Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

A Comprehensive Overview of Machine Learning Based Feature Extraction Techniques for Hyperspectral Image Classification

MD Samiul Haque
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
samiultahsin2001@gmail.com

MD Hasib Nadim
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
hasibnadim0@gmail.com

Tanver Ahmed
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
tanverahmed.cse@gmail.com

Md. Muktar Hossain
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
mmuktar997@gmail.com

A.S.M Delwar Hossain
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
delwar.hossain.vu@gmail.com

***Abstract*—Hyperspectral imaging using remote sensing techniques captures important details about the things on Earth by exploiting hundreds of adjacent, tiny spectral bands. The performance is hindered when all the bands are taken into consideration for categorization. Therefore, it is essential to lower the HSI bands, generally by feature selection and extraction. One popular unsupervised feature extraction method is Principal Component Analysis (PCA). Nevertheless, it ignores the local structure of the data in favor of taking into account global variance. Another feature reduction technique Nonnegative Matrix Factorization (NMF), approximates the data in a low-dimensional subspace. The Incremental PCA (IPCA) exploits Singular Value Decomposition (SVD) to transform data to lower space and is suitable for large datasets. Another dimensionality reduction technique Factor Analysis (FA) eradicates band-to-band correlation preserving the vital spectral information. This study investigates the performance among these for feature extraction techniques for effective HSI categorization. The rigorous analysis proves FA as the superior among the other techniques providing an Overall Accuracy of 92.70%, while PCA, NMF and IPCA provide 82.36%, 82.44% and 80.15% respectively.**

***Index Terms*—Remote Sensing, Hyperspectral Image, PCA, NMF, IPCA, Factor Analysis**

I. Introduction

Several hundred narrow and adjacent spectral wavelength bands consist hyperspectral images (HSIs). This feature of HSIs makes it possible to observe the materials of the earth in great detail. The HSIs must be captured using remote sensing techniques. A range of spectral wavelengths between 0.4 and 3.0 µm is used to capture the HSI bands [1], [2]. Pharmaceutical, agriculture, geology, medical research, mining, food quality monitoring, surveillance systems, and the detection of counterfeit goods are notable industries that use HSI [3].

At the expense of more space and computational complexity, HSI's narrow spectral resolution allows for the categorization of earth materials [4]. Two dimensions, X and Y, are used to represent the spatial domain of earth materials and number of spectral bands or spectral information is represented F. Thus, $X \times Y \times F$ forms a hypercube.

Four popular dimensionality reduction methods are the subject of this study: Factor Analysis (FA), Non-negative Matrix Factorization (NMF), Principal Component Analysis (PCA), Incremental PCA (IPCA). PCA is a well-known and conventional technique that maximizes variance by identifying uncorrelated orthogonal components, which makes it a very useful tool for spectral decorrelation. By using batch-wise processing to handle large datasets, IPCA expands on PCA and is appropriate for situations with limited computational resources. On the other hand, NMF breaks down data into non-negative, additive components, providing interpretability and being especially useful for identifying part-based structures. FA, on the other hand, assumes that data variability results from a smaller number of unobserved factors combined with noise, emphasizing the modeling of underlying latent factors.

Although principal component analysis (PCA) is the most widely used unsupervised linear FE method in hyperspectral imaging, Cheriyadat and Bruce [5] assert that PCA is not appropriate for efficient Feature Extraction (FE) for HSI classification, the main reasons are as follows: (i) PCA takes into account the global variance of the entire HSI, which makes it difficult to extract high-quality local features from a particular data distribution; (ii) PCA is controversial in the visible and near-infrared bands [6]; and (iii) the top extracted features, known as principal components (PCs), cannot ensure that they contain unique, useful information about the entire HSI [2]. However, Incremental Principal Component Analysis (IPCA) is frequently used as a substitute of Principal Component Analysis (PCA) while the dataset to be deconstructed is too large to be stored in memory [7]. In contrast to PCA, IPCA uses a preset amount of memory to build a low-rank estimation for the input data, regardless of the number of input data sets. While it still depends on the input data features, memory use may be controlled by adjusting the batch size. So that IPCA can improve feature extraction and provide more effective classification. IPCA can find a similar data projection to PCA, even though it only processes a small number of samples at a time. Generally speaking, IPCA is designed for large datasets that require incremental approaches and barely fit in the primary memory.

