CONVOLUTIONAL NEURAL NETWORKS FOR HIGH SPATIAL RESOLUTION REMOTE SENSING IMAGE CLASSIFICATION

1. Sakhi. G[1], 2., R. Balasubramanian[2] , R.Nedunchezhian[3]

[1]*Research Scholar, Manonmaniam Sundaranar University,Tirunelveli - 627 012,*
sakthihit@gmail.com

[2]*Professor, Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli - 627 012,*
rbalus662002@yahoo.com

[3]*Professor, Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore - 641 014,*
rajuchezhian@gmail.com

Abstract

Hyperspectral remote sensing based image classification is found to be a very widely used method employed for scene analysis that is from a remote sensing data which is of a high spatial resolution. Classification is a critical task in the processing of remote sensing. On the basis of the fact that there are different materials with reflections in a particular spectral band, all the traditional pixel-wise classifiers both identify and also classify all materials on the basis of their spectral curves (or pixels). Owing to the dimensionality of the remote sensing data of high spatial resolution along with a limited number of labelled samples, a remote sensing image of a high spatial resolution tends to suffer from something known as the Hughes phenomenon which can pose a serious problem. In order to overcome such a small-sample problem, there are several methods of learning like the Support Vector Machine (SVM) along with the other methods that are kernel based and these were introduced recently for a remote sensing classification of the image and this has shown a good performance. For the purpose of this work, an SVM along with Radial Basis Function (RBF) method was proposed. But, a feature learning approach for the classification of the hyperspectral image is based on the Convolutional Neural Networks (CNNs). The results of the experiment that were based on various image datasets that were hyperspectral which implies that the method proposed will be able to achieve a better performance of classification compared to other traditional methods like the SVM and the RBF kernel and also all conventional methods based on deep learning (CNN).

***Keywords**: Remote Sensing Image Classification, Hyperspectral Images, Artificial Neural Network (ANN), Support Vector Machine (SVM), Radial Basis Function (RBF) and Convolutional Neural Networks (CNNs).*

I. INTRODUCTION

The Hyperspectral remote sensors were able to capture images having pixels that are represented as spectral vectors that range from visible spectral to the signatures and this further enables a High spatial resolution data of remote sensing as a powerful tool for image classification [1]. By a combination of imaging with spectroscopy technology, a hyperspectral remote sensing may become both spatially and spectrally continuous. The hyperspectral data is now becoming a very valuable tool to monitor the surface of the earth and have been used in a wide range of applications. There is yet another incomplete list that includes environmental sciences, chemical imaging, astronomy, physics, surveillance, mineralogy and agriculture. A technique commonly used in such applications will be the classification of pixels in the hyperspectral data. In case this is exploited successfully, hyperspectral data will be able to yield an accuracy in terms of a higher yield and will also be more detailed in terms of class taxonomies. There are, however, many crucial issues in classifying hyperspectral data: 1) the curse of dimensionality, owing to the high spectral channels; 2) a limited number of samples of labelled training; and 3) a large spatial variability of that of spectral signature [2].

The conventional methods of High spatial resolution remote sensing image classification were based on the spectral information. The typical classifiers are the ones based on the distance measure, the K-Nearest Neighbours (KNN), the logistic regression and the maximum likelihood criterion. The accuracy of classification of such methods is normally not satisfactory owing to the "small-sample problem": which is a sufficient number of samples of training that are not available for the spectral bands of a high number. This type of imbalance between the spectral bands and their high dimensionality is called the Hughes phenomenon. The spectral redundancy was observed to be some of the spectral bands of the hyperspectral data that may be correlated. Also, these algorithms of classification exploit all spectral information that fails in capturing all critical spatial variability that was perceived for the data of high resolution normally resulting in a lower performance. For improving the performance of classification, the intuitive idea was designing classifiers making use of spatial and spectral information in the pixel-level classifiers. The spatial information further provides some more discriminant information which is connected to the size and shape of various structures that when exploited properly results in classification maps that are accurate [3].

