

# Particle Swarm Optimization Based Deep Learning Model for Scene Classification of Remote Sensing Images

Jagroop Singh<sup>1\*</sup>, Gurjeet Singh<sup>2</sup>

<sup>1,2</sup>Department of Electronics and Communication Engineering, Amritsar College of Engineering and Technology, Punjab, India

**Abstract:** Classification of remote sensing images is an open area of research. Recently deep learning models are extensively utilized to classify the images. To overcome the overfitting issue, an ensemble deep learning model is proposed. The hyper-parameter of the proposed model is tuned using particle swarm optimization. Initially, the features of remote sensing images are extracted. Theater, feature selection approach is used. The extracted features are then used to build the model by using the particle swarm optimization-based ensemble deep learning model. Extensive experiments are performed using the proposed model. The comparative analysis show that the proposed model outperforms the existing model.

**Keywords:** Deep learning, Ensemble model, Hyper-parameters, Image classification.

## 1. Introduction

Remote Sensing (RS) image processing has had a very long tradition since the 1840s. It began with the invention of the camera more than 150 years ago. As noted by Avery and Berlin [35] as well as Baumann [36], “the RS imagery can be defined as any process whereby the Information is gathered by the reflectance of light energy from an external source such as the sun without being in contact with it by any device.” It is like how our human eyes work. Remote sensing has become associated more specifically with the Earth's surface monitoring electromagnetic Spectrum of satellite in nowadays [37], and the electromagnetic spectrum is shown in Fig. 1.

Through the fast development and wide utilization of RS satellites, RS image has become a major tool for data acquisition on the entire Earth surface. The revisit time ranges from a couple of days to a matter of hours. Many GIS applications integrate RS images for various analyses, particularly for those involved in natural resources.

The RS sensors can be divided into two types - passive and active. The passive sensors do not supply energy to the objects being detected, and they are mostly used for measuring and recording

The reflection of light off Earth objects' features. Aerial photography is a major form of remote sensing by passive sensors, and it can collect from visible to near-infrared

wavelengths, or even longer wavelengths from the solar radiation. In contrast, the active sensors supply their own source of energy, flash photography and Radio Detection and Ranging (RADAR) systems are examples.

The RADAR system emits energy with wavelengths in the microwave section of the electromagnetic spectrum, as shown in Fig. 1, the reflection of this energy from the earth's surface produces RADAR images.

### A. Overview of Remote Sensor Application

All kinds of satellites are available in the universe. Some satellites orbit the Sun, the moon, Jupiter, mars, earth, etc. A few satellites provide valuable information to Earth. Different kinds of radio waves are captured by the communication satellites (CoS), which pass on the information to the entire world in a fraction of a second; to capture the natural images tuned into maps, resource satellite (ReS) performed well and giving valuable information to the scientists regarding street view, distance between the source and destinations reports; Navigation Satellites (NaS) help pilots and sailors to transfer and receive signals from the aircrafts and ships for any further urgent resource; for spying, communicating, navigating to any other countries, Military Satellites (MiS) capture image in the radio waves and send them to the forces; to know about the space and planets, scientific satellites capture important data in the image formats to find asteroids, black holes, new planets, etc. To find any weather issues on the earth, Weather Satellites (WeS) send the data in the image format to source tsunamis, earth quakes, storms, etc. All these satellites are artificial satellites, which have developed by humans.

For information related to the earth, artificial satellites and Synthetic Aperture Radar (SAR) provide important data for remote sensing image processing applications. Current sensors display clear resolution images which may have precise shape, good quality in edges, and contours. Therefore, remote sensing images contain both high and low frequency bands and are highly reliable. The PAN and MS images provide useful data in remote sensing applications. PAN images can be attained with high spatial resolution, which relies on various kinds of

\*Corresponding author: [jagroopdhillon602@gmail.com](mailto:jagroopdhillon602@gmail.com)

satellites and composed in the comprehensive range of visual wavelength, concentrated in a combination of black and white colors. For better spectral resolution, the MS image attained has less spatial resolution and relies on the kind of satellites.



