Semantic Segmentation on Hyperspectral Images: Band Selection and Deep Learning Approaches

Assou Samir IBENJELLAL Malak FATARA Oumaima EL AZHARI
Mehdi ZAKROUM
International University of Rabat
(assou-samir.ibenjellal, malak.fatara, oumaima.elazhari)@uir.ac.ma

January 20, 2025

Abstract

Semantic segmentation of hyperspectral images (HSI) is crucial for applications like remote sensing and precision agriculture. This study combines exploratory data analysis, band selection, and deep learning models for HSI segmentation. Two band selection techniques—Max Relevance Min Redundancy (MRMR) and a deflection coefficient-based method—were used to reduce dimensionality while preserving essential spectral information. Models developed include a Multi-Layer Perceptron (MLP), DeepLabV3, and U-Net, with performance evaluated using the F1-score. While band selection improved computational efficiency, it did not significantly enhance accuracy. Advanced models underperformed the MLP due to the small dataset size (100 images), underscoring the need for larger datasets or effective augmentation strategies to train deep learning models for HSI segmentation.

1 Introduction

Hyperspectral imaging (HSI) is a powerful technology that captures detailed spectral information across hundreds of contiguous bands, making it a valuable tool for material identification and classification. Its applications span diverse fields such as remote sensing, environmental monitoring, and precision agriculture. However, working with HSI data comes with unique challenges. The high dimensionality of the data significantly increases computational demands, while the limited availability of labeled samples makes it difficult to train accurate and robust machine learning models. These issues present major obstacles to effectively leveraging HSI for tasks like semantic segmentation.

This paper focuses on optimizing semantic segmentation pipelines for HSI data by addressing these key challenges. Specifically, it explores how to reduce the dimensionality of HSI data without losing critical information and investigates the effectiveness of advanced models in handling this reduced data. The study seeks to answer three key questions: How can dimensionality reduction techniques improve computational efficiency while preserving essential spectral information? What impact do different band selection methods have on model performance? Can deep learning architectures be effectively adapted to work with reduced-dimensional HSI data?

To tackle these questions, we propose a pipeline with three main contributions. First, we perform an exploratory data analysis (EDA) to examine the dataset's characteristics, including spectral variability and redundancy across bands. This analysis lays the groundwork for identifying the most informative spectral features. Second, we implement two band selection methods to address the high dimensionality of HSI data. The first is Max Relevance Min Redundancy (MRMR), which selects bands with high discriminative power and minimal redundancy. The second is a deflection coefficient-based method that ranks bands based on their spectral variability. Finally, we evaluate these methods using both machine learning and deep learning approaches. We employ Multi-Layer Perceptrons (MLPs), DeepLabV3 and U-Net are trained on the full dataset as well as subsets of bands selected by MRMR and the deflection coefficient method to assess how dimensionality reduction impacts classification performance.

We evaluate the proposed methods using the F1-score on both training and test datasets to assess their generalization capabilities and robustness. While band selection techniques are typically known to improve segmentation accuracy, their full potential is less evident in this study due to the limited amount of data available. This limitation is particularly pronounced for deep learning models like DeepLabV3 and U-Net, which require substantial data to fully leverage the benefits of dimensionality reduction. Despite these constraints, the study highlights the promise of integrating band selection methods with advanced modeling approaches for optimizing segmentation pipelines in hyperspectral imaging.

The structure of the paper is as follows. Section 2 reviews related work in hyperspectral imaging and semantic segmentation. Section 3 describes the dataset, band selection approaches and the model development. Section 4 presents the experimental setup and results, followed by limitations in Section 5. Finally, conclusions and future directions are outlined in Section 6.

2 Related Work

Deep learning has significantly advanced hyperspectral image (HSI) classification by addressing challenges like high-dimensionality and the need for efficient processing methods. Hyperspectral imaging, with its broad spectral bands, is valuable for applications such as land cover classification, environmental monitoring, and material detection. However, its complexity calls for specialized methods to manage both spectral and spatial information effectively.

