Section A-Research paper

Classification of Images Using Hybrid Convolutional Neural Networks

[1]K.Ramalakshmi,
[1]asprkrama@gmail.com, P.S.R. Engineering College, Sivakasi-626140, Tamilnadu, India
[2]L.Krishnakumari
Ramco Institute of Technology, Rajapalayam – 626117, Tamilnadu, India
[3]P.A.Mathina
P.S.R. Engineering College, Sivakasi-626140, Tamilnadu, India
[4]G.Theivanathan
Velammal Engineering College, Chennai – 600 066, Tamilnadu, India
[5]S.Samboornalaxmi
P.S.R. Engineering College, Sivakasi-626140, Tamilnadu, India

ABSTRACT

The classification of the land-cover like forest, urban, water/ice, farm- land i.e. crop, oil-slick is important for controlling deterioration of the environment and destruction of wetland, for urban region planning, natural resources monitoring and for collecting information on possible future disasters. Land-cover classification can be implemented using images acquired from various types of sensors. Synthetic Aperture Radar (SAR) is a radar that collects information from the earth surface and generates high resolution images of wide areas. The convolutional neural network, is a specialized type of neural network model designed for working with two-dimensional image data, although they can be used with one-dimensional and three-dimensional data. The proposed method adopts the idea of deep neural networks and presents a hybrid ResNet and VGG with Support Vector Machine (SVM) based convolutional neural network to classify the land cover from high-resolution remote sensing imagery. Pretrained versions of ResNet, VGG, and AlexNet were used for the classification in order to compare and assess the suggested method's superiority. Greater accuracy of 96.3 %is offered by the hybrid ResNet-VGG withSVM model than by the pretrained ResNet and VGG. Compared to the existing models, the proposed model shows better performance in terms of accuracy, precision, recall and dice-index.

Keywords:
Land-cover classification,High resolution images,Convolutional neural network, Synthetic Aperture Radar, Deep neural networks Support Vector Machine, Greater accuracy.

1. INTRODUCTION

Hyperspectral imagery (HSI) differs from its earlier counterpart, multispectralimagery, in two key ways. The first difference is the number of spectral bands collectedby hyperspectral sensors. Multispectral sensors typically collect less than ten bands ofspectral information per pixel. Hyperspectral imagery contains hundreds of bands ofspectral information per pixel. The second difference is that multispectral imagery having so few bands, selects wavelengths that are considered the most informative for a particular application; thus,thebandsarenon-contiguous. Hyperspectral image analysis has developed over the last ten years into one of the most potent and rapidly expanding technologies in the world of remote sensing. Synthetic Aperture Radar (SAR) can obtain rich information of earth surface under all-time and all-weather conditions. Land cover classification or segmentation using SAR images could be of great importance in a variety of Earth observation applications, such as urban planning, natural disasters monitoring, and environmental resource management (Li et al,2012, Rossi et al, 2015 &Cantorna et al, 2019). Compared to other types of remotely sensed data, hyperspectral imagery offers the possibility for more precise and detailed information extraction. The classification of synthetic aperture radar (SAR) pictures is a current topic in the analysis of SAR data.Shi et al(2021) developed a novel architecture of a semiautomated process to label high-resolution single-pol SAR images. Besides, a well pixel-labeled object-level semantic segmentation dataset FUSAR-Map is constructed. Overall accuracy of 92.76% is achieved with this method.

Deep Convolutional Neural Networks are the highly developed category of deep learning approaches. It is an effective and unbeaten technique, especially for learning visual illustrations. Yu et al. (2017) illustrated a CNN model for Hyper Spectral Image classification comprising of each three convolutional and normalization layers, each of which are followed by a dropout and an average pooling layers. Hu et al. (2015) developed a CNN model to categorize Hyper Spectral Images which also comprised of the convolutional, max pooling and fully connected layers. Li et al. (2017) pooled a CNN with an intense learning machine to carry out HSI categorization. Zhang et al. (2016) established a deep learning structure based on a traditional Convolutional Neural Network for Hyper Spectral Images categorization. This structure comprises of three mechanisms: dimensionality reduction using principal component analysis, feature extraction using deep Convolutional Neural Networks and classification using logistic regression classifier. Numerous features extracted using the CNN and the PCA are combined; and then, a classifier is used to classify the features. In addition, Chen et al. (2016) illustrated the deep learning based models for HSI categorization using Convolutional Neural Networks. Numerous CNN structures are used to categorize Hyper Spectral Images. Numerous

convolutional layers and pooling layers are illustrated to extract the time and frequency domain features, that attained precise categorization of HSIs.