In the last two decades, there were various techniques of machine learning that includes the Artificial Neural Networks (ANNs) along with the SVMs which were applied successfully to the classification of hyperspectral images. Particularly, the neural architectures have been demonstrating a lot of potentials to model the mixed pixels resulting in a low spatial resolution of multiple scattering and hyperspectral cameras. But at the same time, there are many limitations that are involved with the ANNs using a Back Propagation (BP) algorithm which is a very popular technique to be a learning algorithm. The development of a Neural Network (NN) model for developing hyperspectral data and this is an expensive procedure as the hyperspectral images which are typically represented to be the three-dimensional cubes that have several spectral channels. Additionally, the ANNs will need a great deal of turning of hyper parameter like the hidden layers, the nodes found in each layer, the rate of learning and so on. Recently, the approaches based on the SVM that was used extensively for the classification of the hyperspectral image as the SVMs outperformed the traditional neural and statistical methods. This is because the maximum likelihood along with the Multi-Layer Perceptron Neural Network (MLPNN) based classifiers. Also, the SVMs have been demonstrating a great performance in classifying the hyperspectral data when there is a very low number of training samples that are made available. But the parameters of the SVM (the kernel parameters and the regularization) will have to be tuned for the performance of an optimal classification [4].

Deep learning is a new approach that aims for an artificial intelligence. The deep learning methods have built a network which has many layers that are typically deeper compared to the three layers. The Deep Neural Network (DNN) may further represent some complicated data but it can get quite difficult to train this network. Owing to the non-availability of a suitable training algorithm, it was quite challenging to harness such a powerful model. This Deep learning involves a new class of models attempting to learn various levels in data representation helping in taking advantage of some input data like text, speech and image. The deep learning model has been usually initialized through some unsupervised learning and this was followed by the fine-tuning in a manner that is supervised. These high-level features that may be understood from all the low-level features. Such learning results in the extraction of invariant and abstract features that were beneficial for various tasks like target detection and classification [5].

There are some more models in deep learning found in the literature that include the Deep Belief Network (DBN), the Stacked Auto-Encoder (SAE) and finally the CNN. In recent times, the CNNs were identified as some good alternatives to the other models of deep learning and for this work, the SVM and the RBF kernel along with the high spatial resolution based on the CNN. The rest of the investigation has been organized thus. Section 2 discusses all related work found in the literature. Section 3 explains the methods employed. Section 4 discusses the results of the experiment and the conclusion is made in Section 5.

II. RELATED WORKS

Zhao and Du [6] had proposed a Spectral–Spatial Feature-based Classification (SSFC) jointly makes use of the reduction of dimension along with the techniques of deep learning for the extraction of both spatial and spectral extraction. For the purpose of this framework, there was a balanced and local discriminant embedding algorithm that was proposed for the extraction of features. At the same time, the CNN was used to identify all spatial-related features at high levels. After this, a fusion feature is duly extracted by the stacking of spectral features with spatial features. Lastly, there is a multiple-feature-based classifier which has been trained for classification of the image.

Yu et al., [7] had made a proposal for another efficient architecture of the CNN for boosting the discriminative capability in a hyperspectral classification of images. This CNN infrastructure also has some advantages. First, there are various methods of traditional classification which need some hand-crafted features, and the CNN model that is employed has been designed for dealing with the issues of a hyperspectral image analysis used in an end-to-end manner. Next, all parameters of a CNN model will be optimized from a smaller training set and the problem of overfitting will be alleviated for a certain extent. Lastly, for getting a better deal using the hyperspectral image information, the 1*1 convolutional layers were adopted having an average layer of pooling with a dropout rate that was large which were employed in the entire process of the CNN.

Zhang et al., [8] had made a proposal for another new Dual-Channel CNN (DC-CNN) framework which was used for an accurate spectral and spatial classification made on the hyperspectral images. For this framework, there was a one-dimensional CNN (1D CNN) which was used for an automatic extraction of the hierarchical and the spectral features along with the two-dimensional CNN (2D CNN) that was applied for the extraction of the space-related features along with a softmax regression classifier that was used for combining both the spectral and the spatial features with a softmax regression classifier which was used for combining both the spectral and the spatial features and duly predict their results of classification. In order to overcome this problem of a limited training sample availability in the hyperspectral images, a simple method of augmentation was proposed by the authors and this was both effective and efficient in improving the accuracy of a hyperspectral image classification.

