A Comparison Between Support Vector Machine (SVM) and Convolutional Neural Network (CNN) Models For Hyperspectral Image Classification

A Comparison Between Support Vector Machine (SVM) and Convolutional Neural Network (CNN) Models For Hyperspectral Image Classification

Hayder Hasan , Helmi Z.M.Shafri* , Mohammed Habshi

Department of Civil Engineering, Faculty of Engineering, University Putra Malaysia,43400UPM

*helmi@eng.upm.edu.my

Abstract. This study presents a methodology model for the spectral classification of hyperspectral images. The applied methodology, first extracts neighbouring spatial regions via a suitable statistical support vector machine (SVM-Linear) architecture, support vector machine radial basis function (SVM-RBF) and Deep Learning (DL) architecture that comprises principal component analysis (PCA) and convolutional neural networks (CNN) and then applies a soft max classifier. PCA is introduced to reduce the high spectral dimensionality, noise, and redundant information of the input image. The SVM-Linear, SVM-RBF and CNN model is used to extract useful high-level features automatically given that it provides results comparable with each other, including hyperspectral image classification. However, because the CNN , SVM models alone may fail to extract features with different scales and to tolerate the large-scale variance of image objects, the presented methodology uses PCA optimization for spatial regions to construct features that can be then used by the SVM and CNN model to classify hyperspectral images. Experimental results obtained in the classification of the Hyperspec-VNIR Chikusei datasets show that the performance of the presented model is competitive with that of other DL and traditional machine-learning methods. The best overall accuracy of the presented methodology for the Hyperspec-VNIR Chikusei datasets is 98.84 % in SVM-RBF model.

1. Introduction

Land cover and land use (LULC) mapping using remote sensing data is a long-standing challenge due to various and mix features especially in heterogeneous urban areas. LULC information is widely used in Geographic Information System (GIS) applications, including urban planning, environmental studies, natural hazard assessment, transportation management, and city designs. Thus, accurate, high quality and detailed maps are in demand everywhere and at any time. While remote sensing images and techniques, as well as field surveying, have been the primary sources of determining land use features, in-field measurements of ground truth data collection for attributing those features has always been a challenging step regarding time, money, as well as information reliability. Hyperspectral imaging (HSI) is an active area in the field of remote sensing research given that it provides rich spectral and spatial information depending on system type [1]. In recent years, the application of a wide range of methods, such as kernel machines [2], morphological profiles [3, 4], and deep learning (DL) [5-7] in HSI classification has been extensively investigated. Research on the development of new and advanced classification models mainly aim to boost classification accuracies while

maintaining computational time within a reasonable range. Accurate classification maps are used in a wide range of applications, including urban mapping and planning, crop identification and analysis, environmental management, and mineral exploration.

Two main critical issues, in particular, influence the classification results obtained for HSI data [8]. First, the high spectral dimension, which may reach hundreds of spectral bands, of images introduces challenges (e.g., Hughes phenomenon) that ultimately limit the accuracy of the classification results. In heterogeneous regions, different objects, for example, parking lots or roofing materials, may exhibit a similar spectral signature. This similarity complicates the classification of HSI data when only spectral information is used. Second, the high spatial resolution of images may decrease interpretation accuracies. Specifically, high spatial resolution may increase intraclass variation and decrease interclass variation in spectral and spatial domains. To address these issues and boost classification accuracies, researchers have attempted to develop and jointly use methods that are based on spectral and spatial information.

Recent efforts in HSI classification have mostly focused on DL methods because of their successful application in other fields, such as computer vision [9], speech recognition [10], and natural language modeling [11]. DL has been utilized in various ways, including supervised [12], semisupervised [13], unsupervised [6], dimensionality reduction [8], and self-learning [14] paradigms, for HSI classification. Recent efforts in HSI classification have mostly focused on DL methods because of their successful application in other fields, such as computer vision [15], speech recognition [16].DL has been utilized in various ways, including supervised [17], semisupervised [18], unsupervised [19], dimensionality reduction [20], and self-learning [21] paradigms, for HSI classification.