Early studies, like "Hyperspectral Image Classification with Deep Learning Models," compared various deep learning models such as 2D-CNN, 3D-CNN, and their enhanced versions (R-2D-CNN and R-3D-CNN) to traditional classifiers. These models showed superior performance, especially R-3D-CNN, which achieved 99.50% accuracy on the Indian Pines dataset, thanks to its ability to capture both spectral and spatial details. However, these models have high computational costs and require large labeled datasets, posing challenges for deployment in low-resource settings. Transfer learning was proposed as a way to overcome the need for large datasets. [1]

Building on these findings, "Deep Learning for Classification of Hyperspectral Data: A Comparative Review" compared different deep learning architectures like CNNs and RNNs, noting CNNs' ability to capture spatial hierarchies that traditional classifiers struggle with. While CNNs demonstrated higher accuracy, they also introduced greater computational demands. The review highlighted the need for better spatial feature integration and raised concerns about the generalization of results across different datasets, laying the foundation for future improvements in HSI classification. [2]

Segmentation of hyperspectral images has benefited from deep learning advancements. "Image Segmentation Using Deep Learning: A Survey" explored several models, including CNNs, FCNs, and U-Net, which excel at providing detailed segmentation boundaries and handling various image scales. However, challenges persist, such as improving boundary precision and computational efficiency for real-time applications, especially in resource-limited settings. [3]

In response to the need for more efficient processing, "Classification of Hyperspectral Images Using a New Fully Convolutional Neural Network" introduced a hybrid approach combining CNNs, PCA, and Extreme Learning Machines (ELM) to address issues like the Hughes effect. This method demonstrated improved classification accuracy while integrating dimensionality reduction techniques to avoid overfitting and excessive computational costs, proving more effective for hyperspectral image classification. [4]

"Methodology for Hyperspectral Band Selection" focused on reducing the computational load by selecting the most informative spectral bands. Using both unsupervised and supervised methods like PCA and Information Entropy, the approach reduces dimensionality without compromising classification accuracy, though it may risk losing some spectral details. This trade-off underscores the importance of balancing efficiency with maintaining key data characteristics.[6]

Finally, The DeepLabv3 framework, discussed in "DeepLabv3", introduces Atrous Spatial Pyramid Pooling (ASPP), enabling the capture of multi-scale features without sacrificing spatial resolution. This technique, while offering improved segmentation, remains computationally intensive, posing challenges for real-time applications in resource-constrained environments. Despite these challenges, it represents a significant step forward in achieving high-quality segmentation boundaries. [5]

In conclusion, the evolution of deep learning models for HSI classification has progressed from basic CNNs to more sophisticated approaches that balance accuracy with computational efficiency. While early models set the stage for breakthroughs, later methods, including 3D convolutions, multi-scale feature extraction, and band selection, have further improved both the accuracy and practicality of these models for real-world applications, setting the stage for future advancements that integrate efficiency, transfer learning, and data augmentation.

3 Methodology

3.1 Data Description

The dataset consists of 100 hyperspectral images organized into two directories: **"images"** (the HSIs) and **"landcover"** (corresponding pixel-wise labels). Each pixel is classified into one of **34 land cover classes**, such as vegetation, urban areas, and water bodies etc.

Hyperspectral images capture more detail than RGB images by recording data across 202 spectral bands. This allows for distinguishing subtle differences in materials that may appear identical in traditional images, and each image has a resolution of 128x128 pixels, making the dataset ideal for land cover classification, environmental monitoring, and resource management.

3.2 Explanatory Data Analysis (EDA)

To gain a deeper understanding of the dataset, we conducted a series of exploratory analyses:

3.2.1 Visual Inspection of Spectral Bands

By plotting individual hyperspectral bands, we observed variations in pixel intensities across different wavelengths and classes (Figure 1). These plots demonstrate how hyperspectral images capture detailed spectral signatures, enabling differentiation between materials that may appear similar in standard RGB images.

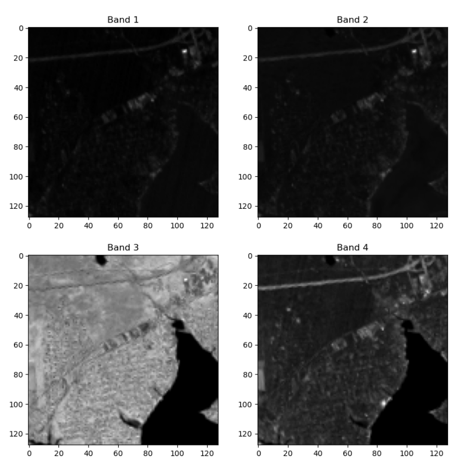


Figure 1: 4 of bands of the first image

The figure highlights intensity values from four spectral bands of a hyperspectral image. The variation across bands reveals the distinct spectral signatures of materials in the scene, providing an initial understanding of the dataset's richness.