Liang and Li [9] had made a proposal of a method for the classification of a remotely-sensed image by employing the sparse representation of the deep learning features. More specifically, the work makes use of the CNN for the extraction of deep features from the image data of higher levels. These deep features have provided a high level of spatial information which is created by the hierarchical structures. Even though the deep features have a high level of dimensionality, they fall under the class-dependent submanifolds and sub-spaces. The deep feature characteristics are investigated using a framework of sparse representation classification.

Li et al., [10] had further proposed another new model of hyperspectral image reconstruction which is based on the CNN for enhancing the spatial features. It further proposes some more new strategies for the band selection for defining the training label having some rich information. After this, there is some hyperspectral data which is trained by the deep CNN for building models having optimized parameters well-suited for the reconstruction of hyperspectral images. Lastly, these reconstructed images are classified using an efficient Extreme Learning Machine (ELM) employing a structure that is simple.

Samat et al., [11] had made an introduction of an ELM for the hyperspectral classification of the image. For overcoming the limitations of the ELM, by input weights and their randomness and also their bias, there were two other new algorithms of the ensemble ELMs (the Bagging-based and the Ada Boost-based ELMs) that were proposed for the task of classification. The results of the experiment with the hyperspectral image data that was real which was collected by the Reflective Optics Spectrographic Image System (ROSIS) along with the Airborne Visible or the Infrared Imaging Spectrometer (AVIRIS) show that this ensemble algorithm proposed can produce great performance of classification in various scenarios in relation to both spectral and the spectral-spatial feature sets.

Ouyang et al., [12] had made a new proposal of a method of hyperspectral image classification which was based on the CNN. The small and convolutional kernels get cascaded along with the pooling operator for performing a feature abstraction. There was a decoding channel that was composed of deconvolution along with the unpooling operators that are established. Any unsupervised reconstruction, which is made by that of the decoding channel introduces the priors to network training and will enhance the abstracted features and their discriminability using the control gate.

Cao et al., [13] had presented another new algorithm of supervised classification used for the hyperspectral images that are remotely sensed integrating the spectral and the spatial information within a unified Bayesian framework. It first formulates the problem of a hyperspectral image classification by looking at it from a Bayesian perspective. After this, the CNN is adopted for learning its posterior class distributions by employing a training strategy that is patch-wise. Once that is done, the spatial information is taken into consideration by means of placing the spatial smoothness earlier on labels. Lastly, there is an iterative updating of the CNN by employing a stochastic gradient descent and all pixel vector class labels are updated by using the algorithm which is an α-expansion mincut-based.

III. METHODOLOGY

For the purpose of this section, the SVM and the RBF kernel with the CNN for the remote sensing data SVM was used to separate the two-class data to learn the optimal decision hyperplane that separates all training examples within a kernel that includes a feature space of a high dimension. There are some extensions of the SVM within the hyperspectral image classification that was presented for improving the performance of classification. In the case of the remote sensing tasks of classification, the SVM is found to be superior to the NN that is used traditionally. The methods of deep learning also can achieve a very promising performance in various fields. In the case of deep learning, the role of the CNNs can be dominant in processing problems that are visual-related.

A. Support Vector Machine (SVM) with Radial Basis Function (RBF) Kernel

An SVM technique had been introduced by Vapnik and further developed in recent years. There are many other techniques to solve problems in regression and classification that have emerged recently. During the last decade, the SVM has become a very important technique of learning in several fields most importantly in text categorization, finance and computational biology. This is because of the built-in mechanisms, that ensure a good generalisation resulting in an accurate prediction that can train the large datasets fast. The primary objective of that of statistical learning was to identify a new description of a dependency that was unknown among objects and their properties. The measurements were also called the input variables and are observable in the various objects of interest. But at the same time, all properties of these objects or the output variables are available for a very small object subset called examples. The primary purpose of the estimation of the dependency existing between the input and the output variables can determine values of that of the output variables for objects of interest [14].