In this study, a compassion between Support Vector Machine (Linear SVM and RBF SVM) and convolutional Neural Networks (CNN) based on programming-based methodology is presented in order to classify hyperspectral (HSI) data. The proposed methodology comprised four steps: (1) dimensionality reduction with PCA, step (2) HSI classification by using SVM model, step (3) the extraction of high-level features by CNN, the final step involved a comparison between two models (SVM and CNN) was applied based on the accuracy assessment.

2. Materials and methods

2.1. Hyperspec-VNIR Chikusei

The Chikusei dataset (Fig. 1) comprises images of agricultural and urban areas in Chikusei, Ibaraki, Japan. Headwall's Hyperspec-VNIR-C imaging sensor took the images on July 19, 2014. This dataset was made public by Dr Naoto Yokoya and Prof. Akira Iwasaki from the University of Tokyo. The dataset comprises 128 bands in the spectral range of 0.363 μm to 1.018 μm. The study scene has a size of 540 pixels × 420 pixels .The central point of the scene is located at coordinates: 36.294946N, 140.008380E. and a spatial resolution of 2.5 m. The reader may refer to Yokoya and Iwasaki (Yokoya & Iwasaki, 2016) for additional details about this dataset. The dataset contains 19 land-cover classes in the dataset with 15,518 training samples and 62,074 testing samples.

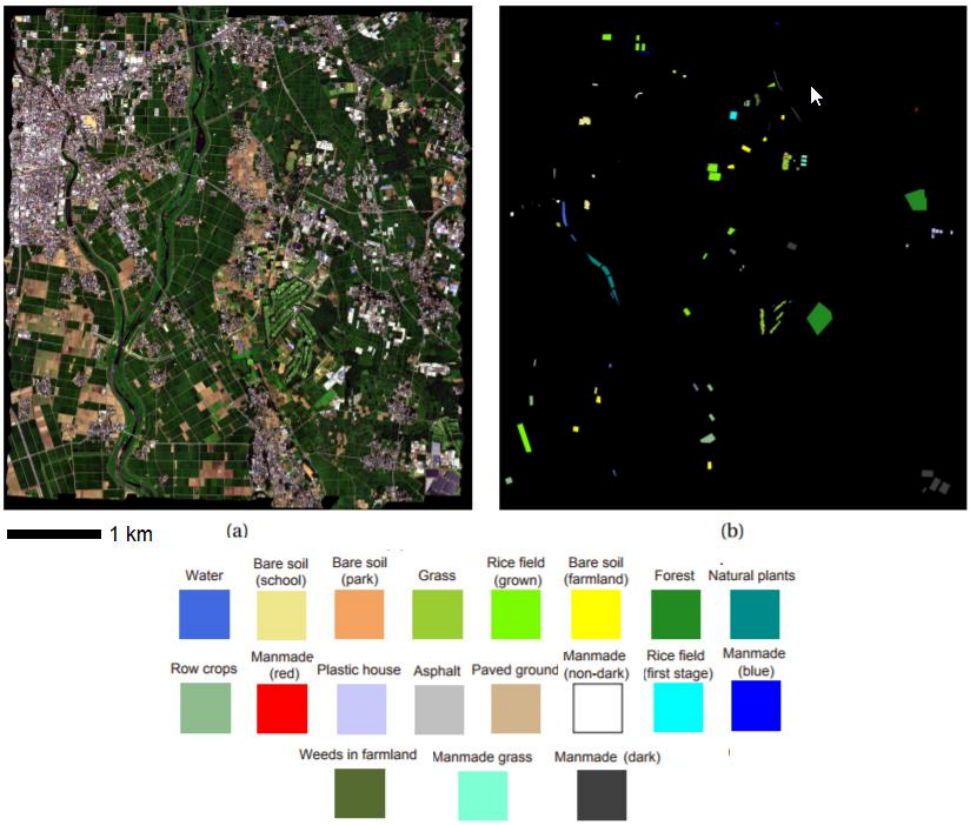

Figure 1. (a) Hyperspectral image of the Chikusei dataset (b) the ground truth of the Chikuseidataset.