This paper is structured with section 2, which exemplify the related works and section 3, illustrates the elaborate functioning of the proposed method. Section 4 depicts the comparison of results obtained by a variety of the existing and the proposed techniques and section 5 discusses conclusion and future directions.

2. LITERATUREREVIEW

Amitrano, D et al (2019), presented a methodology to extract a higher level semantics from synthetic aperture radar images.It is based on fusing contextual data, object-based image analysis, and pixel-based clustering. Datasets using multitemporal SAR over a Colombian forest have yielded preliminary findings. They demonstrated that the synergic exploitation of pixel and object information can provide higher quality land cover results and more information to map users. Li, X et al (2020) proposed a novel multimodal bilinear fusion network (MBFNet), which is used to fuse the optical and SAR features for land cover classification. Following the SACSM's integration into each stream, the global average pooling and global maximum pooling data are bilinearly integrated to provide fine channel attention maps with second-order statistics. The SACSM uses the channel selection process to reconfigure compact feature maps with improved discrimination in addition to automatically highlighting the critical channels of feature maps to increase the representational power of networks. The feature-level fusion method, known as bilinear pooling, creates a second order link between two compact optical and SAR feature maps. Continuous techniques based on normalised compression distance (NCD) estimates were presented by Coca et al. in 2019. With no biases introduced by feature estimators or classifiers, NCD is a similarity metric that is applied straight to the data. Using two TerraSAR-X photos taken over the Japanese city of Sendai and its environs, the experimental findings for the two techniques are calculated. Wijaya, A et al (2009) demonstrated the utility of reduced resolution ALOS PALSAR data for biomass mapping and land cover classification over the tropical forests of Indonesia. The image classifications were carried out using a Machine Learning based classifier, socalled Support Vector Machine (SVM), and a conventional Maximum Likelihood method. Scaled conjugate gradient algorithm and the Kalman filter were used in an ensemble of neural network approach. Both confusion matrices and Kappa statistics were used to evaluate the classification accuracy. It shows that the introduction of SAR textures significantly enhanced the classification accuracies. This study showed that the joint processing of SAR and multispectral data increased the accuracies of biomass estimation and land use classifications. Three datasets

(viz. RISAT-1 hybrid-pol data, RISAT- 1 dualpol data, and Resourcesat-2 data) were evaluated with SVM classifier and compared using three different kernel parameters, such as radial basis function (RBF), Polynomial with degree -2 and Linear. Here, it was observed that the SVM with RBF kernel parameter gave highest Overall Accuracy (OA) of 92.34% for hybrid Pol RISAT-1 data. The SVM with RBF kernel provides an overall accuracy of 76.83% for dual-pol RISAT-1 dataset. SVM has classified the datasets into four classes like Urban, Water, Vegetation, and Bare soil. The evaluation was carried out using confusion matrix for accuracy assessment. Xia, J. et al (2021) provides a representative benchmark of high resolution SAR datasets for building segmentation and reviews the current state-of-the-art segmentation methods. To investigate the segmentation performance of SAR, a comprehensive assessment with different models, encoders, pretraining weights, ensemble schemes are performed. Based on the evaluation, we give some suggestions to improve the segmentation results of SAR datasets. Particular attention should be given to U-Net with the encoder of Efficient Net- B5 with potential improvements of applying ensemble, multichannel mask, and post processing. Unnikrishnanet. al (2018) proposed an AlexNet architecture with less amount of filters to classify the hyperspectral images. This method is experimented with SAT 4 dataset and produced successful accuracy performance compared to the existing methods. Dasari, K. et al. (2018) employed complementary datasets collected over two study regions in the southern Beaufort Sea, Northwest Territories, Canada, to evaluate the capability of dual pol Terra SAR-X and quad pol Radarsat-2 imaging for land cover mapping. The extraction of Kennaugh matrix elements from novel polarimetric SAR processing methods has been used to produce inputs for the Maximum Likelihood Classifier. Results show good potential for general land cover mapping, and that overall accuracy increases when backscatter coefficients from both sensors are utilized (81.4% for eight classes), compared to TerraSAR-X HH/VV (68.9%) or Radarsat-2 HH/HV/VV (73.6%) alone. Kennaugh matrix elements produced similar accuracy gains, however when backscatter coefficients and Kennaugh matrix elements were combined, modest declines were seen. Dong, X et al (2015) proposed a triplet attention feature fusion network (TAFFN) for SAR and optical image fusion classification. To extract spatial and spectral long-range information from the SAR image and optical image, respectively, modules for spatial attention and spectral attention based on self-attention mechanisms are designed. Cross-attention mechanisms are also suggested in order to capture the long-range interaction representation. To further incorporate the complimentary information of SAR and optical pictures, triple attentions are concatenated. Studies using a SAR and optical multimodal dataset show that the suggested method is capable of delivering state-of-the-art results.