The SVM further maps the features non-linearly within an n-dimensional feature space when given a feature set which is capable of getting represented within the space. At the time a kernel has been introduced having a high computation in its SVM algorithm, all inputs will be in the form of the scalar products; after this, a classification was achieved by means of solving as below. This issue is then translated into another convex quadratic problem of optimization where there is a clear answer that is obtained by the convexity. In the case of an SVM, there is an attribute that is the predictor variable along with a feature which is a transformed attribute. There is a set of the features that describes an example as a vector. All features vectors will define a hyper plane and an optimal hyper place will be constructed by the SVM having the aim of separating all vector clusters using a class of attributes that are on one of the sides and other attributes on the other side. This margin duly represents any distance that is between that of the hyper plane and its support vectors. The SVM analysis attempts at positioning a margin in a way in which there is some space left between it and all the support vectors are thus maximized [15].

With a training set of $(x_i, y_i), i1,2,....,l$ wherein the $x_i \in R^n$ and the $y \in \{1,-1\}^l$, SVM can solve the problem of optimization of that of equation (1):

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \qquad (1)$$

This is subject to: $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0$.

Here a function ϕ will map vectors i x in that of a higher dimensional space. C>0 will be the penalty parameter for the error term.

The Lagrangian method is employed for solving the model of optimization and this is just like the method to solve problems in optimization for any other separable case. All dual variables Lagrangian is then maximized as per equation (2):

$$\underset{\alpha}{Max}\, L_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j < x_i . x_j > \qquad (2)$$

This is subject to: $0 \le \alpha_i \le C \; i = 1, \ldots, m \; and \; \sum_{i=1}^{m} \alpha_i y_i = 0$

For the purpose of computing an optimal hyper plane, there is a dual Lagrangian LD (α) that is maximized with regard to a non-negative αi which is subject to the constraints $\sum_{i=1}^{m} \alpha_i y_i = 0$ and the 0 ≤ αi ≤ C. A penalty parameter C, will now be the upper bound on the αi, which is determined by the user.

The kernel function will now be defined as $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$.

The RBF will be given as per equation (3):

$$K(x_i, x_j) = \exp\left(-\gamma \| x_i - x_j \|^2\right), \gamma > 0 \qquad (3)$$

There is a proper parameter setting which is found in kernels that can increase the accuracy of the SVM classification. There are further two more parameters that need to be determined within the SVM model along with that of the RBF kernel: C and the gamma (γ). A gamma parameter will now automatically define the actual distance that can be reached by a training example having low values that mean ‘far’ and some high values that mean ‘close’. C parameter will then give a trade-off for the training examples and their misclassification which is against the simplicity of a decision surface. The lower C values will now ensure a decision surface that is smooth and the higher C will attempt at classifying the examples accurately. The experiments are then conducted for evaluating the performance of the SVM by means of variations of Gamma and the C parameters.

B. Convolutional Neural Network (CNN)

In the case of machine learning the ANN becomes a system of all interconnected neurons passing messages to one another. The NNs are then used for modelling all complex functions and the work deals with the feed-forward networks. In spite of the fact that the NNs represent some complex functions, the loss and its epigraph L may be non-convex. This makes optimization challenging through an approach of gradient descent. For regularizing the loss, and for improving training, the CNNs are found to be a special NN which can impose some restrictions which make sense in image processing [16].

The CNNs are then biologically inspired along with the multilayer classes of the deep learning models using a single neural network which is trained from the image pixel values to the classifier outputs. For these networks, each neuron will be associated with a spatial location (i, j) and this is in relation to its input image. An output $a_{i,j}$ which is associated with the location (i, j) will then be computed as per equation (4):

$$a_{i,j} = \sigma((W * X)_{ij} + b) \qquad (4)$$

Wherein the W denotes the kernel that has learned weights, X denotes the input to a layer and ‘*’ indicates the convolution operation. It is to be noted that this can be a special case of neuron which is in $a = \sigma(wx + b)$ (wherein the x indicates the weight vector, b denotes a scalar called the bias and the σ denotes a new activation function) having the constraints shown below:

- All connections will only extend to a particular spatial neighbourhood which is limited and determined by the size of the kernel;
- The very same filter will be applied to every location which guarantees a translation invariance.