2.2. Proposed Support vector machine for HSI Classification

There are numerous artificial intelligence factors based on supervised learning algorithms which can be applied to detect acoustic signals. An SVM (linear and RBF) classifier is applied in the current detection framework, due to its outstanding generalization capability and reputation in the training data set to achieve high accuracy. This method is based on statistical learning theory and structural risk minimization principle.

The strategy of this classifier is to find an optimal separating hyperplane with the maximum margin between the classes by focusing on the training samples located at the edge of the class distribution.

The formula for the output of a linear SVM is:

$$u = w \cdot x - b \tag{1}$$

Where w is the normal vector to the hyperplane and x is the input vector. The separating hyperplane is the plane u=0. The nearest points lie on the planes u = ±1. The margin m is thus

$$m = \frac{1}{\|w\|_2} \tag{2}$$

RBF (Gaussian) kernels are a family of kernels where a distance measure is smoothed by a radial function (exponential function). This kernel nonlinearly maps samples into a higher dimensional space so it, unlike the linear kernel, can handle the case when the relation between class labels and attributes is nonlinear.

Furthermore, the linear kernel is a special case of RBF since the linear kernel with a penalty parameter C has the same performance as the RBF kernel with some parameters (C, Gamma).

$$K\left(x_i\ , x_j\right) = \exp\left(-\sigma\left\|x_i - x_j\right\|^2\right), \sigma > 0 \tag{3}$$

The adjustable parameter σ plays a major role in the performance of the kernel, and should be carefully tuned. If overestimated, the exponential will behave almost linearly and the higher dimensional projection will start to lose its non-linear power.

In the other hand, if underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in training data. Thus, the behaviour of SVM depends on the choice of the width parameter σ.

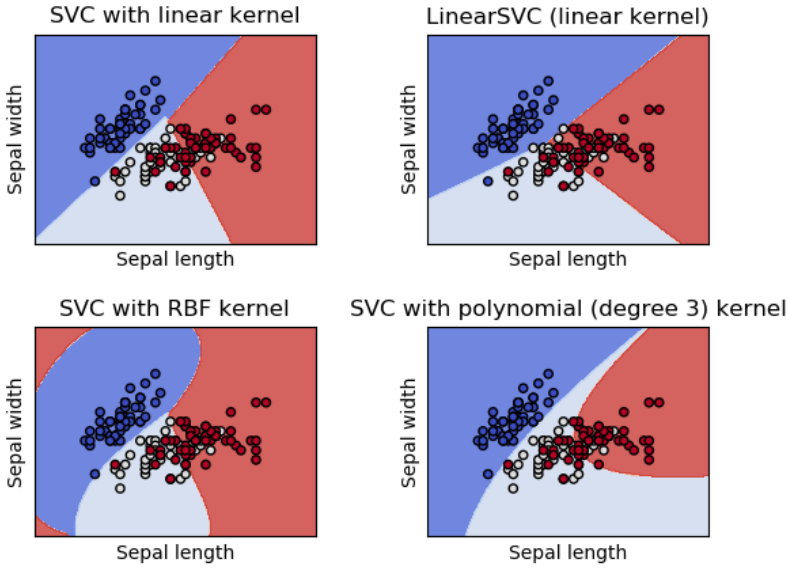

Figure 2. Basic of SVM

2.3. Proposed Convolutional Neural Network for HSI Classification

Previously proposed models for HSI classification can be categorised as pixel-based or patch-based. The former models only the spectral information of the data to perform classification. By contrast, the latter models spectral and spatial (i.e., neighbourhood pixels) information to boost classification performance. In some cases, the spectral profile of individual pixels and spatial features extracted from patches are treated jointly to improve classification accuracies. However, although the mentioned models can efficiently extract low- and high-level features from the spectral and spatial domains, they may fail to model contextual relationships among objects present in a scene accurately. On the other hand, the multi-resolution analysis allows extracting additional contextual features that may boost classification accuracies, such that neither the traditional pixel-based model nor patch-based model can be limited. Therefore, in this research, a convolutional neural network (CNN) for HSI classification is proposed. The proposed classification model in this study is based on CNN spectral features and is abbreviated as CNN. By combining spectral features, the fusion feature is formulated as:

$$S_F = \oplus \left[S_{f1}, S_{f2}, \ldots, S_{fn}\right] \quad (4)$$

Where S_F is the fusion feature vector; S_{f1}, $S_{f2}$and S_{fn} are the features of different scales; n is the number of scales; and $\oplus$ is the concatenation operator.

The flowchart of the proposed CNN is shown in Fig. 2. First, the spectral dimensions of the input HSI data are reduced to cr using a statistical PCA method. Second, the reduced spectral dimension of the HSI is used to generate CNN image patches of sizes $(S_i \times S_i)$, where i ranges from 1 to n. The number of scales (n) and the sizes of the image patches are determined through experiments with the training dataset. However, they can be adjusted in accordance with need. Third, the CNN model is

trained to extract high-level features from each patch collection, and the extracted features are then flattened to produce three one-dimensional (1D) feature vectors. After that, the 1D-feature vectors are combined and used as input for the fully connected network. A software classifier is added on the top of the two dense layers to estimate the probability of each label and predict the label of each pixel.

$$P(Y = i|V, W, b) = s(WV + b) = \frac{e^{W_i V + b_i}}{\sum_j e^{W_j V + b_j}} \quad (5)$$

where V is the combined spectral–spatial feature.W and b denote weight and bias, respectively. $W_i$and $b_i$are the weight and bias connecting the output neuron of class i, respectively. Y and$s$ are the classification result and the softmax function, respectively. The neuron number of the output layer is the total number of classes (Yue et al., 2016).

The CNN model extracts spectral features at multiple scales from an input image; abstracts the features by specific CNN models; and concatenates CNN outputs into a single feature vector, which is used as an input for a dense layer (Fig. 3). For the input hyperspectral image (i.e., after PCA transformation), the neighbourhood pixels around a central pixel i are extracted at n scales. For example, for a $3 \times 3 \times cr$ configuration, CNN extracts 8 neighborhood pixels for the ith pixel, and if $cr = 80$, then the size of the output feature vector (S_n) will be $8 \times 80 = (640,)$. Then, specific CNN models are applied at each scale to extract high-level features that are more abstract thanS_n. The size of the CNN output feature vector (S_f) depends on CNN architecture, specifically, the size of the spatial filters and the pooling operations. Finally, the S_f feature vectors are concatenated into a single feature vector (S_F), which then could be used as an input for a fully connected layer to perform classification.

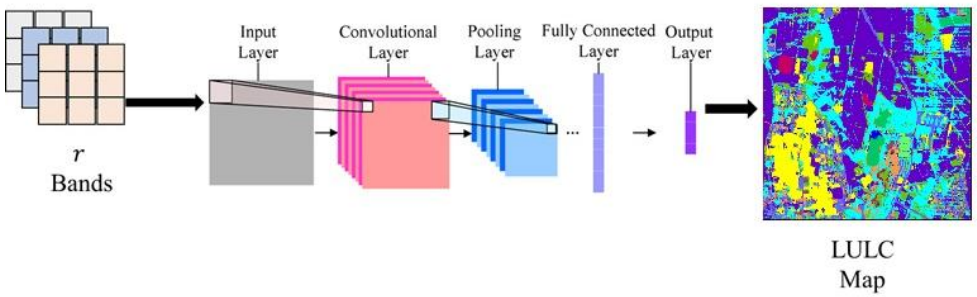

Figure 3. Flowchart of the CNN model and data HSI classification

3. Results and discussion

3.1. Results of SVM and CNN

In SVM linear after training and testing the accuracy, classification the data had a problem of hardware capacity to process it. To solve it, the image were divided into 5 patches. Each patch, was classified .When using SVM_RBF, the computer hanged from processing. It was assumed that the data was huge and required sub setting. This was done, where algorithm managed to process it . However, the algorithm RBF issued a warning and requested to Scale the Data. The scaling change the numerical presentation of data to fit within a Range. The RBF give two options for scaling, (Standard and Minimax). Here, Minimax was applied .The occurred changes of sub setting image data in RBF required to apply the same in Linear. The CNN models that perform spatial feature extraction at different scales.

