9th International Conference on Information Science and Technology (ICIST), Hulunbuir, Inner Mongolia, China; August 2-5, 2019

A Data Augmentation-Assisted Deep Learning Model for High Dimensional and Highly Imbalanced Hyperspectral Imaging Data

[1]Juan F. Ramirez Rochac, [2]Nian Zhang, and [3]Lara Thompson
[1]Dept. of Computer Science & Information Technology, [2]Dept. of Electrical and Computer Engineering, and [3]Dept. of Mechanical Engineering
University of the District of Columbia,
Washington, D.C., USA 20008
Email: jrochac@udc.edu, nzhang@udc.edu, lara.thompson@udc.edu

[4]Timothy Oladunni
Dept. of Computer Science & Information Technology
University of the District of Columbia
Washington, D.C., USA 20008
Email: timothy.oladunni@udc.edu

***Abstract*—Recent advances in remote sensing technologies have led to the fast proliferation of massive and often imbalanced datasets. Direct classification in these datasets becomes difficult, because of the high dimensionality, and the fact that minority classes are overlapped and dwarfed by majority classes. Deep learning is the state-of-the-art in image classification, with applications in face- and text detection, text recognition, as well as voice classification. However, deep learning requires a favorable ratio between dimensionality and sample size. To address high dimensional yet imbalanced datasets, in this paper, we propose the integration of data augmentation, to a deep learning classifier of a high dimensional and highly imbalanced photo-thermal infrared hyperspectral dataset of chemical substances. First, we apply a basic deep machine learning approach using a convolutional neural network (CNN) on the original dataset. Second, we apply principal component analysis (PCA) to reduce dimensionality before applying CNN. Third, we prepend an offline data augmentation step to increase dataset size before applying CNN. After that, we evaluate the performance by calculating the probability of detection (POD), and recall based on true positive (TP), false negative (FN), false positive (FP), and true negative (TN).**

***Keywords*—deep learning; convolutional neural networks, hyperspectral classification; principal component analysis; data augmentation**

I. INTRODUCTION

Machine learning is among the most frequent topics of research in recent publications. Image-based machine learning and deep learning in particular has recently shown expert-level accuracy in medical image classification. In [1], deep convolutional neural networks (CNNs) show potential for classification of skin cancer with a level of competence comparable to dermatologists. Moreover, [2] presented a deep learning algorithm for detecting referable diabetic retinopathy.

It is evident that deep learning is the current state-of-the-art classification technique with applications ranging from medicine to remote sensing to text and voice recognition [3]. In view of numerous feature selection approaches in hyperspectral imagery classification [4]-[12], applicability of deep machine learning algorithms to high dimensional and highly imbalance hyperspectral images in classification remains an unexplored field. There are two ways to reduce this unfavorable ratio between dimensionality and imbalanced datasets: 1) reduce the number of dimensions or features and 2) increase the number of samples. Regarding point 1, above, dimensionality reduction presents a possible solution to reduce the high number of feature vectors. To this end, Principal Component Analysis (PCA) is one of the most commonplace tools [13]. Moreover, in [12] the authors present yet another approach to deal with the imbalanced hyperspectral datasets by adaptively extracting features based on their local and global characteristics. On point 2 above, Data Augmentation presents a well-known technique to expand the number of training samples from the limited number of labelled samples [14]. It has been implemented successfully in many fields from image processing [15] to molecular modelling [16]. Data augmentation thus proves a promising option to deal with high dimensional and highly-imbalanced datasets in deep learning models. The goal of this paper is to explore Data Augmentation before Deep Convolutional Neural Network and to evaluate its classification accuracy.

The remainder of this paper is organized as follows. In Section 2, the high dimensional and highly imbalanced data set is described. In Section 3, the proposed methodology is presented. Principal component analysis (PCA), Data augmentation (DA) and Convolutional Neural Network (CNN) are described. In Section 4, the analysis and results of are presented. In Section 5, the conclusions are given.

