Advancing Hyperspectral Image Classification with Deep Features Learning and Evolutionary Algorithms

Mohammed Bilel Amri [1,2], Dounia Yedjour [3], Mohammed El Amin Larabi [4], Faouzi Berrichi [1]

[1] Agence Spatiale Algérienne, Centre des Techniques Spatiales, Arzew, Algeria – {bamri, fberrichi}@cts.asal.dz
[2] Laboratoire SIMPA, Département Informatique, Université des Sciences et de la Technologie d'Oran, Mohamed Boudiaf USTO-MB, Oran, Algeria – mohammedbilel.amri@univ-usto.dz
[3] Laboratoire ADASCA, Département Informatique, Université des Sciences et de la Technologie d'Oran, Mohamed Boudiaf USTO-MB, Oran, Algeria – dounia.yedjour@univ-usto.dz
[4] Agence Spatiale Algérienne, Algiers, Algeria – malarabi@asal.dz

Keywords: Features Learning, Hyperspectral data, Classification, Convolutional Neural Network, Machine Learning, Evolutionary Algorithms.

Abstract

Hyperspectral Images (HSI) reveal the secrets of land cover at a granular level, capturing hundreds of narrow spectral bands rich in detailed information. However, the sheer dimensionality of these data poses significant challenges to traditional Machine Learning (ML) methods. This research tackles the high-dimensional challenge of HSI classification with an advanced hybrid framework, leveraging the power of Deep Learning (DL), ML and Evolutionary Algorithms (EA) to conquer this challenge and achieve accurate HSI classification. We unleash the data's inherent wisdom via deep Features Extraction (FE) and optimize the representation through EA. Experiments on the Hyperion Earth Observation-1 (EO-1) show that our approach outperforms state-of-the-art ML based methods in analyzing Earth's diverse landscapes. In addition, the experiments conducted on simulated benchmarks validate the superior performance of the proposed approach compared to the baseline ML model in terms of prediction accuracy and F1-score.

1. Introduction

Hyperspectral sensors capture incredibly high-dimensional data by detecting hundreds of narrow spectral bands across the electromagnetic spectrum. However, this rich information contains a large amount of redundancy, posing challenges for interpretation and analysis. Features Extraction (FE) techniques tackle this challenge by transforming the original data into a new set of features that effectively represent key information. It reduces the complexity of the data while retaining its essential properties to become significantly more informative and suitable for various analytical objectives in Remote Sensing (RS) applications.

Features Learning (FL) plays a fundamental role in analyzing RS imagery by identifying and extracting distinct elements that represent specific phenomena of interest. These elements, the building blocks of the image, can take various forms, including geometric shapes, textured patterns, spectral signatures, or statistical measures. Extracting meaningful features is critical for a variety of RS tasks, such as mapping land cover, detecting changes, detecting objects, and classifying spectral data. The choice of features depends largely on the specific application and its goals. Essentially, FL aims to create concise and informative representations of the HSI content, enabling in-depth analysis and unlocking valuable insights.

Supervised FL offers a diverse toolbox for extracting informative features from HSI. Established techniques like Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) (Khalid et al., 2014) offer dimensionality reduction and discriminative feature, respectively. Additionally, Autoencoders (Kim et al., 2021) and Independent Component Analysis (ICA) (Khalid et al., 2014) can uncover hidden patterns and independent features within the data. Decision Trees (DT) (Kim et al., 2021) provide interpretable rules for features selection, while Convolutional Neural Networks (CNNs) (Kim et al., 2021) excel at automatically learning complex, hierarchical features from spatial and spectral information.