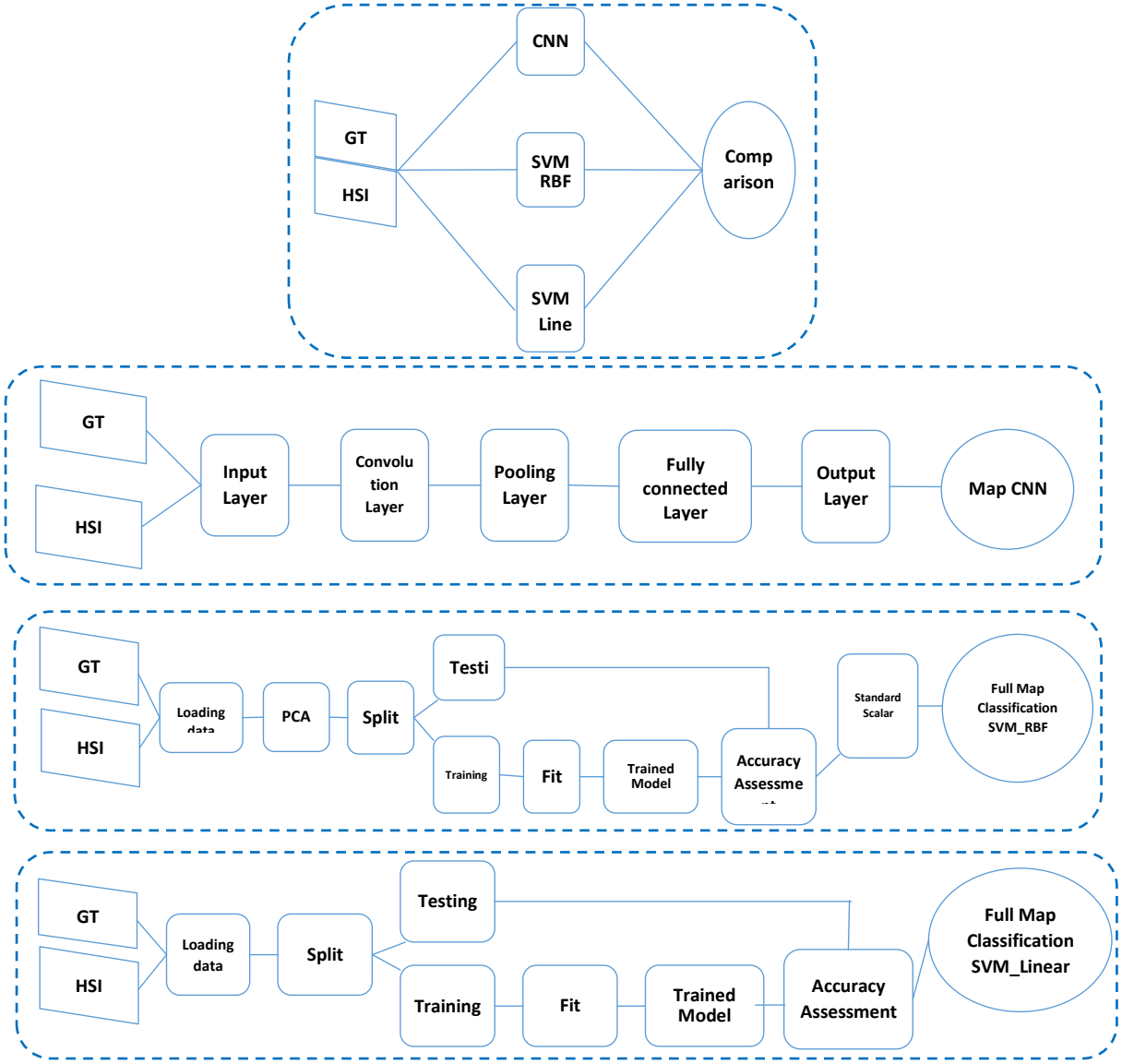

Figure 4. Flowchart Methodology

The input shape of the CNN models depended on the size, spectral band number, and window size of the input dataset. For example, the input shape was (80, 5, 5) for the first CNN model for the Chikusei dataset. Here, the first number was the number of spectral bands of the data, and the second and third numbers indicate the width and the height of the window size, respectively, used to extract neighbouring pixels. Each CNN model had three two-dimensional convolutional layers with some spatial filters of 128, 64, and 32. The spatial filters had a size of 3 × 3. The activation functions in these layers were ReLU. After each convolutional layer, a dropout layer with a threshold of 0.3 was added to reduce model complexity and prevent over fitting. The convolutional layers were followed by a max pooling layer with a pool size of (1 × 1) and a batch normalisation layer. The outputs of the CNN models were flattened, and concatenation merged their results into a single feature vector. Subsequently, two dense layers with 32 neurons each were added, followed by a soft max layer that classified features and predicted the label of the image pixels.

The proposed CNN model can handle multiple scales with different window sizes for each level. The best accuracies were achieved using three scales with sizes of 3 × 3, 5 × 5. The use of a single scale with CNN did not allow the use of the spatial structure of the data and the detection of objects of different scales. The experiments showed that the use of two scales could improve classification accuracies; however, the model does not generalise well to unseen samples. Therefore, the use of at least three scales to achieve high classification accuracies for training and testing samples is suggested. The current work did not investigate a high number of scales because such an investigation requires extensive computing power. Moreover, high numbers of scales are not suggested unless acceptable classification accuracies could not be achieved with the three scales.