3. METHODOLOGY

ConvolutionalNeuralNetworks(CNN)

Oneofthemostpopulartypesofdeepneuralnetworksisknownasconvolutionalneuralnetworks (CNN or ConvNet). A CNN convolves learned features with input data, anduses2Dconvolutionallayers,makingthisarchitecturewellsuitedtoprocessing2D data,suchasimages. CNNs eliminate the need for manual feature extraction, so it is not necessary toidentifyfeaturesusedtoclassifyimages.TheCNNworksbyextractingfeaturesdirectlyfromimages. Therelevantfeaturesarenotpretrained;theyare learnedwhilethenetworktrains on a collection of images. This automated feature extraction makes deep learningmodels highlyaccurateforcomputervisiontasks suchas objectclassification.The main advantage of CNN comparedtoitspredecessorsisthatitautomaticallydetectstheimportantfeatureswithoutany humansupervision.

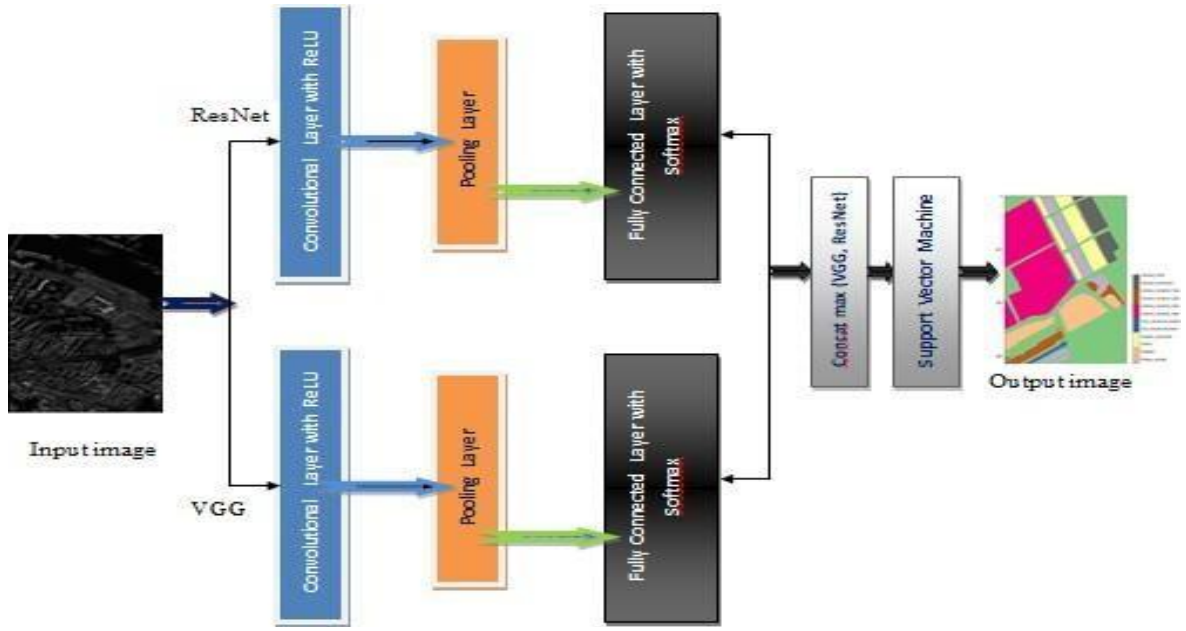

Figure.1.Block diagram of proposed method

CNN is also computationally efficient. It uses special convolution and poolingoperations and performs parameter sharing. Block diagram of proposed method is shown in figure 1.