Fig. 1. Remote sensing applications

### B. Optical remote sensing

Optical remote sensing images can make visible, short-wave infrared and near-infrared bands of the earth's surface. Multiple materials reflect as well as absorb various multiple wavelengths. Wavelength is in the range from 0.3 to 15. Optical remote sensing images consist of four different types; those types are based on the utilization of various spectral bands.

1. PAN optical image: A PAN image sensor is obtained by a singular band detector. This detector is highly sensitive to infrared (IR) wavelengths and radiation in wide spectral and visible ranges. Once the range of the wavelength concerns with the range of the visible, the image will be black and white.
2. Multi-spectral optical image: MS remote sensing image system consists of multiband sensors. MS images obtain a narrow band wavelength range compared to the PAN image system. MS image consists of both the spectral and brightness of the targeted area. However, MS images are low resolution images, which are used for visible and IR wavelengths.
3. Super and Hyper spectral optical images: A Super Spectral (SS) remote sensing image can be acquired by the above 10-spectral bands. The Hyper Spectral (HS) remote sensing image can be acquired by a hundred spectral bands.

### C. Synthetic Aperture Radar (SAR) images

Synthetic Aperture Radar (SAR) is a form of Radio Detection and Ranging (Radar) used for obtaining high spatial resolution images of objects, such as surface of the earth. SAR is used for moving radar antennas over the targeted location for high resolution. From the moving radar antennas, the Microwave (MW) pulses are transmitted to the landscape of the earth. Then SAR system receives the MW energy and forms an image. SAR systems are able to operate in bad conditions, such as stormy weather, dull light, etc. However, a few levels of non-Gaussian noise face trouble in SAR systems. SAR images can help to differentiate the number of operating frequencies to get the same image. SAR systems introduce innovative technologies and new thoughts. SAR is the most important for remote sensing applications.

### D. Light Detection and Ranging (LIDAR) images

Lidar image is the reconciliation between the optical field and radar fields. The Lidar improves because of the exploitation of the spatial properties, spectral properties, and temporal properties of lasers. Lidar is a laser light sensor and used to calculate the distance between the aircraft and surface of the earth. For more accurate height computation is needed for large areas, lidar is costly.

### E. Spatial Classification

In Instantaneous Field of View (IFOV) sensing application, spatial classification is the ground region image. Spatial classification is also called spatial or ground resolution. The ground resolution can be illustrated as the spatial region which shapes one pixel in the satellite image. Spatial resolution has three different resolution systems. Low resolution, Medium resolution, and High resolution. Low resolution is approximated to more than 1km, medium resolution is approximated between 100m to 1km, and high resolution is approximated to less than 5m.

### F. Scene classification of the remote sensing images based on neural network

With different encoding methods, we can use the DCNN for world classification of the remote detecting images and are no longer constrained by the input picture scales. Ergo, we are able to input multiple range types of the exact same remote detecting picture for instruction and use the different range features of the convolution and fully attached layers in the DCNN, which may have powerful abstract representation ability. Ultimately, we use multichannel help vector machine for world classification.

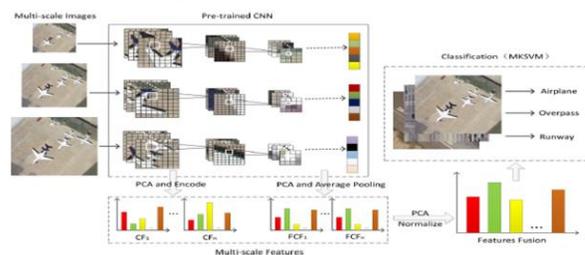


Fig. 2. The structure of scene classification algorithm

## 2. Literature Review

Y. Wang *et al.* (2017) [9] proposed high-resolution polarimetric synthetic aperture radar (SAR) images, texture variation increasingly dominates the scene of those selected targets, creating a major challenge in speckle modeling and target classification. When a texture is incorporated into speckle models, it is often assumed that pixels are independent from each other. In this paper, we allow dependent neighboring pixels and evaluate the texture as a Markov random field (MRF). T. L. Ainsworth *et al.* (2019) [10] exhibited Model-based polarimetric decompositions are often used to make scene classifications from polarimetric SAR imagery. The first Freeman-Durden design has been modified and improved upon multiple times over the past 2 decades. However, quantitative in-depth analyses of those incoherent model-based decompositions have lagged in comparison. T. L. Ainsworth *et*