Ideally, there are many convolution kernels to be learnt at each layer and interpreted to be a set of the detectors of spatial features. All responses to each learned filter are called a feature map.

Departing from its traditional and fully connected layer where each neuron has been connected to the outputs of the earlier layer and a convolutional layer has dramatically brought down the parameters by means of enforcing all the aforementioned constraints. This may further result in a new regularized function and does not lose a lot of generalities. It is to be noted that all convolution kernels are found to be three dimensional as aside from their spatial extent they may go through some feature maps in their earlier layers of by means of their bands within their input image. As the third dimension is inferred from the earlier layer, it is very rarely specified to be in the architectural descriptions and only two of the spatial dimensions are mentioned.

Adding on to the convolutional layers, there are some networks that are state-of-the-art like the Image net that involve a certain degree of down-sampling [17], which denotes a reduction of resolution of all feature maps. The primary goal of that of down-sampling was increasing the neurons and their receptive fields. This is a part of the input image which the neurons are able to “see”. For the purpose of predictions to consider large spatial context, all upper layers will have a large receptive field. This may be achieved by means of increasing the kernel sizes or by means of down-sampling all feature maps to a resolution which is lower. Its first alternative will increase the parameters and the consumption of memory thus making both training and inference prohibitive. The CNNs that are state-of-the-art keep kernels small and also add a certain level of down-sampling to it. This may be accomplished by means of including the pooling layers (considering an average or a maximum of the adjacent locations) or by means of introducing a stride that amounts to be able to skip certain convolutions through the application of a filter once for each of the four locations.

The classification networks normally consist of a layer that is fully connected on the top of convolutions or pooling. The layer has been designed to contain several outputs and produce the final scores of the classification. the success of that of the CNNs is in the fact that these networks are now forced by the construction for learning the hierarchical and contextual translation-invariant features that are useful for the categorization of the images.

This hierarchical architecture belonging to the CNNs has proved to be efficient in learning visual representations. The basic challenge of these visual tasks was to model any intra-class appearance and also shape the object variation. This hyperspectral data having hundreds of spectral channels that can be illustrated as the 2D curves (the 1D array), which is a curve of every class with its own visual shape and is very different from the other classes and even though it is quite difficult for distinguishing certain classes of them with the human eye (such as the gravel and the self-blocking bricks). It is known that the CNNs may be able to achieve better performance than that of a human being in certain problems of vision. The capability further inspires an investigation of the application of the CNNs for a high spatial resolution and remote sensing image classification making use of spectral signatures [18].

In the case of the hyperspectral remote sensing, every hyperspectral image pixel may be taken to be a 2D image that has a height which is equal to 1 (to be 1D audio inputs in their speech recognition). So, the input layer size will be only (n_1, 1), and the n_1 will denote the actual number of bands. The very first layer of hidden convolution C1 will filter the $n_1 \times 1$ of the input data with about 20 kernels of size $k_1 \times 1$. The layer C1 will consist of $20 \times n_2 \times 1$ nodes, and $n_2 = n_1 - k_1 + 1$. There have been about $20 \times (k_1 + 1)$ which are trainable parameters falling between the layer C1 and its input layer. There is a maximum pooling layer known as M2 which is the next hidden layer where the kernel size will be (k_2, 1). The layer M2 consists of $20 \times n_3 \times 1$ nodes, and the $n_3 = n_2/k_2$. There has been no parameter for this layer. A fully connected layer which is F3 consists of n_4 nodes where there are about $(20 \times n_3 + 1) \times n_4$ of trainable parameters that are between this layer and the layer M2. There is also an output layer that has n_5 nodes, along with $(n_4 + 1) \times n_5$ of trainable parameters that are between this particular layer and the layer F3. Thus, the proposed CNN classifier's architecture has a total of $20 \times (k_1 + 1) + (20 \times n_3 + 1) \times n_4 + (n_4 + 1) \times n_5$ parameters which are trainable.

IV. RESULTS AND DISCUSSION

In this section, the SVM with RBF kernel and CNN methods are used. The summary of results as shown in table 1. The classification accuracy, average precision, average recall and Root Mean Square Error (RMSE)as shown in figures 1 to 4.