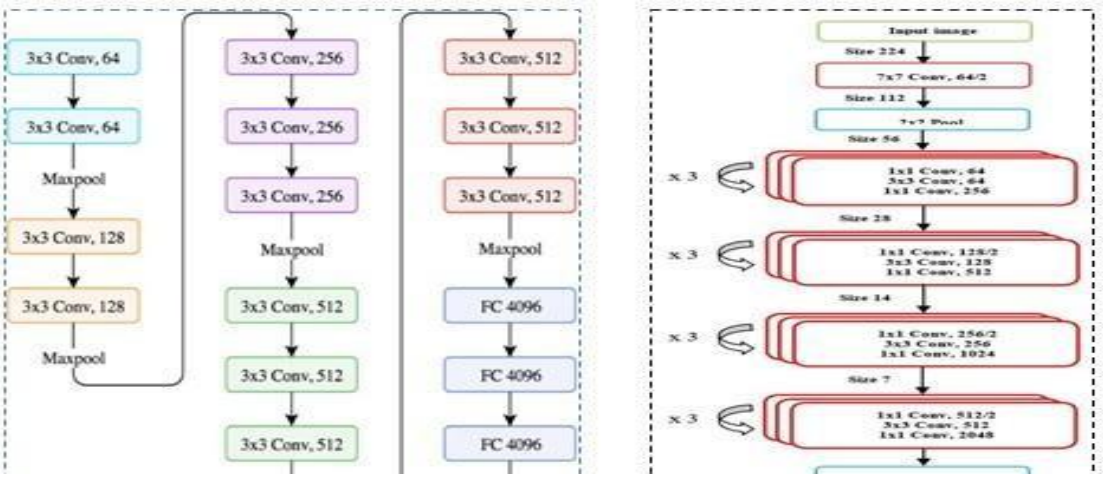

Figure 2.Hybrid ResNet, VGG and SVM structure

Hybrid ResNet, VGG and SVM structure is illustrated in figure 2.A series of convolution and pooling operations, followed by a number of fully connected layers is used in CNN. For performing multiclass classification Softmax is used in the fully connected layer. Another advantage of stacking two convolutions instead of one is the use of two relu operations, and more non-linearity gives more power to the model. The number of filters increases while going for deeper network. The spatial size of the feature maps decrease since involving pooling operation, but the depth of the volumes increases because of using more filters.

Residual Network (ResNet):

A conventional feed forward network with a residual connection is called ResNet. A number of fundamental residual blocks make up the residual network. However, based on the various architectures of residual networks, the operations in the residual block can be altered. These residual blocks replicate the inputs of all the layers and put forward them to the next successive layer. ResNet employs Batch Normalization process as a central part. This Batch Normalization alters the input layer to enhance the network performance. The covariate shift problem is alleviated. Residual Network utilises identity connection, that helps to solve the Vanishing gradient problem of the network.

Visual Geometry Group (VGG)

VGG is notable for its simplicity and is one of the top CNN models. VGG is a 16 layer neural network, not counting the maxpool layers and the softmax at the end. It's also referred to as VGG16. The architecture is the one we worked with above with Stacked convolution and

pooling layers followed by fully connected ANN.

Concatenation

The feature vectors of the Pool 5 and Average Pool layers of the VGG and ResNet architectures are used in this proposed model for each image. Merging feature vectors is done by this layer. To eliminate the size difference between the two feature vectors, the padding process is applied to the feature vector in the ResNet average pool layer and the size is equalized to the VGG feature vector. Therefore, two distinct feature vectors of a single image were combined to create a single feature vector.

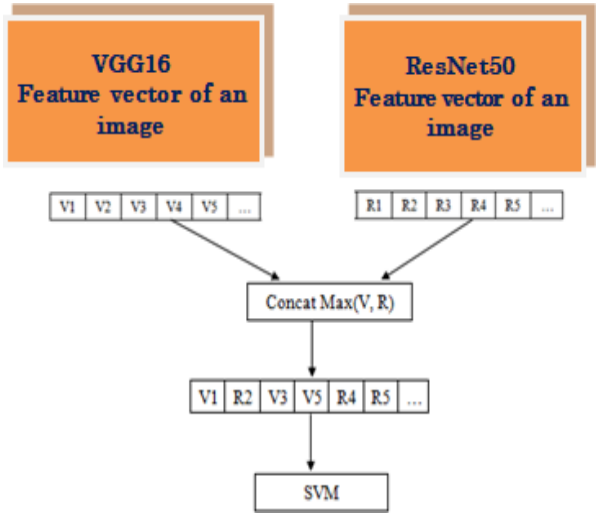

Figure 3. Concatenation of features

Figure 3.illustrates the feature concatenation process. The feature vectors of VGG is represented as V1, V2, V3,... and the feature vectors of ResNet is represented as R1,R2, R3, … Here, the maximum values of the VGG and ResNet features are concatenated for giving as input to the SVM classification.