978-1-7281-2106-2/19/$31.00 @2019 IEEE

II. DATASET

In this paper, an experimental dataset is used to evaluate the proposed approach. This dataset was initially presented in [17], which was acquired by a Photo-Thermal Infrared Imaging Spectroscopy (PT-IRIS) technique which consists of two main components: 1) A quantum cascade laser that causes the substance or analyte to be illuminated and 2) an infrared camera that collects hundreds of images corresponding to different wavelength channels or intensities of light. We called this large collection of images a hyperspectral cube because it basically gathers intensities from across the entire electromagnetic spectrum [18]. Certain analytes and substrates leave a unique fingerprint or spectral signature, which uniquely characterizes the underlying material or element.

When we are in the presence of pure pixels, (i.e., pixels with one single underlying material), classification of the material is achieved by simply matching their signatures. But when the resolution of the infrared camera is not fine enough to separate different substances, one single pixel can capture two or more signatures, and the resulting spectra will be a composite or mixture of their individual pure signatures. The rows are the features and the columns are the instances or samples. It contains a total of eight overlapped of classes, four chemical analytes and four substrates. Since our main objective is to detect analytes, we are going to group all classes into two main groups. Hence our problem becomes a binary classification problem. More details about this dataset can be found in [6]. This hyperspectral dataset showcases high dimensionality and a highly imbalanced sample size.

III. METHODOLOGY

The methodology is divided into four parts: Principal Component Analysis (PCA), Data Augmentation (DA), Convolutional Neural Network (CNN) and performance analysis. The main goal of our experiments is to comprehensively measure classification accuracy for CNN, PCA+CNN and DA+CNN. In this section we will describe these approaches.

A. Principal Component Analysis

Principal Component Analysis (PCA) has been widely used to translate highly-dimensional data into a new feature space with lower dimensionality [13]. The goal of this translation is to reduce the number of dimensions (or feature vectors) while maintaining all relevant information and variability of the data [19]. The first principal component, PCA1, aims to capture the majority of the variability of the data. The second principal component, PCA2, aims to capture the majority of the remaining variability of the data. PCA2, however, has to be orthogonal to PCA1. Likewise, the third principal component, PCA3, has to be orthogonal to both PCA1 and PCA2 while further capturing the majority of any remaining variability of the data. One of the appeals of PCA is that it tends to find the most significant patterns in the data. Another very appealing characteristic of PCA is that it often results in a small number of dimensions. Thirdly, PCA's transformation also brings yet another benefit: it removes some of the noise from the data [20]. Mathematically, we start by defining the dataset as a matrix $\mathbf{D}[n][m]$, where n is the number of samples and m is the number of features. Then we define the covariance **Cov** as the expected value of $d_i * d_j$ where the mean is equal to zero. The result is $\mathbf{R}[n][k]$, where k is the number of new features and $k < m$ and usually $k < n$.

B. Data Augmentation

Data augmentation has been shown to produce promising ways to increase the accuracy of image classification tasks [15]. It also provides a solution to small datasets by generating derived samples using different mathematical transformations. In image classification, common transformations for data augmentation consist of using a combination of mathematical manipulations to modify the training data [62]. For each image, we generate a derived image that is stretched shrunk, zoomed in/out, flipped, rotated, or colored with a different intensity. The two images (original and derived) are fed into the classification model. Basic transformations are depicted in Fig. 1.

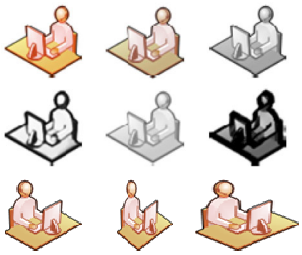

Fig. 1. Data Augmentation: Basic Transformations

However, these transformations cannot be directly applied to spectral images [21], because any stretching, shrinking, or flipping of the original spectra image would result in a spectral or signature that corresponds to a different element or material. Our proposed data augmentation technique is therefore to randomly add Gaussian noise to the spectral dimension. The motivation for this technique comes from signal processing and digital communications domains. Optimal detection under this condition has been proved in the field of wireless communications [22]. In any advanced communication system the received signal is equal to the sent or original signal plus a Gaussian white noise as shown in Equation 1. Both the received signal $x(t)$ and the original signal $y(t)$ are considered random variables and the white Gaussian noise $n(t)$ follows a Gaussian (Normal) distribution with a probability density function presented in Equation 2. Our white noise will follow a Gaussian Normal distribution $N(\mu,\sigma)$ where mu is zero and sigma is 0.0001.

$$x(t) = y(t) + n(t) \quad (1)$$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

C. Convolutional Neural Networks

Convolutional Neural Networks (CNN) is a type of deep machine learning in which a model learns to perform classification tasks directly from the data. The term "deep" refers to the number of layers, hence, the more layers, the deeper learning. Image-based machine learning and deep

learning in particular has recently shown expert-level accuracy in medical image classification [18]. Figure 2 details the layers of our CNN.

