ❖ Activation Function : Tanh, ELU and ReLU
- ❖ Batch Size : 4, 8, 12 and 16
- ❖ Dropout Probabilities : 0.2, 0.3, 0.4 and 0.5

We have obtained 48 different combinations from the above three optimized parameters (4×4×3). In Table 5, ten different possibilities of optimized parameters are given. The model was trained with RMS Properties and learning rate is set to 0.00001 and number of epochs is 10.

Table 5. Result of Optimized RSISC-16 Net Model

S. No.	Batch Size	Activation Function	Dropout	Validation Accuracy
1.	4	ReLU	0.3	95.29
2.	12	ReLU	0.3	93.50
3.	8	ELU	0.4	92.40
4.	12	Tanh	0.2	92.17
5.	4	ReLU	0.4	92.10
6.	8	ReLU	0.5	91.80
7.	4	ReLU	0.5	91.45
8.	4	ELU	0.2	91.10
9.	16	ReLU	0.5	91.10
10.	8	ELU	0.5	87.13

The confusion matrix of the proposed optimized RSISC-16 Net model on NWPU 45-class dataset is shown in Table 6. All the Land Use classes are recognized with good results except wetlands.

Table 6. Confusion Matrix for Optimized RSISC-16 Net Model in (%)

	AI	BE	CA	DE	FO	LK	OP	RI	TC	WL
AI	92.1	0.7	0.7	0.7	0.0	0.0	0.7	4.2	0.0	0.7
BE	0.7	92.8	0.7	0.0	0.7	4.2	0.0	0.0	0.0	0.7
CA	0.0	0.0	97.1	1.4	0.0	0.0	0.0	0.0	1.4	0.0
DE	0.0	0.0	1.4	97.8	0.0	0.0	0.0	0.0	0.7	0.0
FO	0.0	0.0	0.7	0.0	99.2	0.0	0.0	0.0	0.0	0.0
LK	0.0	0.7	0.7	0.7	0.0	97.8	0.0	0.0	0.0	0.0
OP	0.0	0.0	0.0	0.0	0.0	0.0	96.4	3.5	0.0	0.0
RI	0.0	0.0	0.7	0.7	0.0	0.0	0.0	98.5	0.0	0.0
TC	0.7	2.1	0.0	0.0	0.0	0.0	0.0	0.7	96.4	0.0
WL	0.0	0.7	0.0	0.0	10.7	2.8	1.4	2.1	0.0	82.1

Table 7. Performance Metrics of Optimized RSISC-16 Net Model

Class	Accuracy	Precision	Recall	F1-Score
AI	92.14	98.47	92.14	95.2
BE	92.85	95.59	92.86	94.2
CA	97.14	95.1	97.14	96.1
DE	97.85	96.48	97.86	97.16
FO	99.28	89.68	99.29	94.24
LK	97.85	93.2	97.86	95.47
OP	96.42	97.83	96.43	97.12
RI	98.57	90.2	98.57	94.19
TC	96.42	97.83	96.43	97.12
WL	82.1	98.29	82.14	89.49
Total	95.06	95.27	95.07	95.03

Based on the optimized results from RSISC-16 Net model, Table 7 and Figure 5 show the average Precision of 95.29%, Recall of 95.06%, F1-Score of 95.05% and average Accuracy of 95.06% are obtained for the individual remote sensing image classes. The optimized RSISC-16 Net model performance is better when compared to the VGG-16 Net pre-trained model.