3.2. Classification Accuracy of SVM and CNN

In this study, it is shown that SVM overcomes CNN, where it gives best results in classification, the accuracy in PCA- band the SVM linear 97.44%, SVM-RBF 98.84% and the CNN 94.01%, But in the all bands just have accuracy for SVM-linear 96.35% due to the big data hyperspectral image (HIS) The computer was unable to process during running SVM-RBF due to hardware and used algorithm. The RBF classification found to be less accurate compared SVM-Linear. This is due to RBF parameters. Though three of its parameters were optimized for the dataset (C-supports, gamma and Maximum iterations) using Grid Search, these techniques depend on defined ranges of possible values for each parameter.

The same optimization applied for the SVM-Linear. The accuracy assessment do not provide truth compared to Visual interpretation of the overall classification of each image.

3.3. Comparison between SVMs and CNN Methods

The classification performance of the CNN model was compared with those of other methods, such as radial basis function SVM (SVM-linear, SVM-RBF), pixel-based CNN, pixel-pair CNN, and SSDL. The overall accuracies of the methods indicated that the proposed model outperforms all the other methods in the classification of Chikusei datasets. The results shown in Table V demonstrated that CNN achieves the highest classification accuracy (97.44%, 98.72% and 94.01%) for both datasets. The SVM-linear, SVM-RBF and pixel-based CNN models do not utilise the rich spatial information contained by the datasets and instead rely solely on the spectral signature of each image pixel.

The utilisation of additional spatial (from neighbouring pixels) features helps improve the classification accuracies of the methods in land-cover classification. Given that spatial information can improve the classification accuracy, CNN utilises rich spatial features at multiple scales to describe the spatial structure of the data for classifying each pixel in the image. The rich complementary spatial details obtained at different scales helps boost the performance of the proposed classification method.