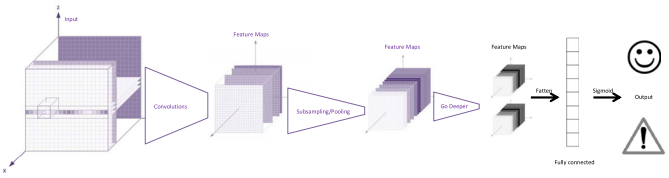

Fig. 2. Deep CNN for Binary Classification

In the experiments, a sequential CNN classifier was used with the following layers:

- Conv2D(75, (1,15), input_shape=(1,1254,1))
- LeakyReLU(alpha=0.1)
- MaxPooling2D(pool_size=(1,6), padding='same')
- Conv2D(150, kernel_size=(1,15), activation='linear')
- LeakyReLU(alpha=0.1)
- MaxPooling2D(pool_size=(1,6), padding='same')
- Conv2D(150, kernel_size=(1,15), activation='linear')
- LeakyReLU(alpha=0.1)
- MaxPooling2D(pool_size=(1,6), padding='same')
- Flatten()
- Dense(300, activation='linear'))
- LeakyReLU(alpha=0.1)
- Dense(units = 1, activation='sigmoid')

D. *Performance Analysis and Tabular Comparison*

To evaluate the performance and compare their accuracy of PCA+CNN versus DA+CNN, we employed the confusion matrix depict in Table I and the classification report to calculate their accuracy.

TABLE I. Confusion Matrix

	Predicted Class	
True Class	***Danger***	***No Danger***
Danger	TP: True Positive	FN: False Negative
No Danger	FP: False Positive	TN: True Negative

As per Table I, we obtain the values of TP, FN, FP and TN, and calculate Error rate, Recall, Sensitivity or Hit rate, Precision, Specificity and False alarm rate. The equations used to compare all approaches are as follows:

- *Probability of Detection* is equal to true positives over (true positives plus false negatives).
- *False Alarm Rate* is equal to false positives over (false positives plus true negatives).
- *Precision* is true positives over (true positives plus false positives).
- *Recall* is true positives over (true positives plus false negatives).
- *Accuracy* is equal to (true positives plus true negatives) over (true positives plus false positives plus false negatives plus true negatives).

K-fold cross validation is also used. Figure 3 is a representation of the data partition for 10-fold cross validation, where 10 % of the total data is used for testing and 90% of the data is used for training the model.

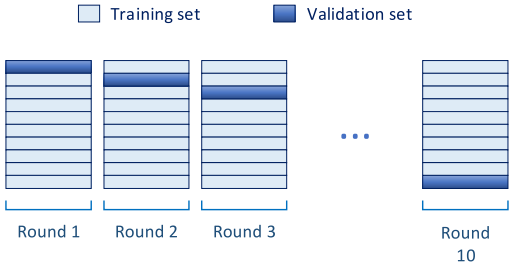

Fig. 3. K-fold Cross Validation Partitions

IV. Experimental Results

In this section all experimental results are presented. We start with a basic convolutional neural network CNN model, then we present the results of using a principal component analysis PCA together with CNN, and then we present the results of our data augmentation-assisted convolutional neural network (DA+CNN) model. We end this section with a tabular comparison of all three models.

A. *Basic Convolutional Neural Network (CNN)*

During this round of experiments all 1254 features were used. The 123 samples were divided into three sets. For testing 20%, about 25 samples, were used. The remaining 80% was further divided into 60% for training data, about 74 samples and 20% for validation data, about 24 samples, as shown in Fig. 6. From the confusion matrix presented in Table II, we can say that the overall accuracy of CNN on the testing data was 84%. This accuracy was calculated by adding the number of correctly predicted classes (19+2) and dividing it by the total number of testing samples (25).