3.2.2 Correlation Analysis of Spectral Bands

A correlation matrix was computed to explore the relationships between spectral bands (Figure 2). The matrix identifies how spectral values vary across the image's wavelength bands.

The Heatmap focuses on correlations from 0.5 to 1.0, highlighting clusters of highly correlated bands (bright regions). These clusters indicate redundancy in the dataset, which is critical for identifying bands that contribute unique information.

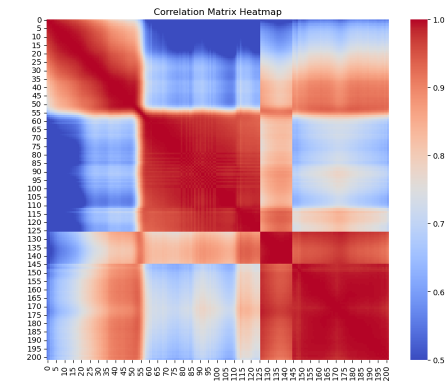


Figure 2: Correlation between bands

3.2.3 Class-Specific Spectral Signatures

As shown in Figure 3, the mean spectral intensities were computed for each class to identify patterns that distinguish one class from another.

Each bar chart illustrates how pixel intensities vary across bands for different classes. Peaks in the spectra indicate the most significant bands for specific classes, though not all classes exhibit clear patterns. These insights are instrumental in selecting the most relevant bands, reducing the number of input features, and enhancing the efficiency of the analysis without compromising accuracy.

The results of this exploratory analysis provide a foundation for subsequent steps, such as band selection and model development. By understanding correlations and class-specific spectral patterns, we can effectively reduce redundancy, focus on the most relevant features, and improve the performance of machine learning models.

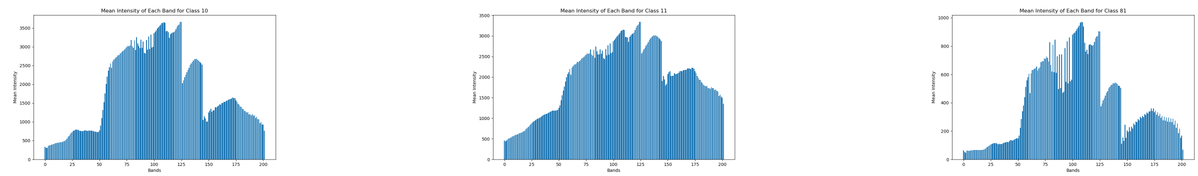
3.3 Band selection

Hyperspectral images consist of numerous spectral bands, many of which may contain redundant or irrelevant information for classification tasks. The process of band selection is crucial to reduce computational complexity, enhance model efficiency, and improve overall performance. In this study, two complementary methods were employed for band selection: **Maximum Relevance Minimum Redundancy (mRMR)** and **the Deflection Coefficient-Based Approach**.

3.3.1 Maximum Relevance Minimum Redundancy (mRMR)

The **mRMR** method focuses on selecting spectral bands that exhibit high relevance to the target class labels while minimizing redundancy among the selected bands. [7]

Relevance: Quantified the relationship between each band and the class labels, prioritizing bands with stronger class discriminative power.



(a) Mean intensity for class 10 (b) Mean intensity for class 11 (c) Mean intensity for class 81

Figure 3: Three images displayed with individual captions in two rows, spanning both columns.

Redundancy: Assessed the correlation between selected bands to ensure that each contributes unique information to the classification task.
Outcome: A subset of bands was identified that provided a balance between maximizing information content and reducing redundancy. This subset formed the basis for further analysis and model training.

$$\mathrm{mRMR}(B_i) = I(C, B_i) - \frac{1}{|S|} \sum_{B_s \in S} I(B_s, B_i)$$

- B_i: The i-th spectral band being evaluated.
- C: The vector of classes (target variable or output labels).
- $I(C, B_i)$: The mutual information between the vector of classes C and the band B_i, representing **relevance**.
- S: The subset of bands already selected.
- $|S|$: The number of bands in the subset S.
- $I(B_s, B_i)$: The mutual information between the already-selected band $B_s \in S$ and the band B_i, representing **redundancy**.
- $\frac{1}{|S|} \sum_{B_s \in S} I(B_s, B_i)$: The average mutual information between B_i and all bands in the subset S, which penalizes redundancy.