In the realm of RS, (Li et al., 2019) introduced ASSFL (adaptive spatial-spectral features learning network), a Deep Learning (DL) model for HSI classification. ASSFL leverages two key components: a CNN that learns spatially-adaptive weights for each pixel, amplifying relevant features based on local context, and a Stacked Autoencoder (SAE) that extracts progressively deeper, abstract features from the data. Building upon existing efforts, (Quan et al., 2020) proposed a CNN architecture for HSI classification capable of extracting informative spectral-spatial features (SSF). (Ladi et al., 2023) introduced PKNNET (Polynomial Kernel Kervolutional Neural Network), a novel DL model designed for spatial FE from HSI. Subsequently, the extracted features are fed into a Support Vector Machine (SVM) for robust classification. (Petrovska et al., 2020) developed a two-stream DL architecture for aerial image scene classification. This architecture leverages: (1) pre-trained CNNs for robust FE, (2) dimensionality reduction techniques to handle high-dimensional features vectors, and (3) features concatenation to create a comprehensive representation for SVM-based classification. (Amri et al., 2024) proposed an innovative approach to improve water body classification from PRISMA hyperspectral data by combining FE with a convolutional extreme learning machine (CELM) and EA, the goal is to simplify complex data while retaining essential information and then automatically optimize model parameters to improve classification accuracy.

This work presents a cutting-edge approach to FL from HSI Hyperion Eo-1. We leverage a CNN trained on a binary classification task to automatically extract discriminative features. Subsequently, these features are employed in conjunction with established Machine Learning (ML) algorithms, further optimized through Evolutionary Algorithms (EA) like Genetic Algorithm (GA) (Holland, 1992) and Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995), for comprehensive HSI classification.

This paper is structured in the following manner. Section 1 presents a comprehensive literature review, delving into the recent advancements in FL techniques specifically within the context of HSI analysis. Section 2 meticulously details our proposed methodology, encompassing the design of the CNN architecture, the employed ML algorithms, and the integration of EA for parameter optimization. Section 3 outlines the experimental setup, including the selected hyperspectral databases and pre-processing procedures. Section 4 then presents and elucidates the obtained results from our experiments, demonstrating the efficacy of our novel approach. Finally, the concluding section summarizes the significance of the study, outlining potential implications and exciting avenues for future research endeavors.

2. Proposed Methodology

This section delves into the application of CNNs for FE, followed by precise classification leveraging a hybrid ML framework complemented by EA. This innovative approach synergistically combines the inherent advantages of CNNs in deciphering and interpreting intricate visual data with the resilience and adaptability of ML and EA for robust classification. Equation (1) represents the raw data extracted from the EO-1 HSI, while Figure 1 visually depicts the proposed methodology adopted in this work.

$$Rd = \{x_1, x_2, , \dots \dots \dots \dots . . x_N\} \tag{1}$$

Where N is number of spectral bands.

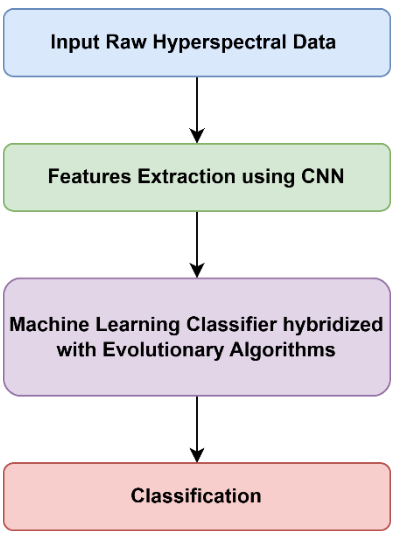


Figure 1. Description of the proposed approach.

2.1 Features Learning

Initially, attention is drawn to CNNs due to their remarkable effectiveness in this domain. This efficacy stems from their innate ability to automatically and hierarchically extract spatial features from raw data. Notably, FE via CNNs involves leveraging these networks to transform raw data into a collection of features that are demonstrably more informative and valuable for diverse tasks, including classification. The CNN architecture used in this work is designed for image classification, it's composed of several layers that progressively extract increasingly extract features from input data. it uses convolutional layers to detect local patterns and pooling layers to reduce dimensionality and fully connected layers for the final classification.
layers are typically followed by non-linear activation functions to introduce non-linearity into the model. This architecture can be adapted to various image classification problems.