TABLE II. Confusion Matrix for CNN

	Predicted Class	
True Class	***Detected***	***Not Detected***
Detected	19	0
Not Detected	4	2

From the classification report presented in Table III, we see that the average precision, recall and f1-scores are 84%. Here precision is a measurement of how often a data sample that was predicted as positive is actually positive. This is a ratio between the number of correctly predicted positive instances and the number of instances predicted as positive instances. Recall is a measurement of how often a positive class instance in the data set was predicted as a positive class instance by the model. This is a ratio between the number of correctly predicted positive instances and the number of truly positive instances. The f1-scores measures the accuracy of the classifier in classifying the data points in a particular class compared to all other classes. The support is the number of samples of the true response that lie in that class.

TABLE III. Classification Report for CNN

	Precision	Recall	F1-score	Support
0	0.83	1.00	0.90	19
1	1.00	0.33	0.50	6
micro avg	0.84	0.84	0.84	25
macro avg	0.91	0.67	0.70	25
Weighted Mean/Sum	0.87	0.84	0.81	25

Figure 4 depicts the loss and Figure 5 shows the accuracy of the CNN over 50 epochs. The figures present the results for training data in blue, the results for validation data in orange and the results for testing in green color. In Fig. 4, we observe that the loss of CNN decreases as the model learns. However, it reaches a plateau after 50 epochs.

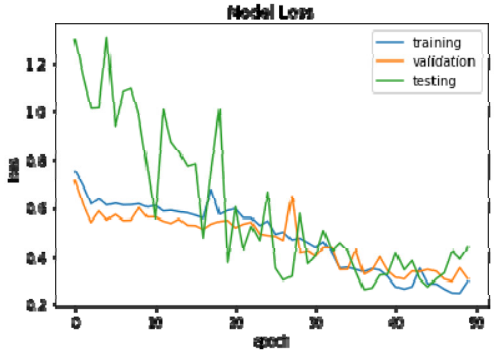

Fig. 4. Model Loss of CNN

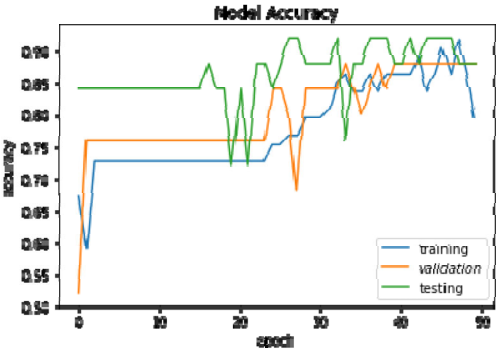

Fig. 5. Model Accuracy of CNN at 50 Epoch

In Fig. 5, we observe that the accuracy of CNN increases as the model learns from the training set. This is typical of any deep learning model. Our CNN model reaches a top precision of 91% on validation data. But it converges at an average value of 84% accuracy on testing data.

B. Principal Component Analysis (PCA+CNN)

In this variant, during the preprocessing stage, Principal Component Analysis (PCA) was applied. Since Scree Plots are an easy way to depict the ratio of the total variance in the data points in terms of principal components, a Scree Plot was employed to determine the number of components to use. From the Scree Plot in Fig. 6, we observed that 16 principal components were enough to represent over 99.99% of the data variability.

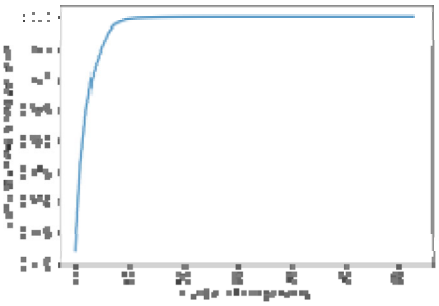

Fig. 6. Scree Plot

During this round of experiments all 16 PCA were used. Similar to the previous round of experiments the 123 samples were divided into three sets, 60% for training, 20% for validation and 20% for testing, 20%. For this variant, CNN was adapted to be able to receive a 16-feature input. No other changes were applied to the original CNN. After running the experiments for PCA+CNN and compiling all cross validation results into a tabular form, the average accuracy of PCA+CNN on our imbalanced dataset was 80%. This accuracy was calculated using the confusion matrix presented in Table IV by adding the number of correctly predicted classes (18+2) and dividing it by the total number of testing samples (25).