3.3.2 Deflection Coefficient-Based Approach

The second method utilized the deflection coefficient, a measure of variability in spectral intensity values across pixels within a band. Bands with higher deflection coefficients were deemed more informative.
Separability between two classes:

$$DC = \frac{(\mu_i - \mu_j)^2}{\sigma_i \cdot \sigma_j}$$

The formula provided is originally designed for two classes, but in our case, we have 34 classes, requiring an adaptation of the process. First, we calculated the deflection coefficient for every pair of classes. Next, the resulting coefficients were sorted in ascending order. For each spectral band, the coefficients from all class pairs were grouped, and the minimum coefficient was selected as the representative value for that band. The bands were then sorted in descending order based on their representative coefficients. Finally, the top N bands were selected, where N = 10 in our case. This approach ensured that the selected bands had the highest discriminatory power for distinguishing among the 34 classes.

3.3.3 Comparison between the two approaches:

To evaluate the efficacy of the band selection methods, the subsets generated by mRMR and the Deflection Coefficient-Based Approach were compared. Some bands were consistently selected by both methods, highlighting their robustness and importance within the dataset. At the same time, each method identified unique bands, demonstrating their complementary contributions to the classification process. The selected bands from both approaches significantly reduced the input feature space, addressing the challenges of high dimensionality. This reduction allowed for faster training of machine learning models while retaining the essential spectral information needed for accurate classification. Additionally, the band selection process provided valuable insights into the spectral properties of the dataset.

3.4 Model Development

In this study, three deep learning models were developed and evaluated for semantic segmentation on hyperspectral images (HSIs): Multilayer Perceptron (MLP), DeepLabV3, and U-Net. Each model was trained and evaluated under three scenarios: (1) using all spectral bands, (2) using bands selected via the mutual information-based Minimum Redundancy Maximum Relevance (mRMR) approach, and (3) using bands selected via the Deflection Coefficient (DC) approach. The following sections detail the implementation of these models and the results obtained.

3.4.1 Multi-Layer Perceptron (MLP)

The MLP was adapted for hyperspectral semantic segmentation by treating each pixel as a 1D vector, where each feature corresponds to the pixel's intensity across all spectral bands. The architecture consisted of an input layer, four hidden layers (256, 256, 128, and 64 neurons), and an output softmax layer for multi-class classification. Each hidden layer employed Batch Normalization, ReLU activation, and a 25% Dropout rate to mitigate overfitting. The model was trained using the Adam optimizer, categorical cross-entropy loss, and an exponential decay learning rate over 50 epochs, with a batch size of 512.

Performance was evaluated across the three band selection scenarios. The results indicated that band selection consistently enhanced performance by reducing redundancy and emphasizing the most discriminative spectral information. Notably, the MLP model performed particularly well with the selected bands, showing improved accuracy and F1 scores compared to using all spectral bands. This suggests that the MLP model is well-suited for handling small datasets with focused feature selection.[8]

3.4.2 DeeplabV3

DeepLabV3, a state-of-the-art model for semantic segmentation, leverages Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale contextual information. The network begins with convolutional layers that reduce spatial resolution, followed by the ASPP module, which applies parallel atrous convolutions with varying dilation rates (6, 12, and 18 in this implementation). This approach effectively captures features at multiple receptive fields. Additionally, global average pooling integrates global context, enhancing the model's ability to capture broader spatial relationships.

Band selection was applied using both the mRMR and Deflection Coefficient (DC) methods, effectively reducing the 202-band input to a more manageable subset of the most relevant spectral bands. The outputs from the ASPP module's parallel paths are concatenated and passed through convolutional layers for feature refinement. Finally, the model upsamples features to the original image resolution using bilinear interpolation, followed by a 1x1 convolution with softmax activation to generate per-pixel class probabilities.

The optimized band selection approach allowed DeepLabV3 to retain critical spectral information while minimizing redundancy, significantly improving segmentation accuracy. This model's ability to perform well with multi-scale analysis and precise boundary delineation made it particularly effective for segmenting complex hyperspectral imagery with a 34-class output configuration.