On the other hand, the nonnegative constraints of Nonnegative Matrix Factorization (NMF) allow it to be distinguished from PCA [8]. With its simple iteration-based calculation, it can use the non-negativity of HSI pixel intensity values and extract nuanced information from big datasets [9]. By identifying the underlying factors of the initial image bands and using those factors to represent the spectral information of the original image, the Hughes effect can be overcome using the factor analysis dimension reduction technique. Following that, we employed a multilayer perceptron classifier to categorize every pixel in the input hyperspectral image and a convolutional neural network to integrate the image's spectral and spatial information in a single step. The investigation presented in this literature includes the following contributions-

– An in-depth analysis of how well ML-based feature extraction methods work for HSI classification.
– An insightful comparison between the investigated ML-based approaches.

The remainder of this literature is organized as follows. Section II describes the insights of the PCA, NMF, IPCA and FA. Results analysis has been depicts in Section III, and the conclusion and future work in Section IV.

II. Methodology

A. Principal Component Analysis (PCA)

PCA, a statistical approach for feature extraction used for data compression and dimensionality reduction preserving variance of the dataset as much as possible. To employ [10], [11] the PCA, first a data matrix $\mathbf{D}$ of size $F \times S$ is constructed from the HSI data-cube. Here, F denotes the bands or features and $S = X * Y$ denotes the number of pixels of the data-cube. An object is uniquely identified by each spectral vector, which is expressed as $\mathbf{x}_n = [\mathbf{x}_{n1}\mathbf{x}_{n2}....\mathbf{x}_{nF}]^T$, where $n \in [1, S]$. Using the F spectral bands, the mean spectral vector $\mathbf{M}$ is calculated as follows:

$$\mathbf{M} = \frac{1}{S}\sum_{n=1}^{S}\mathbf{x}_n. \quad (1)$$

The following formula is used to construct the mean adjusted spectral vector $\mathbf{I}_n = \mathbf{x}_n - \mathbf{M}$. Following that, the mean adjusted data matrix $\mathbf{I}$ is written as $\mathbf{I} = [\mathbf{I}_1\mathbf{I}_2....\mathbf{I}_n]$. Our mean adjusted data $\mathbf{I}$'s covariance matrix $\mathbf{Cov}$ is determined using the following mathematical formula:

$$\mathbf{Cov} = \frac{1}{S}\mathbf{I}\mathbf{I}^T. \quad (2)$$

Since PCA relies on the covariance matrix's eigenvalue decomposition, the following mathematical statement is used to calculate the eigenvector and eigenvalues:

$$\mathbf{Cov} = \mathbf{VEV}^T, \quad (3)$$

where $\mathbf{E} = diag[E_1E_2....E_F]$ holds the eigenvalues of $\mathbf{Cov}$ in the main diagonal and $\mathbf{V} = [V_1V_2....V_F]$ describes the respective eigenvectors of the eigenvalues in $\mathbf{E}$. In addition to the eigenvalues being put in decreasing order ($E_1 \geq E_2 \geq ... \geq E_F$), the associated eigenvectors' order is also rearranged.