al. (2018) [11] shown Model-based polarimetric decompositions that have long been applied to identify polarimetric SAR imagery. That extended and successful record has, at times, been interrupted by nagging issues: polarimetric orientation, position results, bad spreading energy, simplistic spreading models, etc. We check model-based polarimetric decomposition techniques for robustness to modifications of the underlying in-scene spreading mechanisms. Xu et al. (2015) [19] have attempted to overcome the color distortion issue by dividing the PAN and MS image pixels into k-means clustering classes, after that multiple regression to compute weights on every cluster of pixels. The summary of low-resolution PAN images is needed because of satellite sensors nonlinear spectral processing; however, few methodologies are not approached well in color distortion. Abdullah et al. (2015) [20] have developed methodologies for the PAN and MS images with Multivariate Empirical Mode Decomposition (MEMD) method. It is the advance of Empirical Mode Decomposition (EMD) method and can avoid mixing mode non-uniqueness and misalignment problems in mode and allow the multivariate decomposition information into its basic oscillatory scale. Moreira et al. (2013) [21] proposed that the most convenient mouth procedure gives a 2D reflectivity guide to the imaged scene. Bright spots on the image indicate the targets having high back scattering signal and dark area indicate smooth surface. Visualizing live SAR facts is not going to provide any handy information on the scene. Hence, this raw data is preprocessed and represents the scene reflectivity. Peng et al. (2014) [22] proposed that the MAP primarily based a spatial filter intended for SAR images by employing Cauchy-Rayleigh concoction type intended for disturbance in addition to rectangular underlying Gamma syndication for that disturbance free signal. A variables are projected working with Expectation-Maximization (EM) criteria whoever the correctness can be in line with the chosen initial beliefs with an iterative start. Lu et al. (2004) [23] have shown that the detecting changes that occurred in remote places of the earth is a very challenging task. Critical changes come under short-term changes where computing time is a main constraint. All-natural disasters like floods, earthquake, and forest fire come under short-term change. Decision and planning are essential tasks after detection. Celik et al. (2009) [24] proposed unsupervised CD criteria employing Primary Element Evaluation (PCA) along with K-means clustering. Even measurements involving low overlap golf prevents are produced in an image along with eigen vectors are produced for anyone block. Then an attribute vectors are produced by way of predicting the eigen vector within a low overlap golf block. K-indicated the criteria must be used so that you can discriminate your attribute vectors in to evolve along with unrevised groups. However, strong answers are obtained throughout a lesser amount of computing time, the outcomes are not improved upon additional by way of tuning. Bazi et al. (2005) [25] presented provided that the not being watched CD solution employs improved Kittler-Illingworth (KI) tolerance underneath the assumption involving Many times Gaussian model. It has the next methods: preprocessing about sound

removal, wood relation, big difference, image creation along about sharp evolves along with unrevised classes. This tolerance structured method does not make the most of the spatial contextual information and facts about the appropriate approximation involving evolved along with unrevised statistics.

### 3. Proposed Algorithm

This section discusses the proposed model. Initially, the potential features of the used dataset are obtained for classification process. The Harris corner detector then utilized.

Particle swarm optimization (PSO) is then used to select the hyper-parameters of the proposed model. The pseudocode of the proposed model is shown in Algorithm 1. Based upon the return parameters by PSO, deep ensemble model is trained for classification purpose. Fig. 3 shows the proposed deep ensemble model. Various set of extracted features are initially assigned to various deep learning models. The results are then aggregated to obtain the final results.

```

Algorithm 1: Particle Swarm Optimization
1 Max_Itr ← Maximum number of iterations
2 n ← Swarm size
3 for i = (1 to n) do
4    $x_i \leftarrow$  initialize the position of the ith particle
5    $L_{best}^i = x_i$  ▷ set the initial local best of the ith particle
6   if  $f(L_{best}^i) > f(G_{best})$  then
7      $G_{best} = L_{best}^i$ 
8    $v_i \leftarrow$  initialize the velocity of the ith particle
9 while (t < Max_Itr) do
10  t = t + 1
11  for i = (1 to n) do
12     $v_i \leftarrow$  Update the velocity of the ith particle
13     $x_i(t) = x_i(t-1) + v_i$  ▷ Update the position of the ith particle
14    if  $f(x_i) > f(L_{best}^i)$  then
15       $L_{best}^i = x_i$ 
16      if  $f(L_{best}^i) > f(G_{best})$  then
17         $G_{best} = L_{best}^i$ 
Output:  $G_{best}$ 