2.2 Classification

In ML, several classifiers are commonly used for image classification, each with its own advantages and specific areas of application. These include the Support Vector Machine (SVM) which is effective for high-dimensional problems, the Decision Tree (DT) known for its ease of interpretation, the *k* Nearest Neighbors (*k*-NN) appreciated for its simplicity and effectiveness on well-separated datasets and the random forest (RF) a set of decision trees that work collaboratively to improve overall accuracy.
In this contribution, we opted for the RF due to its robustness, its ability to handle noisy data and above all its high accuracy widely demonstrated in the scientific literature and through numerous experiments with other machine learning techniques (Kim et al., 2021). In addition, RF has the advantage of limiting overfitting while offering good performance on complex data sets which makes it a particularly suitable choice for our problem.

2.3 Evolutionary algorithms-based optimization

EA can be used to optimize the parameters and architecture of ML model for image classification. The keys advantage of using EA in classification is their ability to efficiently explore the solution space, ability to process complex data, and search for an optimal representation. EA are a class of optimization and search algorithms inspired by the biological process of evolution. They used to solve complex problem and achieve accurate and reliable classification results by mimicking the process of biological evolution.

2.3.1 Genetic Algorithms (GA): GA (Holland, 1992) are stochastic optimization methods that mimic the natural process of biological evolution to efficiently solve complex problems. Inspired by the fundamental mechanisms of Darwin's theory of evolution, these algorithms are based on concepts such as natural selection, crossover (recombination), mutation, and adaptation. The process (Figure 2) begins with the generation of an initial population composed of random individuals, each representing a potential solution to the problem in the form of chromosomes. Each individual is evaluated using an evaluation or fitness function that measures the quality of the proposed solution.
During each iteration, called a generation, a natural selection process is applied to select the best-performing individuals with a greater probability of passing their characteristics on to the next generation. Then, the selected individuals undergo crossover, where parts of their chromosomes are exchanged to produce new individuals called offspring. This mechanism allows the best characteristics of each parent to be combined and new regions of the search space to be explored.
To maintain genetic diversity and avoid the risk of premature convergence towards a local optimum, a mutation process is introduced, consisting of randomly modifying a part of the individual. The selection, crossover, and mutation process are repeated over several generations until a stopping criterion is reached, such as a maximum number of generations or a desired performance level.

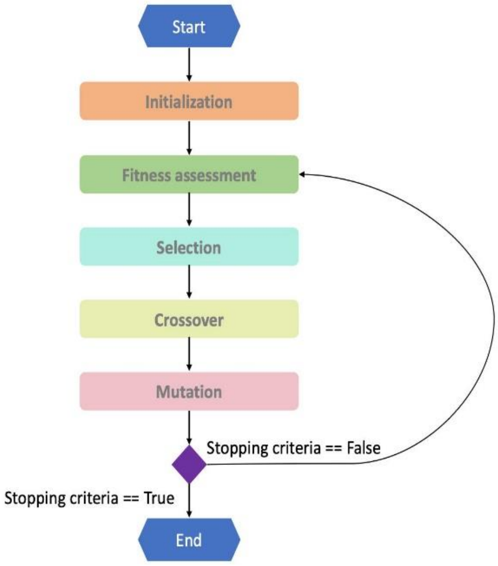


Figure 2. GA algorithm steps.

2.3.2 Particle Swarm Optimization (PSO):

PSO (Kennedy and Eberhart, 1995) is an optimization metaheuristic inspired by the collective behavior of swarms of birds or fish searching for food. In this algorithm, each particle represents a potential solution in the search space, and the set of particles constitutes a swarm. Each particle moves through this space, adjusting its position and velocity at each iteration, guided both by its own experience (its best position found) and by the collective experience of the swarm (the best position found by all particles). Key parameters of PSO include the number of particles, their initial position, their velocity, as well as their personal and global inertia coefficients and attraction factors. A particle's velocity is updated according to an equation that takes into account its previous velocity, the distance from its own best-known position, and the distance from the swarm's best position.