TABLE IV. Confusion Matrix for PCA+CNN

	Predicted Class	
True Class	***Detected***	***Not Detected***
Detected	18	1
Not Detected	4	2

From the classification report presented in Table V, we see that the average precision, recall and f1-scores are 80%. In adding PCA to our basic CNN, the overall performance decreased by 4%. PCA usually performs well in case of high dimensionality. In the presence of imbalanced datasets however, it suffers from the lack of data samples. Deep learning approaches greatly depend on training/learning from massive amount of data samples. PCA+CNN on an imbalanced hyperspectral dataset only achieved average accuracy, which is not desirable for any critical decision making. Since in this paper, we are targeting highly dimensional and imbalanced datasets, the results presented in Table V suggest that PCA+CNN is not the optimal approach.

TABLE V. Classification Report for PCA+CNN

	Precision	Recall	F1-score	Support
0	0.82	0.95	0.88	19
1	0.67	0.33	0.50	6
micro avg	0.80	0.80	0.80	25
macro avg	0.74	0.64	0.66	25
Weighted Mean/Sum	0.78	0.80	0.77	25

Figures 7 and 8 further show the implications of having imbalanced data in terms of model's loss and the model's accuracy over 50 epochs correspondingly. The figures present the results for training data in blue, the results for validation data in orange and the results for testing in green color.

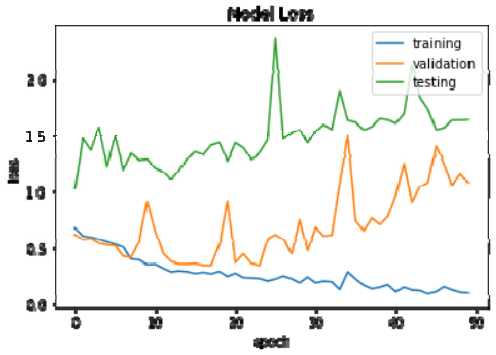

Fig 7. Model Loss of PCA+CNN

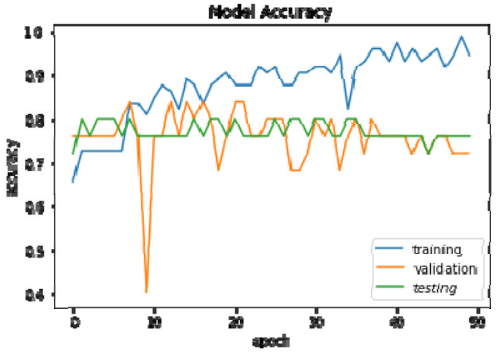

Fig. 8. Model Accuracy of CNN at 50 Epoch

In Fig. 7, we observe that the loss of PCA+CNN decreases as the model learns. However, there is performance degradation on validation and testing data. In Fig. 8, we observe that the accuracy of PCA+CNN increases as the model learns from the training set. But it quickly levels out at 76% on validation data and testing data.

C. Data Augmentation (DA+CNN)

In this variant, during the preprocessing stage, offline Data Augmentation was applied. We achieve this by adding a small random white noise to our hyperspectral dataset. To generate multiple noisy additional samples, we add a Gaussian Normal (mean=0, variance=0.0001) to the spectral features. The idea is to add a small distortion which will help us train our deep learning model with a larger number of samples (original samples + noisy samples) and at the same time it will produce a more robust classifier. The following tables and figures present the experimental results of DA+CNN.

The confusion matrix presented in Table VI, suggests that the average accuracy of DA+CNN on our highly-imbalanced dataset is 90% on testing set. This accuracy was calculated by adding the number of correctly predicted classes (24+4) and dividing it by the total number of testing samples (31). Note that in this case there are more testing instances due to the fact that the original dataset was augmented.