In order to generate the final projection matrix $\mathbf{PM}$, a matrix having dimension of $F \times k$ is formed, where k is the top selected eigenvectors and $k << F$. Next, the following equation is used to generate the projection matrix:

$$\mathbf{PM} = \mathbf{w}^T\mathbf{I}. \quad (4)$$

B. Incremental Principal Component Analysis (IPCA)

IPCA is typically used as a PCA substitute when the dataset to be decomposed is too large to fit in memory. Changing the batch size improves memory usage, but it still depends on the characteristics of the input data. Depending on the size of the input data, it can be substantially more memory efficient than PCA and supports sparse input. Using this method, the dataset is split up into mini-batches that can fit in the memory. The IPCA algorithm is then fed one mini-batch at a time. Mathematically, an IPCA approximates the most notable PC for data arriving sequentially and performs iteration without seemingly computing and storing the covariance matrix.

$$v^{(n)} = u^{(n1)} + \beta^{(n1)}\mathbf{D}_{(:\,,j)}\mathbf{D}^T_{(:\,,j)}u^{(n1)}, \quad (5)$$

$$u^{(n)} = \frac{v^{(n)}}{||v^{(n)}||_2} \quad (6)$$

The j^{th} column of the data matrix, $\mathbf{X}$, is $\mathbf{X}(:\,,j)$ in the equations above, and the step size is $\beta^{(n-1)} > 0$. When $v = \lambda.x$, where x and λ are two corresponding eigenvalues of the covariance matrix $\mathbf{C}$ that satisfy $\lambda.x = \mathbf{C}.x$, $v^{(n)}$ in Equations Equation 5 and Equation 6 denotes the nth step estimate of v. The IPCA method can be thought of as a singular-vector algorithm for the eigenvalue computation problem, where the associated iteration, such as Equations Equation 5 and Equation 6, only evaluates the eigenvector that corresponds to the largest eigenvalue [7]. Before computing the second-order eigenvector, where $u^{(n)}$ is determined by diving$v^{(n)}$ by the Euclidean norm of $v^{(n)}$, the data should be corrected by projecting them onto the perpendicular counterpart space of the section of space covered by $u^{(n)}$.

C. Nonnegative Matrix Factorization (NMF)

To implement NMF for HSI [12]–[14], It is executed utilizing the use of the two-dimensional hyperspectral data matrix's nonnegative spectral band values. $\mathbf{D}$. Considering the data matrix $\mathbf{R} = \mathbf{D}^T$ of dimension $\mathbf{R} = S \times F$ can be approximated through the following linear relationship:

$$\mathbf{R} \approx \mathbf{PQ} \quad (7)$$

NMF's objective is to find reduced rank matrices $\mathbf{P}$ of dimension $S \times k$ and $\mathbf{Q}$ of dimension $k \times F$ that are not negative to this extent that minimizes the following function:

$$f(\mathbf{P},\mathbf{Q}) = \frac{1}{2}||\mathbf{R} - \mathbf{PQ}||_F^2 \quad (8)$$

A selection of $\mathbf{k}$'s value is made so that $k << min(S,F)$. NMF, also referred to as a basis matrix, extracts the data set's inherent properties in $\mathbf{P}$. The basis vectors in $\mathbf{P}$ are combined linearly to estimate the spectral vectors. In lower dimension subspace, the spectral vectors can be projected using a limited number of basis vectors. A variety of methods are employed to approximate $\mathbf{P}$

and **Q** [14], [15]. To compensate for the approximation mistake, penalty terms are applied using Equation (8) as follows:

$$f(\mathbf{P},\mathbf{Q}) = \frac{1}{2}||\mathbf{R}-\mathbf{PQ}||_F^2 + \alpha g_1(\mathbf{P}) + \beta g_2(\mathbf{Q}) \tag{9}$$

here, the regularization parameters are α and β while the penalty parameters are denoted as $g_1(\mathbf{P})$ and $g_2(\mathbf{Q})$

D. Factor Analysis (FA)

This unsupervised feature extraction technique preserves the spatial information while identifying the underlying factors that provide a reduced number of dimensions for the original spectral information of hyperspectral data. In addition to lowering training parameters and compressing the original data for memory efficiency, this dimension reduction technique mitigates the curse of dimensionality issue.