The particle's position is then adjusted based on its new velocity. These two updates are generally expressed by equations (2) and (3) in classical PSO formulations. This simple yet powerful mechanism allows particles to gradually converge toward optimal regions of the search space making PSO an effective tool for optimizing complex functions particularly for problems where the exact solution is difficult to find using deterministic methods.

$$v_{id}^{t+1} = wv_{id}^{t} + c_1 r_{1i}(p_{id} - x_{id}^{t}) + c_2 r_{2i}(p_{gd} - x_{id}^{t}) \quad (2)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (3)$$

Where *t* is t^{th} iterations and d is the dimensional search space. The mass of inertia is represented by w, constant coefficients by c_1 and c_2, r_{1i}, r_{2i} represent random values uniformly distributed in the range [0,1]. Pid and P_{gd} represent the best individual (P_{best}) and global (G_{best}) elements, respectively. During each iteration, both velocity and position are updated in order to search the optimal solution.

Due to its high classification on HSI, in this work RF classifier have been optimized through the EA GA and PSO with the aim of automatically tuning the classifier and extracting the best performance for the considered classification task. The PSO steps are detailed in Figure 3.

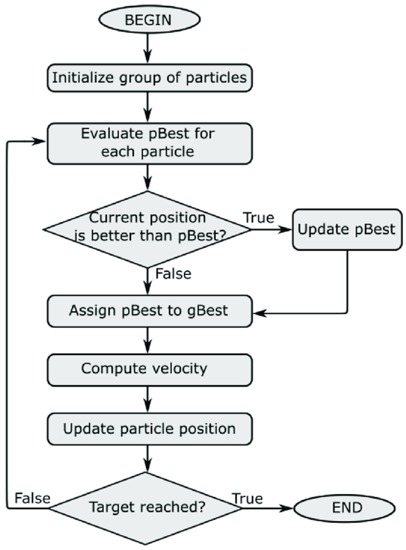


Figure 3. PSO algorithm steps.

This hybridization explores improving the accuracy of EO-1 hyperspectral RS data by optimizing the hyperparameters of a RF model using GA and PSO. The adjusted hyperparameters include the number of trees, the maximum depth, the minimum number of samples to split a node and the minimum number of samples per leaf. These evolutionary algorithms inspired by natural selection are used to efficiently search for optimal combinations of these hyperparameters. This approach improves model performance facilitating more accurate classification.

3. Experiments

This study aims to improve the performance of the EA for the Hyperion EO-1 HSI classification. This will result in obtaining a high-performing classification system by leveraging both the FE capabilities of CNN and the generalization ability of ML classifiers, jointly optimized by EA.

3.1 Dataset

The Hyperion EO-1 HSI dataset is a valuable resource for Earth observation. This instrument on board the EO-1 satellite captures hyperspectral images of remarkable precision, with 242 continuous spectral bands covering wavelengths from 0.4 to 2.5 µm The spatial resolution of 30 meters allows for detailed terrain analysis (ALI, 2003). The selected study area is located in the Oran region, a major metropolis in northwestern Algeria, characterized by a diversity of urban, agricultural and natural landscapes. This region is of particular interest due to its rapid urban development and varied ecosystems. Figure 4 presents a false-color visualization of the data; a technique that can highlight certain terrain features that would not be visible in a natural-color image. This representation is accompanied by its corresponding ground truth[1] which serves as a validated reference for assessing the accuracy of the analyses. The ground

[1]https://github.com/bilelamri687/EO-1_Vegetation_Mapping

truth was established through in situ observations and expert validation, thus ensuring the reliability of the reference data for subsequent analyses.

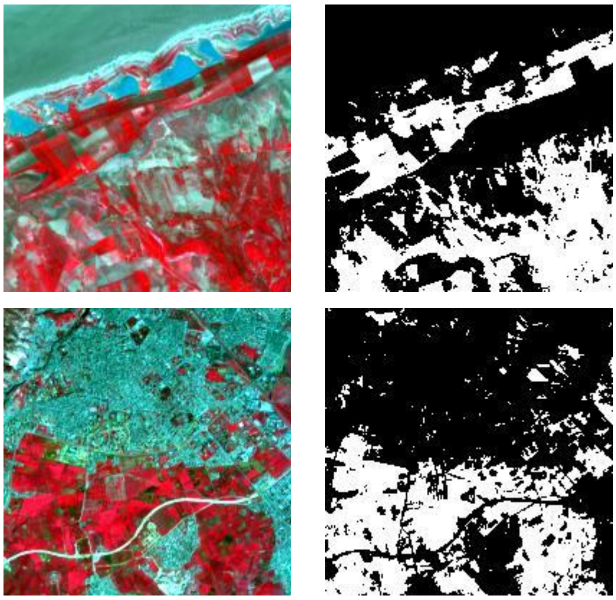


Figure 4. The false color composite from EO-1 HSI. The first column represents the training data, while the second one represents the manually labeled ground truth.