Table 1Summary of Results

	SVM with RBF Kernel	CNN
Classification Accuracy	89.54	92.4
Average Precision	0.8952	0.9238
Average Recall	0.8947	0.9228
RMSE	0.3234	0.1511

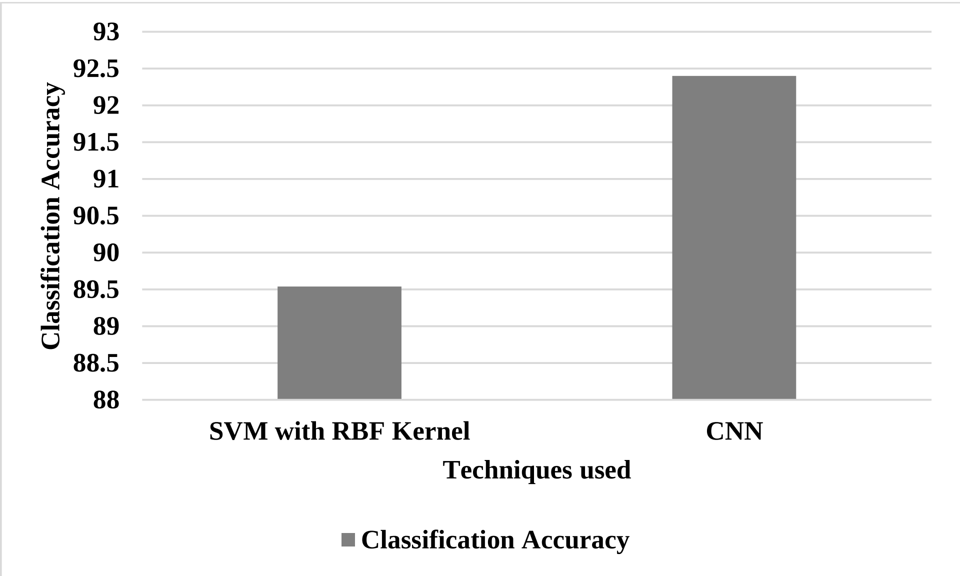

Figure 1Classification Accuracy for CNN

From the figure1, it can be observed that the CNN has higher classification accuracy by 3.14% for SVM with RBF kernel.

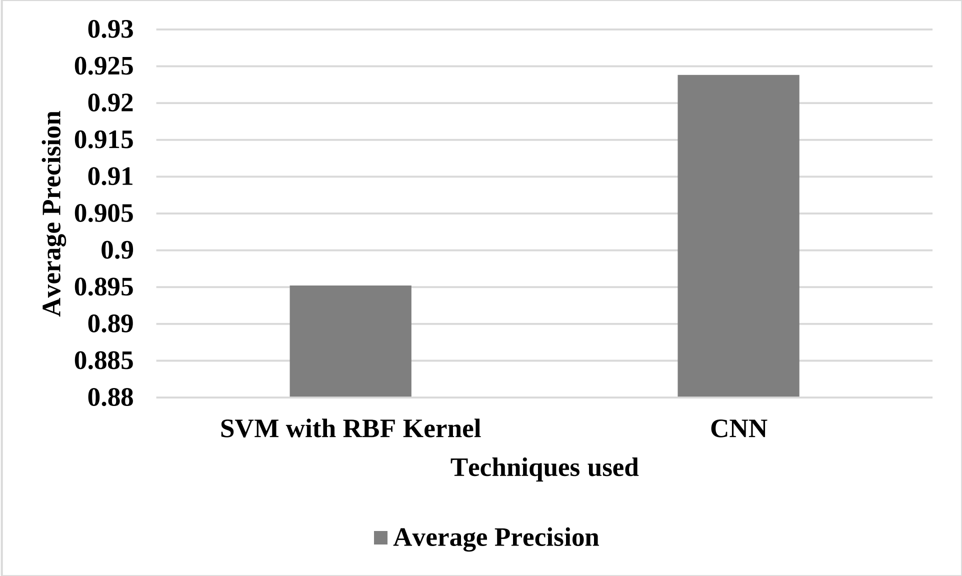

Figure 2Average Precision for CNN

From the figure 2, it can be observed that the CNN has higher average precision by 3.14% for SVM with RBF kernel.