3.4.3 U-net

U-Net, a popular architecture for semantic segmentation originally designed for biomedical images, uses an encoder-decoder structure to capture spatial context by reducing spatial dimensions in the encoder and upsampling in the decoder. In this study, we adapted U-Net for hyperspectral image segmentation with a custom implementation. The encoder consists of four convolutional blocks, with filters increasing from 64 to 512, each followed by max-pooling, batch normalization, ReLU activation, and a 10% dropout rate. The bridge layer processes abstract features using 1024 filters, and the decoder upsamples with transposed convolutions, with skip connections to corresponding encoder layers. The decoder progressively reduces the number of filters from 512 to 64. The output layer is a 1x1 convolution with softmax activation, producing a multi-class output for pixel-wise segmentation. To improve model performance, spectral band selection was applied using both mRMR and DC methods to reduce input dimensionality and focus on the most informative spectral features, minimizing redundancy. This approach enhanced the model's efficiency while maintaining high segmentation accuracy, effectively handling the complexities of hyperspectral image segmentation. Overall, the U-Net model demonstrated its capacity to capture intricate spatial relationships within hyperspectral images, making it particularly effective for complex segmentation tasks with 34-class output configurations. The combined use of U-Net with spectral band selection allowed for significant improvements in both segmentation accuracy and computational efficiency.

4 Results and Evaluation

In this section, we compare the performance of the three models (MLP, DeepLabV3, and U-Net) across different band selection approaches (all spectral bands, mRMR-selected bands, and DC-selected bands). The key evaluation metrics include F1-score, accuracy, precision, and recall. Table 1 presents the performance metrics for each model and configuration.

The results in Table 1 highlight some interesting patterns in how the models performed on both the training and test sets, especially when using different band selection methods. While all three models—MLP, DeepLabV3, and U-Net—performed well on the training set, the test set results reveal challenges, particularly for the more complex models.

The MLP model stood out with the best test set performance across all band selection methods. This makes sense because MLPs are relatively simple models that tend to generalize better when data

Model	Band Selection Method	F1-Score On Train set (%)	F1-Score On Test set(%)
MLP	All Bands	80	81.11
MLP	mRMR	51	53.6
MLP	DC	73	74
DeepLabV3	All Bands	37	36.7
DeepLabV3	mRMR	82.02	18.76
DeepLabV3	DC	77.45	20.73
U-Net	All Bands	52.1	44.45
U-Net	mRMR	53.7	22.09
U-Net	DC	35	37.7

Table 1: Model Performance Comparison

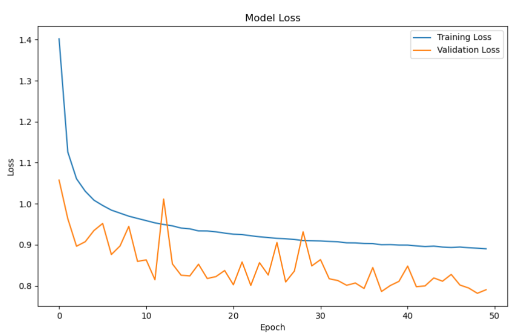


Figure 4: Results of MLP on all the bands

is limited (Figure 4). For example, using DC-based band selection improved the test set F1-score to 74%, which is only slightly lower than the 81.11% achieved with all bands. This shows that DC-based band selection effectively reduced the input size while retaining enough relevant information for the MLP to perform well.

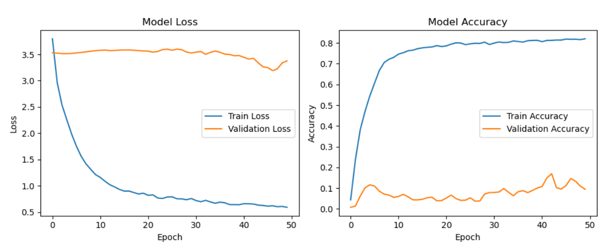


Figure 5: Results of DeepLabV3 using the mRMR

In contrast, DeepLabV3 and U-Net showed a stark drop in performance on the test set despite their strong training results. For instance, DeepLabV3 achieved an impressive F1-score of 82.02% on the training set with mRMR band selection but dropped to only 18.76% on the test set. U-Net faced a similar issue, where its F1-score with mRMR went from 53.7% on the training set to just 22.09% on the test set (Figure 5). This kind of drop suggests overfitting, where the models learned the training data well but struggled to generalize to unseen samples.