3.2 Performance

To evaluate the performance of our model, it is common to use multiple metrics in order to obtain a more comprehensive understanding of its effectiveness. Among these metrics accuracy, precision, recall and F1-score (Amri et al., 2022) as depicted in equations (4),(5),(6), and (7) respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{7}$$

TP, *TN*, *FP* and *FN* are true positive, true negative, false positive and false negative respectively.

4. Results and Discussion

This section presents various experimental results to assess the performance of our approach. Figure 5 and 6 show the evolution of the loss function and accuracy on the training set over the epochs of learning for our CNN model respectively. In the loss function (Figure 5), The blue (training) and orange (testing) curves decrease rapidly at first and then more gradually. Their convergence towards similar values indicates good generalization of the model without apparent overfitting. In the accuracy function (Figure 6), we observe a rapid increase in accuracy at the beginning of training (within the first 25 epochs), starting from around 90% and reaching nearly 97%. The progression then becomes slower but continues, with both curves reaching around 99% accuracy towards the end of training.

The fact that the test and validation curves follow similar trajectories and converge towards very close values indicates that the model generalizes well to data it did not see during training. This excellent performance suggests a robust model with a very low error rate.

Figure 7 illustrates example of test image from the Hyperion EO-1 HSI used to evaluate our classification approach while Figures 8, 9 and 10 show the obtained classification results from random Forest algorithm, random forest with GA and the random forest with PSO respectively.

Figure 11 shows the evolution of the fitness function of our GA over successive generations; the fitness function is defined as the accuracy achieved by the RF classifier on the validation set. We observe a rapid and significant improvement in fitness in the first 5 generations, going from approximately 0.986 to over 0.993. After this initial phase of rapid improvement, the curve stabilizes and forms a plateau, indicating that the algorithm has reached convergence. This trend is typical of genetic algorithms where the greatest improvements occur in the first generations, followed by a phase of slower refinement and then stabilization.

Table1 presents the evolution of accuracy on the training set during the use of the RF classifier, with and without optimization by GA and PSO. We observe the evolutionary approach achieve better generalization on the training set with a final accuracy of 0.997 for RF-GA, 0.995 for RF-PSO compared to 0.990 for baseline model.

The table 2 illustrates and compares the performance metrics of different classification models on the test set (Figure 7). The results show interesting differences in their performances. The standard RF, applied on raw data, achieves high precision (0.951) with an excellent precision/recall balance (0.990/0.950) but its relatively lower F1-score (0.750) suggests limitations in its generalization ability. The approach in (Amri et al., 2023) using Wrapper Feature Selection (WFS) shows an overall accuracy of 0.988 with a better F1-score (0.819). This improvement in F1-score over standard RF indicates that features selection helps to better capture relevant information and reduce noise in the data. The RF-PSO and RF-GA methods based on FE show particularly interesting results. Notably, RF-GA achieves the best overall accuracy (0.994) and the best F1-score (0.832) of all models. This highlights the effectiveness of optimization algorithms such as GA, in improving model performance.

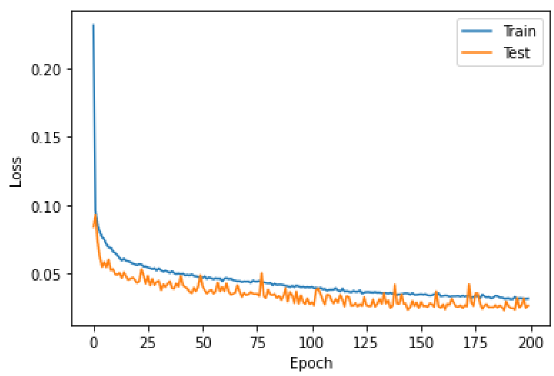


Figure 5. CNN’s training evolution loss plot as a function of epochs.