All the t observed variables $(X_1, X_2, ...X_t)$ in a dataset are assumed to be linearly dependent on m unobservable, common factors $(F_1, F_2, ...F_m)$ in a factor analysis [16]. For instance

$$X_i - \mu_i = l_{i1}F1 + l_{i2}F2 + ... + l_{im}F_m + \varepsilon_i \tag{10}$$

l_{ij} denotes the weight of ith variable of the jth factor. μ_i represents $XX-i$'s mean value. The matrix version of the factor model for n observations,

$$X - \mu = LF + E \tag{11}$$

Factors that raise the probability of producing the covariances matrix of the original data will be identified using a maximum likelihood estimate approach. An independent sample of the data is assumed from a multivariate normal distribution with a covariance matrix in the form of $LL^T + \psi$ and a mean vector μ, where ψ is the covariance matrix of E [17] . To find MLE estimators μ, L and ψ, use the log likelihood function [16],

$$\begin{aligned}\mathscr{L}(\mu, L, \psi) = &-\frac{np}{2}\log 2\pi - \frac{n}{2}\log\left|LL^T + \psi\right| \\ &- \frac{1}{2}(X_i-\mu)^T(LL^T+\psi)(X_i-\mu)\end{aligned} \tag{12}$$

On the basis of Barlett approach, factor score matrix F is obtained using

$$F = \left(L^T\psi^{-1}L\right)^{-1}L^T\psi^{-1} \tag{13}$$

Finally, X is projected into a new subspace by,

$$Y = F^T(X-\mu) \tag{14}$$

The aforementioned technique can be used to reduce a hyperspectral picture cube X with dimensions of $W \times H \times \lambda$ to $W \times H \times B$, where $B \leq \lambda$.

III. Experimental Result and Analysis

A. Description of Dataset

The hyperspectral image of the Indian Pines in northwest Indiana serves as the basis for all of our research. The image captured by AVIRIS sensor consists of 145×145 pixels and 220 spectral bands resulting Indian Pine datset. The wavelengths of the spectral bands span from 400 to 2500 *nm*, and the spectral precision of the HSI data is 10 *nm*. The available ground truth consists of 16 non-contradictory classes [18]. When comparing feature extraction methods, all 10249 pixels from each of the 16 classes are considered. A 50:50 split of the total pixels is made up of 5124 training pixels and 5125 testing pixels. Half of the pixels in each class are in the training set, while the remaining half are in the testing set.

TABLE I: OA comparison of the investigated techniques

No of Features	PCA	IPCA	NMF	FA
5	72.21%	66.87%	71.40%	86.44%
10	75.82%	76.51%	75.73%	91.24%
15	76.10%	76.55%	76.29%	91.61%
20	77.00%	75.53%	76.66%	92.60%
26	77.50%	73.97%	76.96%	**92.70%**
30	80.31%	74.65%	77.17%	92.21%
35	80.68%	75.30%	78.24%	91.88%
40	80.90%	78.67%	80.68%	91.90%
45	81.64%	78.44%	81.93%	91.84%
50	**82.36%**	**80.18%**	**82.44%**	91.61%

B. Experimental Setup

One of the main challenges for feature extraction techniques is figuring out how many Principal Components (PCs) are best for the best classification results. We have made analysis from 1 to 50 extracted features for PCA, NMF, IPCA and FA separately. The overall accuracy (OA), precision score, recall score, and f1 score are computed for classification using the Scikit-learn Python library's support vector machine (SVM) with an RBF kernel. As a classifier model, SVM is well accepted and used by many researchers in the field of HSI classification. SVM is utilized because it may benefit from margin-based criteria and is incredibly resistant to the Hughes phenomenon [3], [19], [20]. The optimal gamma (γ) and C parameter values for efficient SVM model training and testing are determined through the use of 10-fold cross validation.