Support Vector Machine (SVM)

SVM is a supervised learning algorithm. Regression analysis and classification both make extensive use of it. In supervised learning, a classification requires firstly the training of the classifier by means of the instances from a training set, which is formed by the images used for training. Each of these instances is defined by one or more features, and a class label that represents the correct classification of each instance. The goal of SVM is to train a classifier able to predict unknown class labels of instances. In order to achieve it, SVM finds an optimal hyperplane that separates the different classes in the training set and obtains the SVM classifier. The instances from the test set, which is made up of the images to be evaluated, are then classified between the various classes using the SVM classifier. The primitive features of each

original patch are automatically classified and grouped into clusters using a support vector machine (SVM). The aim is to obtain a feature-based image patch classification by assigning a semantic label to each patch using a user-oriented terminology of real-world categories. The classification accuracy of images is evaluated by computing characteristic metrics (like precision, recall, Dice index) and comparing the classification results of new images with reference data.

4. Experimental Results:

Data set details

The Pavia University dataset is a collection of hyperspectral images that were captured by the ROSIS-3 reflective optics system imaging spectrometer over the Italian city of Pavia. 115 spectral bands and 610340 pixels make up the picture

. The asphalt, meadows, gravel, trees, metal sheet, bare soil, bitumen, brick, and shadow are among the 9 classes that make up the image's 42,776 tagged samples.

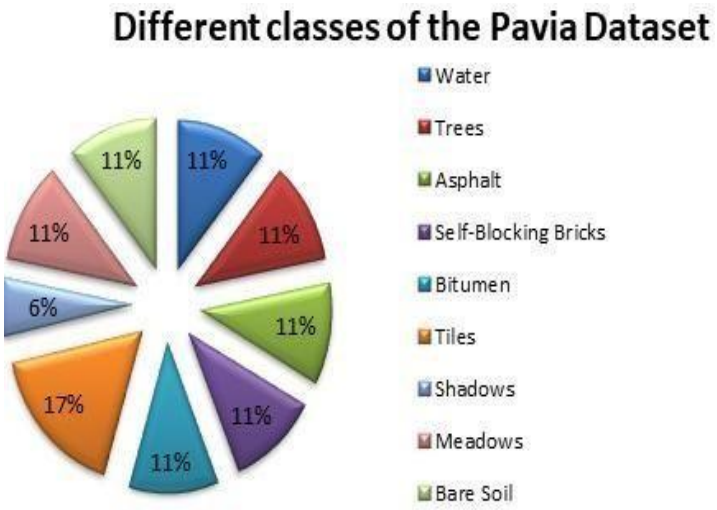

Figure 4.Different classes of the Pavia Dataset

The graphical representation of the ground truth classes for the Pavia centre scene and their respective samples numbers is represented in figure 4.

Comparison of performances

The performance comparison depicted in the table 2 demonstrates the success of the proposed model. The figure 4 shows the performance comparison of existing and proposed methods. It shows that the proposed algorithm achieved a considerable amount of accuracy improvement than existing methods.

Table 1. Comparison of performances of CNNModels

CNN Models	Accuracy(%)	Precision(%)	Recall (%)	Dice Index (%)
AlexNet	95.1	90.23	95.81	92.85
ResNet	94.88	88.55	94.6	92.3
VGG	95.92	90.98	96.16	93.26
VGG, ResNet and SVM	96.31	91.65	97.1	94.32