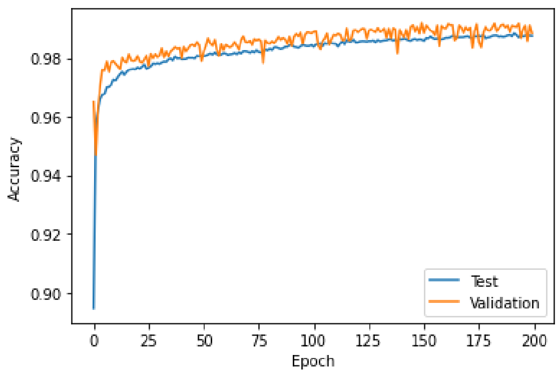
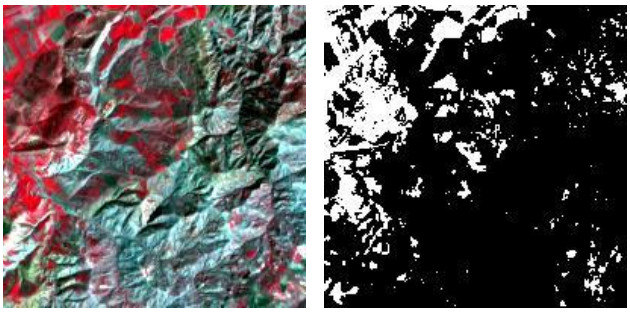


Figure 6. CNN's training evolution accuracy plots as a function of epochs.



(a) (b)

Figure 7. Test image of EO-1 HIS for model evaluation, (a) Original test data, (b) Ground truth.

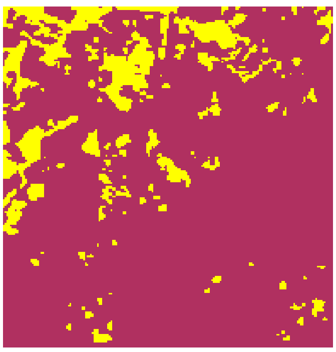


Figure 8. Examples of Classification results by RF.

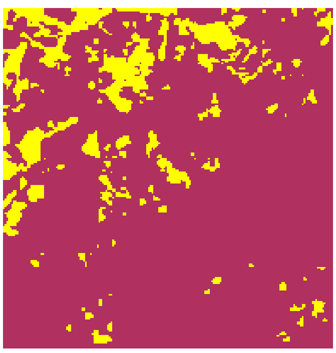


Figure 9. Examples of Classification results by RF-GA

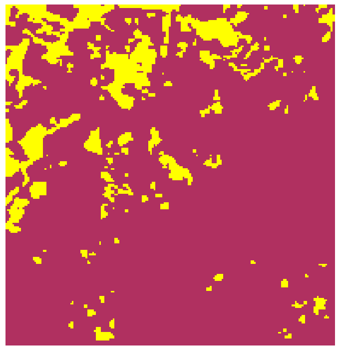


Figure 10. Examples of Classification results by RF-PSO.

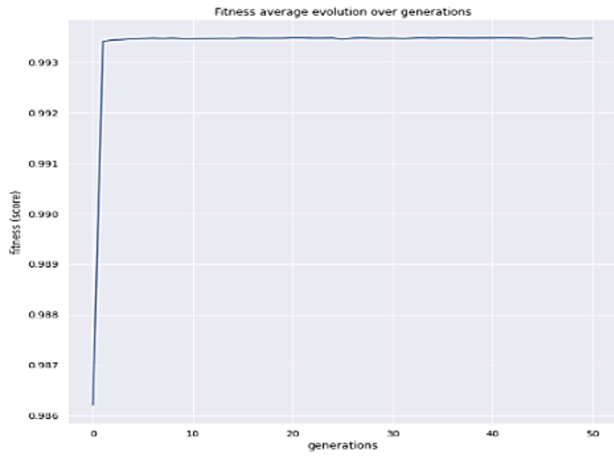


Figure 11. Evolution of the fitness function of genetic algorithm.