C. Performance analysis

The table I illustrates the comparison between the four investigated dimensionality reduction techniques- PCA, IPCA, NMF, and FA in terms of OA. With 26 features, FA continuously reaches the best accuracy, reaching a peak of 92.70%, proving its supremacy in feature reduction for this dataset. PCA also provides a decent performance but with more number of features than FA reaching 82.36% at 50 features. With a peak performance of 82.44% at 50 features, NMF exhibits consistent progress and attains accuracy levels that are equivalent to those of PCA and FA at larger feature counts. Despite having the lowest overall accuracy, IPCA becomes better with more features, reaching 80.18% with 50 features. The OA comparison can be observed from the figure 1. The comparison in terms of precision, recall and F1 scores can be obtained from the figure 2, 3, and 4 respectively. All these analysis proves FA as the most effective feature extraction approach for the Indian Pine dataset followed by PCA and NMF, and IPCA.

IV. Conclusion

Factor Analysis has proved as an efficient dimensionality reduction technique for HSI classification, providing the highest OA with a strong ability to preserve important information. Additionally, PCA works well, especially as the number of features rises, making it a reliable substitute at greater dimensions.NMF is suitable for preserving classification accuracy, as evidenced by its performance being equivalent to PCA at increased feature

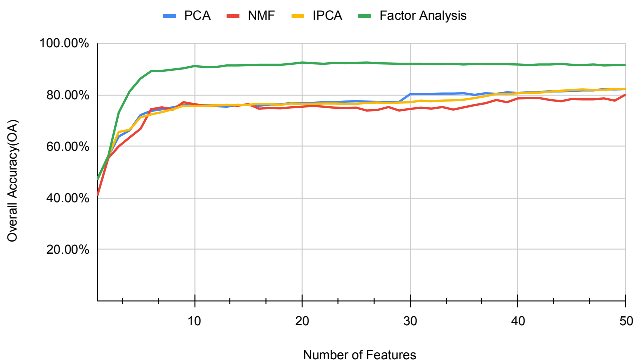

Fig. 1: Overall Accuracy comparison for PCA, NMF, IPCA and FA

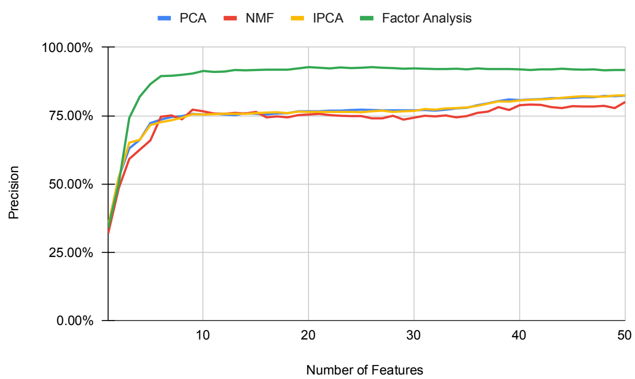

Fig. 2: Precision comparison for PCA, NMF, IPCA and FA

sets. On the other hand, IPCA performs worse overall but steadily gets better as the number of features rises. These results show that FA is the best technique for dimensionality reduction in this situation, followed by PCA and NMF, however IPCA might not be as good for this use case. Future research will investigate deep learning based model for HSI classification.