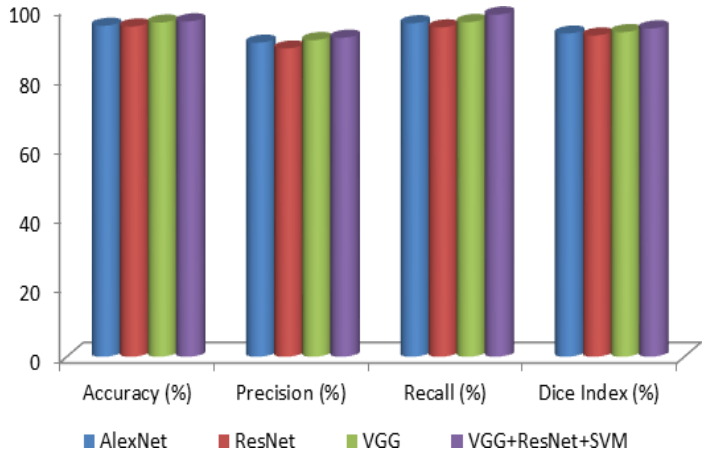

Figure.5 performance comparison of proposed and existing methods

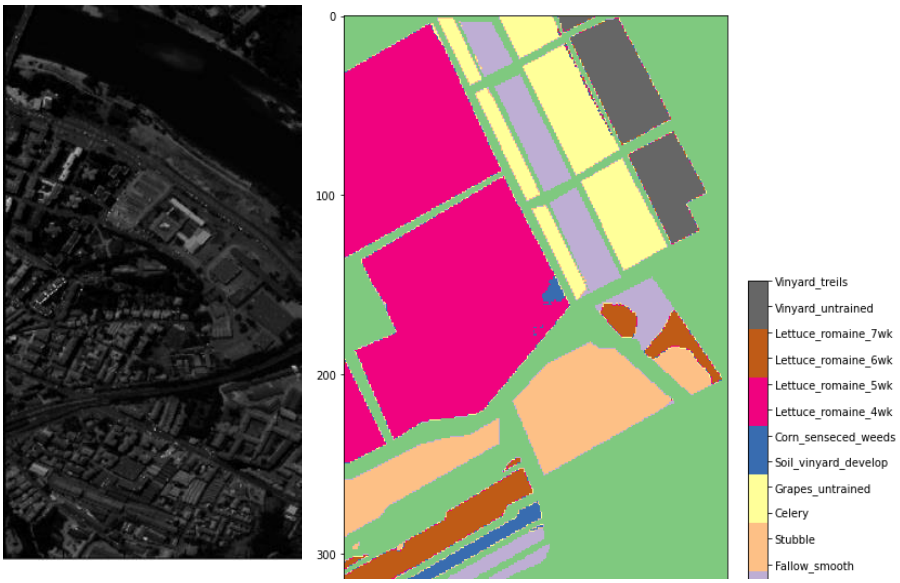

Figure 6 (a) Input image (b) Output image

Figure 6 (a) and (b) shows the input and output of the proposed method.

Figure 7 shows the training and testing accuracies and the losses during the training phase of the model. It also shows that this model was able to converge to a local minimum within 50 epochs.