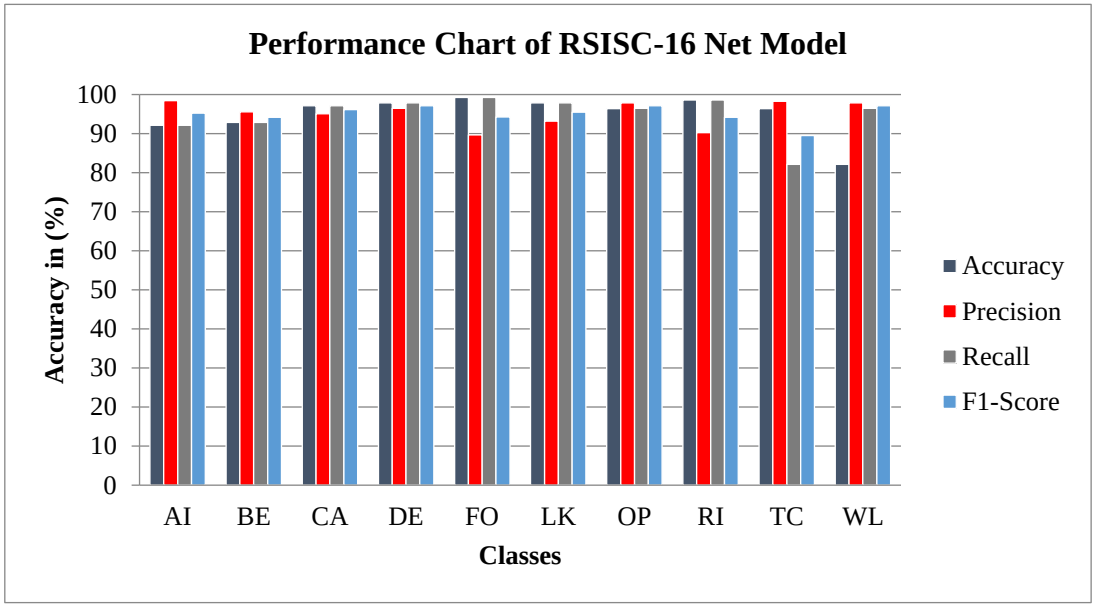

Figure 5. Performance Chart of RSISC-16 Net Model

5. Conclusion

Scene Classification in remote sensing images is a challenging problem because of small interclass variations, large intra-class variations and complex backgrounds. For getting good results, we have introduced an optimized RSISC-16 Net model for classification of remote sensing images. The proposed RSISC-16 Net model is extended by the architecture of VGG-16 Net base model and fine tuning the hyperparameters such as batch size, dropout probabilities and activation function of deep CNN by Talos optimization tool. We have observed that batch size of 4, activation function ReLU with dropout probabilities of 0.3 gives highest accuracy of 95.06%. The experimental results of optimized RSISC-16 Net model performance are superior when compared to the VGG-16 Net pre-trained model.

References

[1] Shafaey, M. A., Salem, M. A. M., Ebied, H. M., Al-Berry, M. N., and Tolba, M. F. (2019). Deep Learning for Satellite Image Classification. Advances in Intelligent Systems and Computing, Vol. 845, pp. 383–391.

[2] P.Deepan, L.R. Sudha, "Remote Sensing Image Scene Classification using Dilated Convolutional Neural Networks", International Journal of Emerging Trends in Engineering Research, Vol. 8, No.7, pp.3622-3630, 2020.

[3] Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition, Proceedings of the International Conference on Learning Representations (ICLR), pp. 1-8.

[4] Cheng, G. and Han, J., (2016). A survey on object detection in optical remote sensing images, ISPRS Journal of Photogrammetry and Remote Sensing, Vol.117, pp.11-28.

[5] Chen, J., Huang, H., Peng, J., Zhu, J., Chen, L., and Li, W., (2020). Convolution Neural Network Architecture Learning for Remote Sensing Scene Classification, arXiv:2001.09614v1, pp. 1-10.

[6] Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., and Zou, H., (2018). Multi-scale object detection in remote sensing imagery with convolutional neural networks, ISPRS Journal of Photogrammetry and Remote Sensing, pp. 1-20.

[7] P Deepan and L.R. Sudha, (2020).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.

[8] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P., (2015). Gradient-based learning applied to document recognition, Proceedings of the IEEE, vol. 86, pp. 2278–2324.
[9] O'Shea, K., and Nash, R., (2015). An Introduction to Convolutional Neural Networks, International Journal of Computer Vision and Pattern Recognition,pp. 2-11.
[10] Liu, X., Zhou, Y., Zhao, J., Yao, R., Liu, B., and Zheng, Y., (2019). Siamese Convolutional Neural Networks for Remote Sensing Scene Classification, IEEE Geoscience and Remote Sensing Letters, pp. 1-5.
[11] Deepan, P., Abinaya, S., Haritha, G., and Iswarya, V., (2018). Road Recognition from Remote Sensing Imagery using Machine Learning, International Research Journal of Engineering and Technology,pp. 3677-3683.
[12] Cheng, G., Yang, C., Yao, X., Guo, L., and Han, J., (2018). When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs, IEEE Transactions on Geoscience and Remote Sensing,pp. 1-11.
[13] Qayyum, A., Malik, A., Saad, N.M., Iqbal, M., Abdullah, M.F., Rasheed, W., Abdullah, T., and Jafaar, M., Scene classification for aerial images based on CNN using sparse coding technique, International Journal of Remote Sensing,pp. 1-24.
[14] Zhang, W., Tang, P., and Zhao, L., Remote Sensing Image Scene Classification Using CNN-CapsNet, Remote Sensing, pp. 1-22, 2019.
[15] Chaib, S., Liu, H., Gu, Y., and Yao, H., Deep Feature Fusion for VHR Remote Sensing Scene Classification, IEEE Transactions on Geoscience and Remote Sensing, Vol.2(10) : 1-10, 2017.
[16] Deepan, P., and Sudha, L.R., Fusion of Deep Learning Models for Improving Classification Accuracy of Remote Sensing Images, Journal Of Mechanics of Continua and Mathematical Sciences, Vol. 3(12): 189-201, 2019.
[17] Dong, Y., and Zhang, Q., A Combined Deep Learning Model for the Scene Classification of High-Resolution Remote Sensing Image, IEEE Geoscience and Remote Sensing Letters:1-5, 2019.
[18] Scott, G.J., Marcum, R.A., Davis, C.H., and Nivin, T.W., Fusion of Deep Convolutional Neural Networks for Land Cover Classification of High-Resolution Imagery, IEEE Geoscience and Remote Sensing Letters, vol.2(5): 1-5, 2017.
[19] Trivas,W., and Li, R., An Ensemble of Convolutional Neural Networks Using Wavelets for Image Classification, Journal of Software Engineering and Applications, pp.69-88, 2018.
[20] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research, Vol. 15, pp. 1929-1958.
[21] Bargshady, Ghazal, Zhou, X, Deo, Ravinesh C, Soar, Jeffrey, Whittaker, Frank and Wang, Hua, The modeling of human facial pain intensity based on Temporal Convolutional Networks trained with video frames in HSV color space., Applied Soft Computing, 97, 2020.
[22] Xiao, Luwei, Xue, Yun, Wang, Hua, Hu, Xiaohui, Gu, Donghong and Zhu, Yongsheng, Exploring Fine-grained Syntactic Information for Aspect-based Sentiment Classification with Dual Graph Neural Networks, Neurocomputing. Vol. 1(3), pp. 1-14, 2021.