```

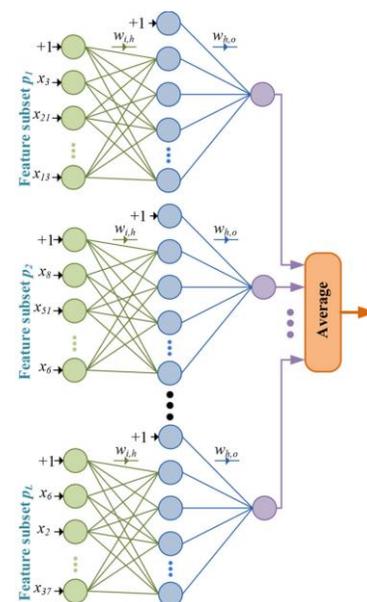


Fig. 3. Proposed ensemble model

### 4. Performance Metrics

This section discusses the various performance metrics used in this research work to evaluate the performance of the existing and proposed image captioning models. The proposed algorithm is tested on various stages. The algorithm is applied using various performance indices like Accuracy, F-measure, Sensitivity, Specificity, Kappa statistics Testing time.

#### A. Specificity analysis

Specificity calculates the degree between true negative rate, and false positive rate. It defines weather, our system can identify nonauthentic captions in an efficient manner or not. It should be maximized. Specificity ( $S_p$ ) can be mathematically defined as:

$$S_p = \frac{T_n}{T_n + F_p}$$

Table 1  
Specificity analysis of the proposed model

Images	Existing	Proposed
Im1	0.9745	0.9835
Im2	0.9802	0.9842
Im3	0.9708	0.9728
Im4	0.9747	0.9822
Im5	0.983	0.9899
Im6	0.9825	0.9855
Im7	0.9762	0.9852
Im8	0.9714	0.9804
Im9	0.9792	0.9872
Im10	0.9826	0.9861
Im11	0.9832	0.9882
Im12	0.9811	0.9861
Im13	0.9743	0.9763
Im14	0.9715	0.9765
Im15	0.9821	0.9881

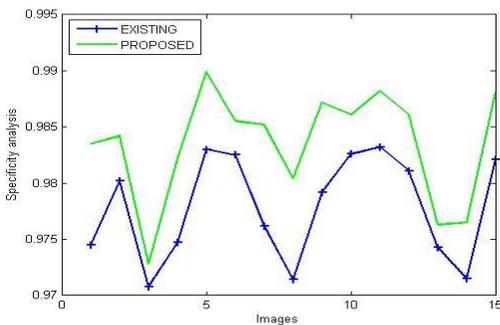


Fig. 4. Specificity analysis of the proposed model

Table 1 and Fig. 4 show the specificity analysis of the proposed model. It clearly shows that the proposed model outperforms the existing model by 1.4293%.

#### B. Kappa statistic

Cohen’s kappa statistic ( $k$ ) is a performance metric which computes the inter-rater precision (i.e., reliability) between True ( $T$ ) and Negative ( $N$ ) classes. Inter-rater precision is considered if your data raters provide a similar scores to the same data item such as captions in this research work.  $k$  is a more efficient performance metric compared to the simple percent agreement computation, as  $k$  considers possibility of

200 agreements occurring by chance.  $k \in 2 [0; 1]$  and should be maximum.

Table 2  
Kappa statistic

Images	Existing	Proposed
Im1	0.9844	0.9859
Im2	0.9807	0.9887
Im3	0.9755	0.9778
Im4	0.9705	0.9767
Im5	0.9804	0.9889
Im6	0.9812	0.9902
Im7	0.9752	0.9772
Im8	0.9762	0.9782
Im9	0.9731	0.9751
Im10	0.9721	0.9756
Im11	0.9817	0.9867
Im12	0.9721	0.9761
Im13	0.9753	0.9798
Im14	0.9792	0.9882
Im15	0.9796	0.9836

Table 2 and Fig. 5 show the specificity analysis of the proposed model. It clearly shows that the proposed model outperforms the existing model by 1.4293%.