Table 1. Comparison Accuracy of HSI Data

	Accuracy %		
Hyper data	SVM_Linear	SVM_RBF	CNN
PCA-band	97.44%	98.84%	94.01%

* The computer was unable to process during running SVM-RBF due to hardware and used algorithm.
** Same as RBF the computer was unable to process during running CNN due to hardware and used algorithm.

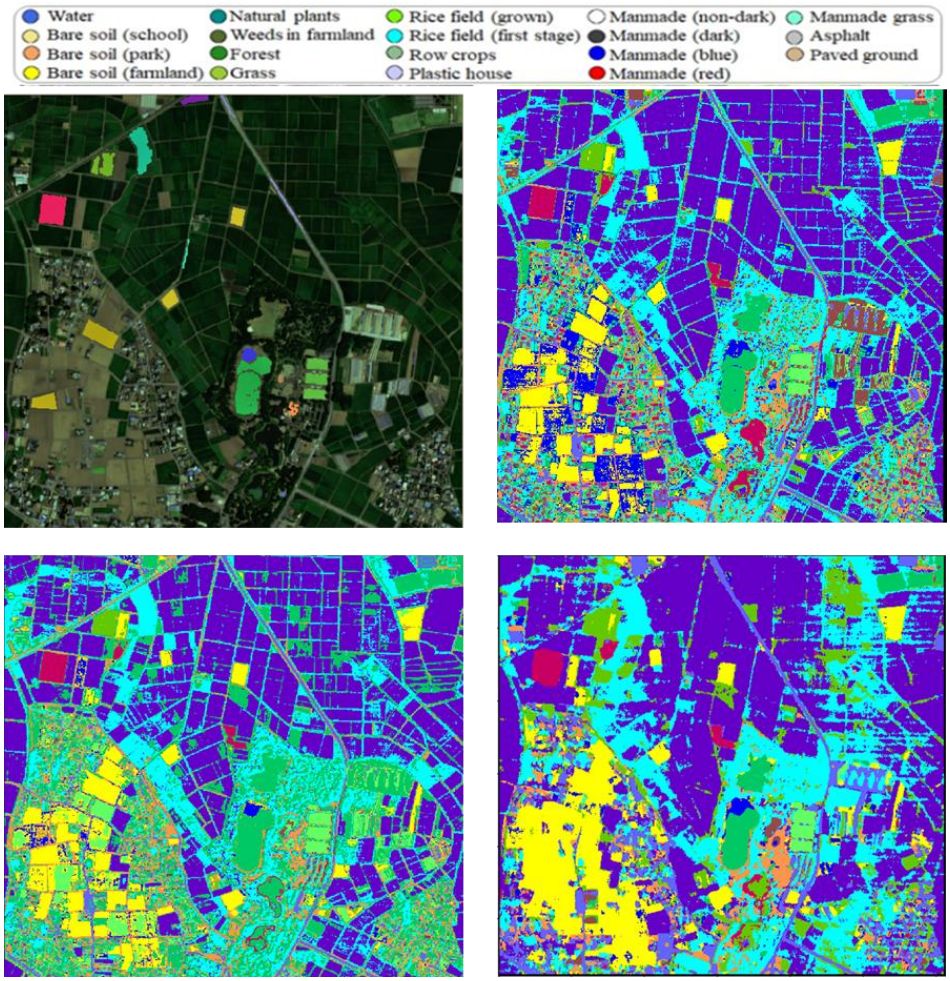

Figure 5. HSI classification for SVM_Linear, SVM_RBF and CNN

Table 2. Comparison of the performances of SVM-Linear, SVM-RBF, CNN model and other state-of-the-art methods.

Data set	SVM-Linear	SVM-RBF	CNN
Indian Pines	88.55%	87.60%	90.16%
Salinas	90.59%	91.66%	92.60%
U Pavia	92.70%	90.52%	98.72%
Chikusei	97.44%	98.84%	94.01%

4. Conclusions

In SVM linear after training and testing the accuracy, classification the data had a problem of hardware capacity to process it. To solve it, the image were divided into 5 patches. Each patch, was classified.