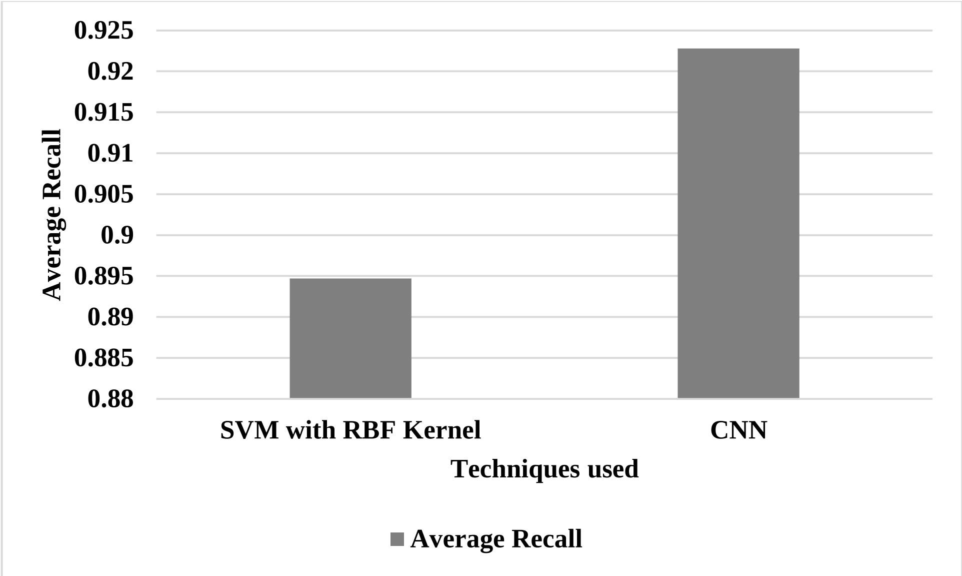

Figure 3Average Recall for CNN

From the figure 3, it can be observed that the CNN has higher average recall by 3.09% for SVM with RBF kernel.

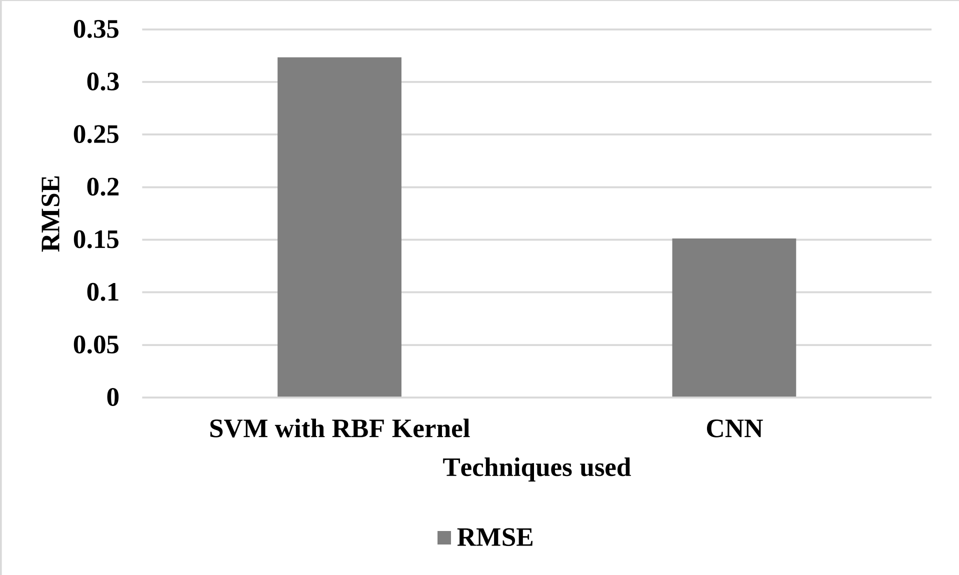

Figure 4 RMSE for CNN

From the figure 4, it can be observed that the CNN has lower RMSE by 72.62% for SVM with RBF kernel.