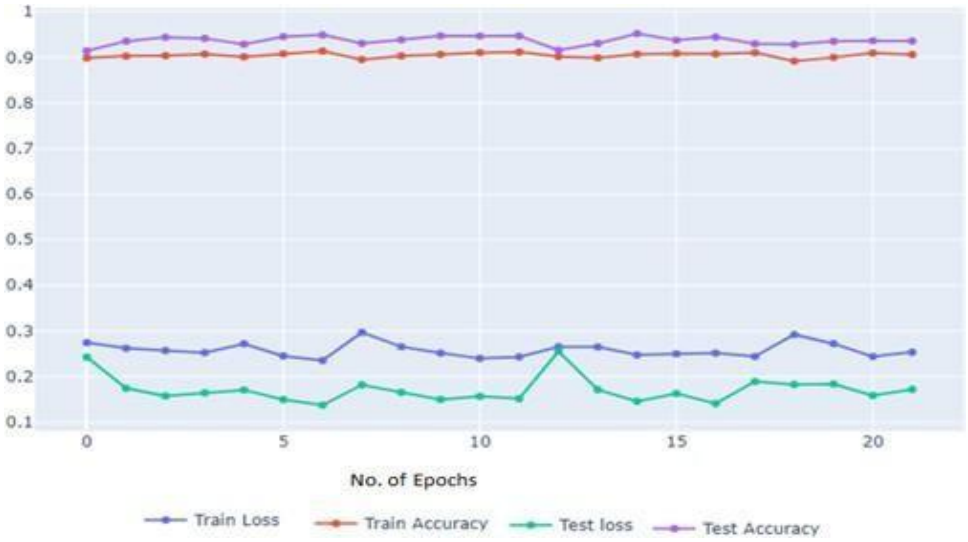

Figure 7. Training and testing losses and accuracies

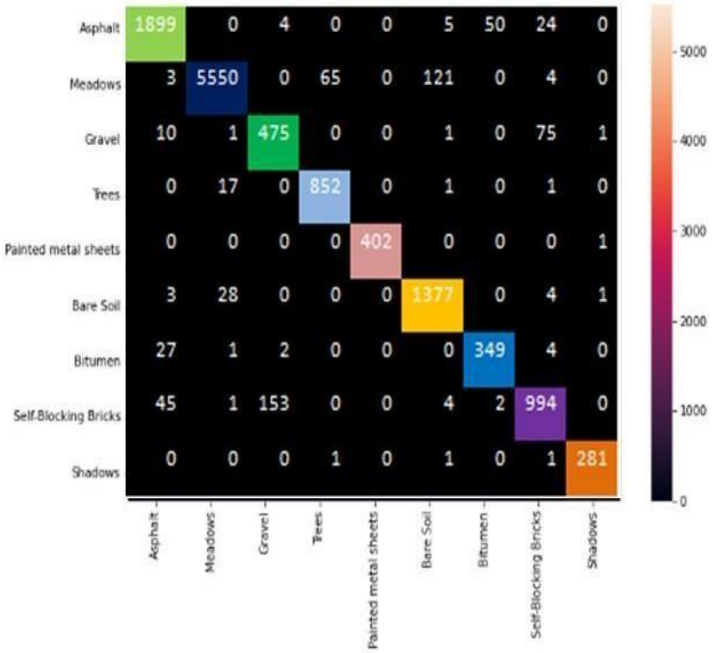

Figure 8. Confusion matrix for the classification

Figure 8 shows the confusion matrix of the proposed classification approach. The values entered in the matrix form represent the summary of the correct and incorrect estimates with count values for each class.

Table.3. Performance parameters of the proposed method

	Precision	recall	f1-score	support
1. Asphalt	0.95	0.95	0.95	1982
2. Meadows	0.98	0.96	0.97	5743
3. Gravel	0.75	0.85	0.80	555
4. Trees	0.93	0.98	0.95	871
5. Painted metal sheets	0.99	0.99	0.99	403
6. Bare Soil	0.91	0.98	0.94	1413
7. Bitumen	0.88	0.92	0.90	383
8. Self-Blocking Bricks	0.90	0.83	0.86	1199
9. Shadows	0.99	0.98	0.98	284
accuracy			0.96	12833
Macro avg	0.92	0.94	0.93	12833
weighted avg	0.95	0.95	0.95	12833

Table 3 shows the performance parameters of the proposed method. The foremost motive of involving this convolutional neural network models for image classification is to realize the accuracy performance when training the models. All over the algorithm, the total number of forward and backward passes or iterations taken by training dataset is called an epoch. The amount of epochs is called as hyperparameter. Each sample present in the training data had a consideration to update its parameters one time for the duration of an epoch. In VGG model, the network groups the multiple convolution layers with reduced kernel size rather than having a single convolution layer with huge kernel size and had an average pooling layer. This reduced the number of features at the output and the use of ReLU layers increases the learning rate. When training a neural network a large value of input space is mapped with a small input space, which causes the derivative disappearance. The residual neural networks mitigates these effects by means of using ReLU activation function and hosting short paths that can bring gradient all over the range of deep neural networks. The ResNet structure does not require all the neurons to fire in each epoch. It significantly decreases the training period in addition improves accuracy performance and reduce the losses. Once if a feature is learnt, then it doesn't attempt to learn the same once more but relatively attentions on learning other different features. Low validation loss in the ResNet indicates that there is not any over-fitting occurs during training. This approach improves the performance of the model training. Also, SVM performs better when a clear separation margin exists between the classes and which is most efficient in large dimensional spaces as well as it is memory efficient classifier. This improves the classification performance. The improvement of the performance of the proposed method is due to the average pooling layer of VGG, ResNet feature vector evaluation and the SVM algorithm for classification.