When using SVM_RBF, The computer was unable to process during running SVM RBF. It was assumed that the data was unable and required sub setting. This was done, where algorithm managed to process it. However, the algorithm RBF issued a warning and requested to Scale the Data. The scaling change the numerical presentation of data to fit within a Range. The RBF give two options for scaling, (Standard and Minimax) which in here, Minimax was applied.

The occurred changes of sub setting image data in RBF required to apply the same in Linear. Although, RBF is well known of its accuracy compared with Linear Algorithm, Linear found to more tolerant with big imageries; as well simplicity to apply compared RBF. The RBF classification found to be less accurate compared SVM Linear. This is due to RBF parameters. Though three of its parameters were optimized for the dataset (C-supports, gamma and Maximum iterations) using Grid Search, these techniques depend on defined ranges of possible values for each parameter. The same optimization applied for the SVM Linear.

The accuracy assessment do not provide truth compared to Visual interpretation of the overall classification of each image. The CNN approaches of classification requires to define a Deep Neural network Model. This model defined as simple model to be comparable with SVM. Though, how CNN works require preparing data to be processed by kernel window. In this study, the Kernel window defined 5 x 5 pixel based previous studies.

The CNN required to use PCA to reduce the dimensions of the data that is reducing noisy bands. However, the approach required augmentation to increase the dataset training samples.

Though the CNN accuracy is 94.01%, the visual interpretation contradict such accuracy, where SVM classifiers have shown better accuracy performance.

Acknowledgement
The authors gratefully acknowledge Space Application Laboratory, Department of Advanced Interdisciplinary Studies, and the University of Tokyo for providing the Chikusei hyperspectral data.

References

[1] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, pp. 879-893, 2012.

[2] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple Kernel Learning for Hyperspectral Image Classification: A Review," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 6547-6565, 2017.

[3] E. Aptoula, M. C. Ozdemir, and B. Yanikoglu, "Deep learning with attribute profiles for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 1970-1974, 2016.

[4] J. Xia, P. Ghamisi, N. Yokoya, and A. Iwasaki, "Random Forest Ensembles and Extended Multiextinction Profiles for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2017.

[5] E. Li, P. Du, A. Samat, Y. Meng, and M. Che, "Mid-level feature representation via sparse autoencoder for remotely sensed scene classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, pp. 1068-1081, 2017.

[6] R. Kemker and C. Kanan, "Self-taught feature learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 2693-2705, 2017.

[7] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework," *IEEE Transactions on Geoscience and Remote Sensing*, 2017.

[8] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, pp. 4544-4554, 2016.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[10] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, 2013, pp. 6645-6649.

[11] R. Sarikaya, G. E. Hinton, and A. Deoras, "Application of deep belief networks for natural language understanding," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, pp. 778-784, 2014.
[12] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, 2015, pp. 4959-4962.
[13] X. Ma, H. Wang, and J. Wang, "Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 120, pp. 99-107, 2016.
[14] P. Ghamisi, Y. Chen, and X. X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 1537-1541, 2016.
[15] 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).
[16] Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In Acoustics, speech and signal processing (icassp), 2013 ieee international conference on (pp. 6645-6649). IEEE.
[17] Makantasis, K., Karantzalos, K., Doulamis, A., & Doulamis, N. (2015, July). Deep supervised learning for hyperspectral data classification through convolutional neural networks. In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International (pp. 4959-4962). IEEE.
[18] Ma, X., Wang, H., & Wang, J. (2016). Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning. ISPRS Journal of Photogrammetry and Remote Sensing, 120, 99-107.
[19] Kemker, R., & Kanan, C. (2017). Self-taught feature learning for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 55(5), 2693-2705.
[20] 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, 54(8), 4544-4554.
[21] Ghamisi, P., Chen, Y., & Zhu, X. X. (2016). A self-improving convolution neural network for the classification of hyperspectral data. IEEE Geoscience and Remote Sensing Letters, 13(10), 1537-1541.