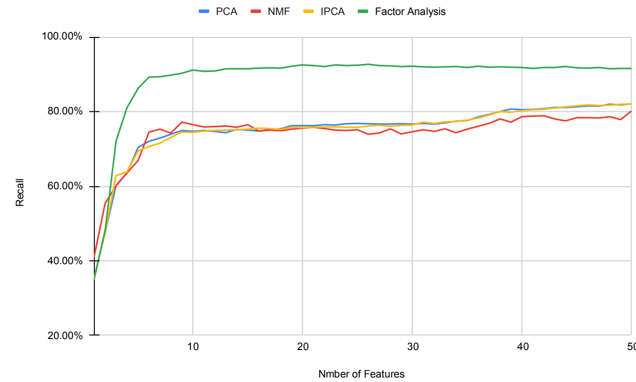

Fig. 3: Overall Accuracy comparison for PCA, NMF, IPCA and FA

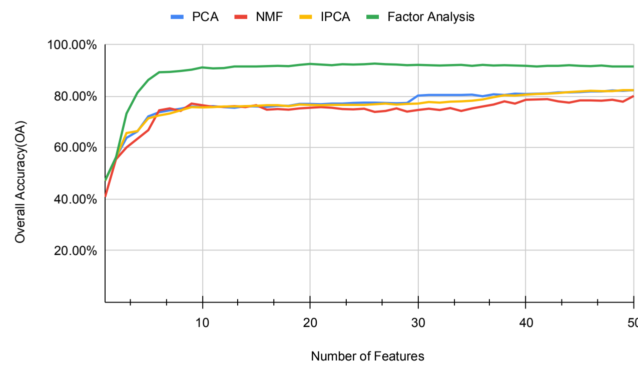

Fig. 4: F1 score comparison for PCA, NMF, IPCA and FA

REFERENCES

[1] J. Richards, *Remote Sensing Digital Image Analysis*, 10 2013.

[2] M. P. Uddin, M. A. Mamun, M. A. Hossain, and M. I. Afjal, "Improved folded-pca for efficient remote sensing hyperspectral image classification," *Geocarto International (just-accepted)*, pp. 1–23, 2022.

[3] J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, and J. Han, "Novel folded-pca for improved feature extraction and data reduction with hyperspectral imaging and sar in remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 93, p. 112–122, 07 2014.

[4] B. Guo, S. Gunn, R. Damper, and J. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 4, pp. 522–526, 2006.

[5] R. Dianat and S. Kasaei, "Dimension reduction of optical remote sensing images via minimum change rate deviation method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 198–206, 2010.

[6] X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*, 01 2006.

[7] S. Ahmed, M. Haque, M. Marjan, M. P. Uddin, and M. I. Afjal, *Segmented-Incremental-PCA for Hyperspectral Image Classification*, 06 2023, pp. 550–563.

[8] M. H. Bari, T. Ahmed, M. I. Afjal, A. M. Nitu, M. P. Uddin, and M. A. Marjan, "Segmented nonnegative matrix factorization for hyperspectral image classification," in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2023, pp. 1–5.

[9] R. Zdunek, "Hyperspectral image unmixing with nonnegative matrix factorization," in *2012 International Conference on Signals and Electronic Systems (ICSES)*, 2012, pp. 1–4.

[10] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surv Land inf Syst*, vol. 62, 01 2002.

[11] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Pca-based feature reduction for hyperspectral remote sensing image classification," *IETE Technical Review*, vol. 38, no. 4, pp. 377–396, 2021.

[12] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," *Adv. Neural Inform. Process. Syst.*, vol. 13, 02 2001.

[13] S. Tsuge, M. Shishibori, S. Kuroiwa, and K. Kita, "Dimensionality reduction using non-negative matrix factorization for information retrieval," vol. 2, 02 2001, pp. 960 – 965 vol.2.

[14] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.

[15] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, 2013.

[16] S. M and G. Sadashivappa, "Hyperspectral image classification using convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 12, 01 2021.

[17] S. Kassim, H. Hasan, A. Ismon, and F. Asri, "Parameter estimation in factor analysis: Maximum likelihood versus principal component," *AIP Conf Proc*, vol. 1522, pp. 1293–1299, 04 2013.

[18] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, "220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3," Sep 2015.

[19] M. Li, S. Zang, B. Zhang, S. Li, and C. Wu, "A review of remote sensing image classification techniques: the role of spatio-contextual information," *European Journal of Remote Sensing*, vol. 47, no. 1, pp. 389–411, 2014.

[20] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, pp. 1778 – 1790, 09 2004.