TABLE VI. Confusion Matrix for DA+CNN

	Predicted Class	
True Class	***Detected***	***Not Detected***
Detected	24	0
Not Detected	3	4

In Table VII, the classification report for DA+CNN is presented, we observe that the average precision, recall and f1-scores are 94%., 79% and 83%. After the incorporation of our data augmentation approach to CNN, the overall performance in terms of accuracy increased by 10%. Note that traditional data augmentation techniques, shown in Fig. 4, were not used here since they cannot be directly applied to spectral images. Data augmentation usually improves resistance to noise. And more importantly to our research, offline data augmentation can also be used to increase the sample size. Deep learning approaches can favorably benefit from more data. Our results indicate that in fact DA+CNN outperform both basic CNN and PCA+CNN.

TABLE VII. Classification Report for DA+CNN

	Precision	**Recall**	**F1-score**	**Support**
0	0.89	1.00	0.90	24
1	1.00	0.57	0.73	7
micro avg	0.90	0.90	0.90	31
macro avg	0.94	0.79	0.83	31
Weighted Mean/Sum	0.91	0.90	0.89	31

Figure 9 and Figure 10 depict the DA+CNN's loss and accuracy over 20 epochs correspondingly. The figures present the results for training data in blue, the results for validation data in orange color. In Fig. 9, we observe that the loss of DA+CNN decreases as the model learns more and more from the noisy data. Also, there is performance improvement on validation data. In Fig. 10, we observe that the accuracy of DA+CNN increases as the model learns from the training set. After 15 epochs it slowly improves and by 20 epochs it reaches a top accuracy of 93% and an average of 90%on testing data. Note that both figures follow a logarithm curve.

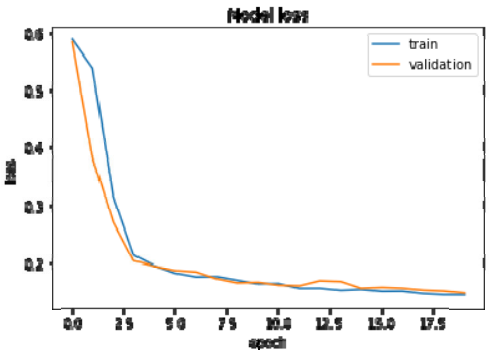

Fig. 9. Model Loss of DA+CNN

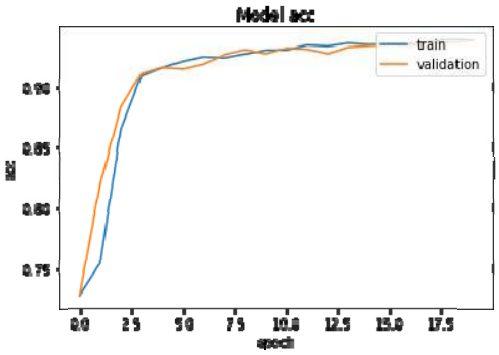

Fig 10. Model Accuracy of DA+CNN at 20 Epoch

By looking at these curves, we observe that DA+CNN quickly reaches and passes the results of CNN and PCA+CNN. Another observation is that the curves for DA+CNN, as is shown in Fig. 9 and Fig. 10, are significantly smoother and less jerky than the curves for CNN and PCA+CNN. This suggests that DA+CNN is a more robust against noise since it learned from noisy data.

CONCLUSIONS

Data augmentation-assisted deep learning using a convolutional neural network (DA+CNN) was explored on a hyperspectral dataset. Both PCA and Data augmentation methods were used to preprocess classification input and predict with a comparable degree of accuracy. Initially, a three-layer CNN was designed and performance was measured in terms of accuracy. Then this was compared to output generated with PCA together with CNN, or DA together with CNN. For PCA+CNN, the overall performance decreased by 4% as compared to CNN alone. For DA+CNN, the overall performance in terms of accuracy increased by 10%. In this last approach, the hyperspectral images were preprocessed using offline data augmentation and it was analyzed with a confusion matrix and the average accuracy (90%) was comparable to results in [6]. In the future, deep learning and hyperspectral images should be further explored paying particular attention to online data augmentation, spectral adaptive filtering and recurrent neural networks.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation (NSF) grants, HRD #1505509, HRD #1533479, and DUE #1654474, and Department of Defense (DoD) grant #W911NF1810475.