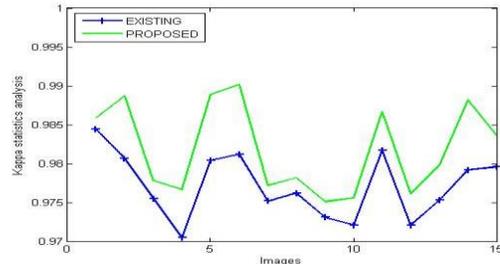


Fig. 5. Kappa statistics analysis of the proposed model

Table 3  
Testing time analysis of the proposed model

Images	Existing	Proposed
Im1	1.112	0.938
Im2	2.186	2.038
Im3	2.268	2.03
Im4	1.104	0.899
Im5	1.9	1.73
Im6	1.818	1.629
Im7	1.871	1.661
Im8	1.495	1.325
Im9	2.46	2.227
Im10	2.12	1.963
Im11	1.042	0.876
Im12	1.441	1.238
Im13	2.389	2.187
Im14	1.176	0.937
Im15	1.865	1.718

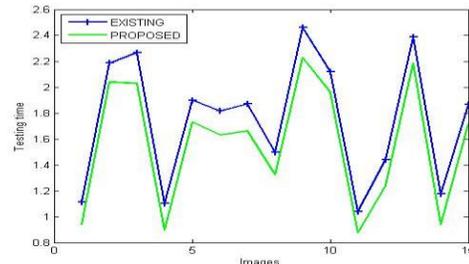


Fig. 6. Testing time analysis of the proposed model

Table 3 and Fig. 6 show the testing time in seconds analysis of the proposed model. It clearly shows that the proposed model takes lesser time than the existing model by 1.3822%.

### 5. Conclusion

An efficient ensemble deep learning based ensemble classification model is proposed for remote sensing images. Initially, we have loaded the remote sensing image dataset. The feature of the collected dataset has been then obtained. Thereafter, the dataset has been divided into training and testing fractions. Potential features of the dataset for training purpose have been selected. Thereafter, random particles were initialized to optimize the SVM model. Compute velocity and position values of PSO. Thereafter, fitness of particles has been computed by using the initial parameters of SVM. Global best (gbest) and particle (pbest) have been utilized to select the solutions.

Thereafter, recompute the velocity and fitness values of selected solutions. The objective was to select gbest solution. Depending upon termination criteria final trained model was achieved. Comparative analysis indicate that the proposed approach outperforms the competition models.

In near future, we will use other metaheuristic techniques to enhance the work. Also, deep transfer learning models can be used to extend the proposed work.