Models	Training accuracy
Random Forest (RF)	0.990
RF-PSO	**0.995**
RF-GA	**0.997**

Table 1. Accuracy on the training set of different techniques

Models	Accuracy	Precision	Recall	F1
Random Forest (RF)	0.951	0.990	0.950	0.750
Method (Amri et al., 2023)	0.988	0.903	0.901	0.819
RF-PSO	0.968	0.926	0.912	0.732
RF-GA	**0.994**	**0.956**	**0.952**	**0.832**

Table 2. Comparison of classification model evaluation metrics.

5. Conclusion

Beyond traditional methods, this research proposes a novel framework for HIS classification powered by Deep Features Learning and EA to optimize the hyperparameters of ML models. This innovative hybridization leverages the ability of deep neural networks to automatically extract relevant features from HSI data while leveraging the exploratory power of EA to efficiently tune model hyperparameters. By capturing the essence of important features and adaptively optimizing model configurations, this

data driven approach produces a robust and generalizable model capable of better interpreting the complexity and spectral richness provided by HSI data.
The resulting model exhibits increased robustness and better generalization significantly improving the ability to extract complex information from HSI data through intelligent exploration of the search space by EA thus avoiding stagnation in local optima. Future work could incorporate more sophisticated evolutionary mechanisms and explore new deep neural network architectures to further enhance performance and generalization.

References

ALI, L.H., 2003. Earth Observing-1 Extended Mission.

Amri, M.B., Larabi, M.E.A., Bakhti, K., Meroufel, H., 2022. Plastic Greenhouses Detection from Alsat-2A Satellite Data Using Mask R-CNN, in: 2022 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS). IEEE, pp. 82–85.

Amri, M.B., Yedjour, D., Larabi, M.E.A., 2023. Evolutionary Algorithms Based Feature Selection For Remote Sensing Image Classification, in: 2023 8th International Conference on Frontiers of Signal Processing (ICFSP). Presented at the 2023 8th International Conference on Frontiers of Signal Processing (ICFSP), pp. 79–82. https://doi.org/10.1109/ICFSP59764.2023.10372938

Amri, M.B., Yedjour, D., Larabi, M.E.A., Berrichi, F., 2024. Spatial Pattern Recognition for Precise Water Body Extraction: Integrating PRISMA Hyperspectral Data with Evolutionary Machine Learning Algorithms.

Holland, J.H., 1992. Genetic algorithms. Scientific american 267, 66–73.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: Proceedings of ICNN'95-International Conference on Neural Networks. IEEE, pp. 1942–1948.

Khalid, S., Khalil, T., Nasreen, S., 2014. A survey of feature selection and feature extraction techniques in machine learning, in: 2014 Science and Information Conference. IEEE, pp. 372–378.

Kim, Daekyum, Kim, S.-H., Kim, T., Kang, B.B., Lee, M., Park, W., Ku, S., Kim, DongWook, Kwon, J., Lee, H., 2021. Review of machine learning methods in soft robotics. Plos one 16, e0246102.

Ladi, P.K., Kakita, M.G., Dash, R., Ladi, S.K., 2023. PKNNet: a novel feature learning architecture for vegetation mapping using remote sensing hyperspectral image classification. Earth Sci Inform 16, 1091–1107. https://doi.org/10.1007/s12145-022-00923-3

Li, S., Zhu, X., Liu, Y., Bao, J., 2019. Adaptive spatial-spectral feature learning for hyperspectral image classification. IEEE access 7, 61534–61547.

Petrovska, B., Zdravevski, E., Lameski, P., Corizzo, R., Štajduhar, I., Lerga, J., 2020. Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification. Sensors 20, 3906.

Quan, Y., Dong, S., Feng, W., Dauphin, G., Zhao, G., Wang, Y., Xing, M., 2020. Spectral-spatial feature extraction-based CNN for hyperspectral image classification, in: IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 485–488.