V. CONCLUSION

Owing to the hyperspectral data's inherent nature, the good features and their discernment of all good features of a hyperspectral image classification that is challenging. The approaches to deep learning need datasets of a large scale and their size have to be proportional to the parameters that are employed by the network for avoiding over fitting to learn the network. Thus the work proposes another novel method which is based on the CNN for a remote sensing classification of images that are inspired by an observation of the remote sensing image classification that may be implemented through human vision. The primary aim of the classification of the support vector was devising an efficient method to learn the good separating hyperplanes found in a feature space which is of a high dimension. Generally, all SVM based classifiers get a better accuracy of classification compared to the other techniques of pattern recognition. Generally, an RBF kernel will be the very first choice. The kernel will nonlinearly map all samples within the high dimensional space. Thus, unlike the linear kernel, the case where relation among class labels and the attributes that are nonlinear can be handled. The CNN makes use of the local connections for the efficient extraction of spatial information and the shared weights for the reduction of parameters. This work is an exploration of making use of the CNNs for the spatial resolution remote sensing and image classification that has a great performance. The results prove that a CNN has a higher level of accuracy of classification by about 3.14% for the SVM with the RBF kernel.

REFERENCES

1. Tao, C., Pan, H., Li, Y., & Zou, Z. (2015). Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. IEEE Geoscience and remote sensing letters, vol.2, no.12, pp. 2438-2442.

2. Chen, Y., Lin, Z., Zhao, X., Wang, G., &Gu, Y. (2014). Deep learning-based classification of hyperspectral data. IEEE Journal of Selected topics in applied earth observations and remote sensing, vol.7, no.6, pp. 2094-2107.
3. Li, Y., Zhang, H., & Shen, Q. (2017). Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sensing, vol.9, no.67, pp.2-21.
4. Chen, C., Li, W., Su, H-., & Liu, K. (2014). Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine. Remote sensing, vol.6, no.6, pp.5795-5814.
5. Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, vol.54, no.10, pp.6232-6251.
6. Zhao, W., & Du, S. (2016). Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing, vol.54, no.8, pp.4544-4554.
7. Yu, S., Jia, S., & Xu, C. (2017). Convolutional neural networks for hyperspectral image classification. Neurocomputing, vol.219, pp. 88-98.
8. Zhang, H., Li, Y., Zhang, Y., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. Remote Sensing Letters, vol.8, no.5, pp.438-447.
9. Liang, H., & Li, Q. (2016). Hyperspectral imagery classification using sparse representations of convolutional neural network features. Remote Sensing, vol.8, no.99, pp.1-16.
10. Li, Y., Xie, W., & Li, H. (2017). Hyperspectral image reconstruction by deep convolutional neural network for classification. Pattern Recognition, Vol. 63, pp. 371-383.
11. Samat, A., Du, P., Liu, S., Li, J., & Cheng, L. (2014). Ensemble Extreme Learning Machines for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(4), 1060-1069.
12. Ouyang, N., Zhu, T., & Lin, L. (2018). Convolutional Neural Network Trained by Joint Loss for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters. pp.1-5
13. Cao, X., Zhou, F., Xu, L., Meng, D., Xu, Z., & Paisley, J. (2018). Hyperspectral image classification with Markov random fields and a convolutional neural network. IEEE Transactions on Image Processing, vol.27, no.5, pp.2354-2367.
14. Apostolidis-Afentoulis, V., &Lioufi, K. I. (2015). SVM classification with linear and rbf kernels. pp. 1-7.
15. Renukadevi, N. T., & Thangaraj, P. (2013). Performance evaluation of SVM–RBF kernel for medical image classification. Global Journal of Computer Science and Technology. Volume, 13 Issue 4, pp.14-19

ISSN NO: 0972-1347

16. Maggiori, E., Tarabalka, Y., Charpiat, G., &Alliez, P. (2017). Convolutional neural networks for large-scale remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, vol.55, no.2, pp.645-657.
17. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097-1105.
18. Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification, Journal of Sensors, Volume 2015, PP.1-12