### References

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.
- [2] Huang, Xin, Xuehua Guan, Jón Atli Benediktsson, Liangpei Zhang, Jun Li, Antonio Plaza, and Mauro Dalla Mura. "Multiple morphological profiles from multicomponent-base images for hyperspectral image classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, no. 12 (2014): 4653-4669.
- [3] X. Liu, S. Jia, M. Xu and J. Zhu, "Statistical Fusion-Based Transfer Learning for Hyperspectral Image Classification," *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019, pp. 1108-1111.
- [4] Y. Wang, T. L. Ainsworth and J. Lee, "MRF modeling of textured speckle for polarimetric SAR image classification," *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, 2017, pp. 5287-5290.
- [5] T. L. Ainsworth, J. Lee and Y. Wang, "Assessment of Model-Based Polarsar Decompositions," *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019, pp. 5019-5022.
- [6] T. L. Ainsworth, J. Lee and Y. Wang, "Model-Based Polarsar Decompositions: Virtues and Vices," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, 2018, pp. 8120-8123.
- [7] Yang, Zhou, Xiao-dong Mu, and Feng-an Zhao. "Scene classification of remote sensing image based on deep network and multi-scale features fusion." *Optik* 171 (2018): 287-293.
- [8] John E Ball, Derek T Anderson, and Chee Seng Chan. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing*, 11(4):042609, 2017.
- [9] SudipanSaha, Francesca Bovolo, and Lorenzo Bruzzone. Change detection in image time-series using unsupervised lstm. *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [10] Peng, Q. & Zhao, L. (2014), 'SAR image filtering based on the Cauchy Rayleigh mixture model', *Geo science and Remote Sensing Letters*, IEEE, 11(5):960–964.
- [11] Lu, D., Mausel, P., Brondzio, E., & Moran, E. (2004), 'Change detection techniques', *International Journal of Remote Sensing*, 25(12):2365–2401.
- [12] Celik, T. (2009), 'Multiscale Change Detection in Multitemporal Satellite Images', *IEEE Geoscience and Remote Sensing Letters*, 6(4):820–824.
- [13] Bazi, Y., Melgani, F., & Al-Sharari, H. D. (2010), 'Unsupervised Change Detection in Multispectral Remotely Sensed Imagery with Level Set Methods', *IEEE Transactions on Geoscience and Remote Sensing*, 48(8):3178–3187.
- [14] Bhuiyan, M. I. H., Ahmad, M. O., & Swamy, M. N. S. (2007), 'Spatially Adaptive Wavelet-Based Method Using the Cauchy Prior for Denoising the SAR Images', *IEEE Transactions on Circuits and Systems for Video Technology*, 17(4):500–507.
- [15] Ranjani, J. J. & Thiruvengadam, S. J. (2011), 'Generalized SAR Despeckling Based on DTCWT Exploiting Interscale and Intracscale Dependences', *IEEE Geoscience and Remote Sensing Letters*, 8(3):552–556.
- [16] Argenti, F., Bianchi, T., Lapini, A., & Alparone, L. (2012), 'Fast MAP Despeckling Based on Laplacian-Gaussian Modeling of Wavelet Coefficients', *IEEE Geoscience and Remote Sensing Letters*, 9(1):13–17.
- [17] Gu, Wei, ZhihanLv, and Ming Hao. "Change detection method for remote sensing images based on an improved Markov random field." *Multimedia Tools and Applications* 76.17 (2017): 17719-17734.
- [18] Zhu, Zhe. "Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications." *ISPRS Journal of Photogrammetry and Remote Sensing* 130 (2017): 370-384.
- [19] Jay Nandy, Wynne Hsu, and Mong Li Lee. Towards maximizing the representation gap between in-domain & out-of-distribution examples, 2020.
- [20] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks, 2018.
- [21] Xiao Xiang Zhu, Jingliang Hu, Chunping Qiu, Yilei Shi, Jian Kang, LichaoMou, Hossein Bagheri, Matthias H'aberle, Yuansheng Hua, Rong Huang, et al. So2sat LCZ42: A benchmark dataset for global local climate zones classification, 2019.
- [22] G. Chen, W. Choi, X. Yu, T. X. Han, and M. Chandraker, "Learning efficient object detection models with knowledge distillation," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, December 4-9, 2017, Long Beach, CA, USA,
- [23] I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., 2017, pp. 742–751. <https://proceedings.neurips.cc/paper/2017/hash/e1e32e235eeef1f970470a3a6658dfdd5-Abstract.html>
- [24] Q. Lu, X. Huang, J. Li and L. Zhang, "A Novel MRF-Based Multifeature Fusion for Classification of Remote Sensing Images," in *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 4, pp. 515-519, April 2016.
- [25] S. Jia and J. Xian, "Multi-Feature-Based Decision Fusion Framework for Hyperspectral Imagery Classification," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, 2018, pp. 5-8.
- [26] O. Sen and H. YalimKeles, "Scene Recognition with Deep Learning Methods Using Aerial Images," *2019 27th Signal Processing and Communications Applications Conference (SIU)*, Sivas, Turkey, 2019, pp. 1-4.
- [27] X. Xu, J. Li, X. Huang, M. Dalla Mura and A. Plaza, "Multiple Morphological Component Analysis Based Decomposition for Remote Sensing Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 3083-3102, May 2016.
- [28] X. Xu, J. Li and M. Dalla Mura, "Remote sensing image classification based on multiple morphological component analysis," *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Milan, 2015, pp. 2596-2599.
- [29] B. Deng, S. Jia and D. Shi, "Deep Metric Learning-Based Feature Embedding for Hyperspectral Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1422-1435, Feb. 2020.