This behavior isn't surprising given the complexity of DeepLabV3 and U-Net. These models are designed to handle intricate spatial and contextual information, but they need a lot of labeled data to fully realize their potential. With the limited data in this study, they likely overfit to the training set, capturing patterns that don't generalize well. Additionally, the band selection methods, while effective for reducing input dimensionality, may have excluded features that these models rely on for their advanced spatial-contextual processing.

On the other hand, the MLP's strong performance reinforces its suitability for small datasets. Band selection methods like DC and mRMR seem to have complemented the MLP well by reducing redundancy in the data and focusing on the most important features. This balance likely helped the MLP avoid overfitting and achieve better generalization.

Overall, these results highlight the trade-offs between model complexity and data availability. While DeepLabV3 and U-Net are powerful models, they need more data to fully leverage their capabilities. In future work, expanding the dataset or using techniques like transfer learning or data augmentation could help these models perform better. At the same time, simpler models like MLP continue to be a reliable choice for scenarios where data is limited.

5 Limitations

While this study provides valuable insights into hyperspectral image segmentation, several limitations must be acknowledged.

- **Limited Labeled Data:** The small size of the labeled dataset posed challenges in achieving robust generalization, particularly for deep learning models that require large amounts of data.
- **Computational Complexity:** The high dimensionality of hyperspectral images increases the computational burden, making it difficult

to experiment with more complex models and larger datasets.

6 Conclusion

In conclusion, this study underscores the importance of semantic segmentation in hyperspectral image analysis, tackling the challenges of high dimensionality and complex spatial-spectral features. Our key contributions include comprehensive exploratory data analysis, integration of advanced band selection methods (mRMR and Deflection Coefficient), and the use of state-of-the-art models like MLP, DeepLabV3, and U-Net, demonstrating the significant impact on improving segmentation performance. The promising results indicate potential applications in fields such as agriculture, environmental monitoring, and urban planning. To further enhance model performance, future work could focus on incorporating larger and more diverse datasets to improve generalization, leveraging pretrained models for better feature extraction and reduced training time, exploring multi-task learning to increase model versatility, and using more bands (beyond the 10 bands implemented by our solution) for a more detailed analysis. Looking ahead, the segmented images could be used for more advanced tasks like land cover classification or change detection, paving the way for broader real-world applications in hyperspectral image processing.

Acknowledgments

We would like to express our sincere gratitude to our supervisor, Prof. Mehdi Zakroum, for their invaluable guidance and support throughout the course of this academic project. Their expertise and constructive feedback were crucial in shaping the direction of our work. We are also thankful to the PhD student, Ikram El Hajri, for their insightful input on the concept of deflection coefficient, which helped deepen our understanding in this area. This work would not have been possible without the collective effort and dedication from all involved.

References

[1] Yang, Xiaofei, et al. "Hyperspectral image classification with deep learning models." IEEE Transactions on Geoscience and Remote Sensing 56.9 (2018): 5408-5423.

[2] Audebert, Nicolas, Bertrand Le Saux, and Sébastien Lefèvre. "Deep learning for classification of hyperspectral data: A comparative review." IEEE geoscience and remote sensing magazine 7.2 (2019): 159-173.

[3] Minaee, Shervin, et al. "Image segmentation using deep learning: A survey." IEEE transactions on pattern analysis and machine intelligence 44.7 (2021): 3523-3542.

[4] Li, Jiaojiao, et al. "Classification of hyperspectral imagery using a new fully convolutional neural network." IEEE Geoscience and Remote Sensing Letters 15.2 (2018): 292-296.

[5] Chen, Liang-Chieh. "Rethinking atrous convolution for semantic image segmentation." arXiv preprint arXiv:1706.05587 (2017).

[6] Bajcsy, Peter, and Peter Groves. "Methodology for hyperspectral band selection." Photogrammetric Engineering & Remote Sensing 70.7 (2004): 793-802.

[7] Zhao, Zhenyu, Radhika Anand, and Mallory Wang. "Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform." 2019 IEEE international conference on data science and advanced analytics (DSAA). IEEE, 2019.

[8] Taud, Hind, and Jean-Franccois Mas. "Multilayer perceptron (MLP)." Geomatic approaches for modeling land change scenarios (2018): 451-455.