5. CONCLUSION

In order to categorize the land cover from high-resolution remote sensing data, the proposed method takes the concept of deep neural networks and proposes a hybrid ResNet and VGG-based convolutional neural network (VGG-CNN). The effectiveness of our proposed solution can be attributed to two key factors. The Average pool layer's VGG, ResNet feature vector evaluation comes first. In order to classify this, the feature with the highest value across two feature vectors was chosen. The second is the use of SVM, a powerful classification algorithm.To compare and judge the superiority of the recommended strategy, the classification was performed using pretrained versions of ResNet, VGG, and AlexNet. The hybrid ResNet-VGG SVM model offers more accuracy compared to the pretrained ResNet and VGG. The suggested model outperforms the already available models in terms of accuracy, precision, recall, and dice-index. In future, the proposed hybrid model can be implemented in medical images and transfer learning can also be applied in the pretrainedResnet and VGG models to further improvement of the performances.

REFERENCES

1. Dasari, K., Lokam, A. “Exploring the Capability of Compact Polarimetry (Hybrid Pol) C band RISAT-1 data for Land Cover Classification”. IEEE Access, 1–1, 2018.
2. Xia, J., Yokoya, N., Adriano, B., Zhang, L., Li, G., Wang, Z. “A Benchmark High-Resolution GaoFen-3 SAR Dataset for Building Semantic Segmentation”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 5950–5963, 2021.
3. Wijaya, A., Gloaguen, R. “Fusion of ALOS Palsar and Landsat ETM data for land cover classification and biomass modeling using non-linear methods”, 2009 IEEE International Geoscience and Remote Sensing Symposium, 2009.
4. Coca, M., Anghel, A., Datcu, M. “Unbiased Seamless SAR Image Change Detection Based on Normalized Compression Distance”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 1–9, 2019.
5. Dong, X., Quegan, S., Yumiko, U., Hu, C., Zeng, T. “Feasibility Study of C- and L-band SAR Time Series Data in Tracking Indonesian Plantation and Natural Forest Cover Changes”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(7), 3692–3699, 2015.
6. Amitrano, D., Guida, R., Iervolino, P. “High Level Semantic Land Cover Classification of MultitemporalSar Images Using Synergic Pixel-Based and Object-Based Methods”, IGARSS 2019 -

2019 IEEE International Geoscience and Remote Sensing Symposium, 2019.

7. Li, X., Lei, L., Sun, Y., Li, M., Kuang, G. "Multimodal Bilinear Fusion Network With Second-Order Attention-Based Channel Selection for Land Cover Classification", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 1011–1026, 2020.
8. Shi, X., Fu, S., Chen, J., Wang, F., &Xu, F. "Object-Level Semantic Segmentation on the High-Resolution Gaofen-3 FUSAR-Map Dataset", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 3107–3119, 2021.
9. G. Li, D. Lu, E. Moran, L. Dutra, and M. Batistella, "A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region", ISPRS J. Photogrammetry Remote Sens., vol. 70, pp. 26–38, 2012.
10. C. Rossi and E. Erten, "Paddy-rice monitoring using tandem-x", IEEE Trans. Geosci. Remote Sens., vol. 53, no. 2, pp. 900–910, 2015.
11. D. Cantorna, C. Dafonte, A. Iglesias, and B. Arcay. "Oil spill segmentation in SAR images using convolutional neural networks. A comparative analysis with clustering and logistic regression algorithms", Appl. Soft Comput., vol. 84: No. 105716, 2019.
12. Yu S, Jia S, Xu C, "Convolutional neural networks for hyperspectral image classification", Neurocomputing 219:88–98, 2017.
13. Hu W, Huang Y, Wei L, Zhang F, Li H, "Deep convolutional neural networks for hyperspectral image classification". J Sens. No. 258619, 2015.
14. Li Y, Hu J, Zhao X, Xie W, Li J, "Hyperspectral image super-resolution using deep convolutional neural network", Neurocomputing. 266 :29-41; 13, 2017.
15. Zhang L, Zhang L, Du B, "Deep learning for remote sensing data: a technical tutorial on the state of the art.", IEEE Geosci Remote SensMagn 4(2):22–40, 2016.
16. Chen Y, Jiang H, Li C, Jia X, Ghamisi P, "Deep feature extraction and classifcation of hyperspectral images based on convolutional neural networks", IEEE Trans Geosci Remote Sens 54(10):6232–6251, 2016.
17. AnjuUnnikrishnan, Sowmya V, Soman K P,"DeepAlexNet with Reduced Number of Trainable Parameters for Satellite Image Classification",Procedia Computer Science, Vol. 143: 931-938, 2018.