REFERENCES

[1] A. Esteva, et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017.

[2] V. Gulshan, et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," vol. 304, pp. 649–656, 2016.

[3] J. Schmidhuber, "Deep learning in neural networks: an overview," Neural Networks, vol. 61, pp. 85–117, 2015.

[4] N. Zhang and K. Leatham, "Feature selection based on SVM in photo-thermal infrared (IR) imaging spectroscopy classification with limited training samples" *WSEAS Transactions on Signal Processing*, vol. 13, 2017, no. 33, pp. 285-292.

[5] N. Zhang and L. A. Thompson, "An intelligent clustering algorithm for high dimensional and highly overlapped photo-thermal infrared imaging data," *Fall 2016 ASEE Mid-Atlantic Regional Conference*, Hofstra University, Hempstead, NY, October 21-22, 2016.

[6] G. Taşkın, H. Kaya, and L. Bruzzone, "Feature selection based on high dimensional model representation for hyperspectral images," in *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2918-2928, June 2017.

[7] J. F. Ramirez Rochac, N. Zhang, and P. Behera, "Design of adaptive feature extraction algorithm based on fuzzy classifier in hyperspectral imagery classification for big data analysis," *The 12th World Congress on Intelligent Control and Automation (WCICA 2016)*, Guilin, China, pp. 1046 – 1051, June 12-15, 2016.

[8] N. Zhang, "Cost-sensitive spectral clustering for photo-thermal infrared imaging data," *2016 Sixth International Conference on Information Science and Technology (ICIST)*, Dalian, pp. 358 – 361, May 6-8, China, 2016.

[9] Q. Zhang, Y. Tian, Y. Yang, and C. Pan, "Automatic spatial–spectral feature selection for hyperspectral image via discriminative sparse multimodal learning," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 261-279, Jan. 2015.

[10] J. F. Ramirez Rochac and N. Zhang, "Reference clusters based feature extraction approach for mixed spectral signatures with dimensionality disparity," *10th Annual IEEE International Systems Conference (IEEE SysCon 2016)*, Orlando, Florida, pp. 1 – 5, April 18-21, 2016.

[11] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1723-1740, March 2019.

[12] Juan F. Ramirez Rochac and Nian Zhang, "Feature extraction in hyperspectral imaging using adaptive feature selection approach," *The Eighth International Conference on Advanced Computational Intelligence (ICACI2016)*, Chiang Mai, Thailand, pp. 36-40, 2016.

[13] I. T. Jolliffe, "Principal component analysis," Springer-Verlag, 2nd Edition, October, 2002.

[14] 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.

[15] J. Wang and L. Perez, "The effectiveness of data augmentation in image classification using deep learning," Cornell University, arXiv.org, 2017.

[16] E. Jannik Bjerrum, "SMILES enumeration as data augmentation for neural network modeling of molecules," Cornell University, arXiv.org, 2017.

[17] R. Furstenberg, C. A. Kendziora, J. Stepnowski, S. V. Stepnowski, M. Rake, M. R. Papantonakis, V. Nguyen, G. K. Hubler, and R. A. McGill; "Stand-off detection of trace explosives by infrared photo-thermal spectroscopy," Applied Physics Letters, December 2008.

[18] D. Landgrebe, "Hyperspectral image data analysis," IEEE Signal Process. Mag., vol. 19, no. 1, pp. 17–28, 2002.

[19] M. A. Carreria-Perpinan, "A review of dimension reduction techniques," Technical Report CS-96-09, Dept. of Computer Science, University of Sheffield, 1997.

[20] I.K. Fodor, "A survey of dimension reduction tehniques", Technical Report UCRL-ID 148494, LLNL June 2002.

[21] E. Jannik Bjerrum, M.Glahder, and T.Skov, "Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics," Cornell University, arXiv.org, 2017.

[22] P. Cotae and S. Yalamanchili, "On the optimized sensor location performances in the presence of additive white Gaussian noise," *2008 IEEE International Conference on System of Systems Engineering*, Singapore, 2008.