Deep Learning-Based Grassland Mapping with Sentinel-2: Prioritizing Key Spectral Bands and Time Periods

Konstantinos Christofi[a,b], Charalambos Chrysostomou[a], Iason Tsardanidis[c], Michalis Mavrovouniotis[a], Charalampos Kontoes[c], and Diofantos G. Hadjimitsis[a,b]

[a]ERATOSTHENES Centre of Excellence, 82 Franklin Roosevelt, 3012, Limassol, Cyprus
[b]Dept. of Civil Engineering and Geomatics, Cyprus University of Technology, Limassol, Cyprus
[c]Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Athens, Greece

ABSTRACT

Accurate grassland mapping is essential for biodiversity conservation and sustainable land management, yet remains challenging due to the spectral and temporal variability of grassland ecosystems. This study presents a Deep Learning approach for grassland classification using multi-temporal Sentinel-2 imagery, incorporating a dynamic feature selection mechanism to prioritize informative spectral bands and time periods. In order to allow the model to adaptively focus on discriminative temporal spectral patterns, we compare a baseline neural network with a modified design that learns to weight input features dynamically. Our findings demonstrate that the feature selection model achieves superior performance (Accuracy: 0.954 ± 0.004, MCC: 0.726 ± 0.027) compared to both the baseline network and single-date models, highlighting the importance of temporal diversity in grassland classification.

Keywords: Grassland Classification, Remote Sensing, Machine Learning, Sentinel-2 Optical Imagery, Feature Selection

1. INTRODUCTION

Grasslands are among the most critical terrestrial ecosystems, supporting a variety of plant and animal species, and are essential for carbon sequestration, and contributing to soil conservation and water cycle regulations Bengtsson et al. (2019); Boval and Dixon (2012); Ojima et al. (1993). In order to maintain biodiversity and encourage sustainable land use, European policies like the Common Agricultural Policy (CAP) emphasize on protecting permanent grasslands d'Andrimont et al. (2018). Therefore, accurate grassland classification and monitoring are essential for ecological forecasting, climate policy implementation, and informed environmental management.

However, grasslands using remote sensing data still remains quite difficult. The spectral signatures of the diverse vegetation species that compose grassland environments can vary across both space and time. Particularly when employing conventional techniques or static representations of vegetation, seasonal dynamics, environmental disturbances, and subtle spectral differences between vegetation types, complicates the classification process Weeks et al. (2013); Yong et al. (2003).

To address these challenges, recent studies have explored the use of satellite-based Earth Observation combined with Machine Learning (ML) and Deep Learning (DL) algorithms. The Copernicus Program's Sentinel-2 satellite offers high-resolution optical imagery across 13 spectral bands in the visible, red-edge, near-infrared (NIR), and shortwave-infrared (SWIR) regions. These bands offer detailed insights into vegetation health, structure, and growth cycles. Leveraging this spectral richness enables improvements in detection of subtle vegetation differences that are relevant to grassland classification.

Further author information: (Send correspondence to C.C.)
C.C.: E-mail: charalambos.chrysostomou@eratosthenes.org.cy, Telephone: +357 25 002908

Eleventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2025), edited by Andreas Christofi, et al., Proc. of SPIE Vol. 13816 · · doi: 10.1117/12.3073215

In this study, we expand on this basis by classifying grassland versus non-grassland parcels using multi-temporal Sentinel-2 reflectance data with Deep Neural Networks (DNN). Unlike prior work that often relies on static or handcrafted feature selection, we incorporate a dynamic feature selection mechanism into our model. This mechanism learns to weight spectral-temporal features during training, effectively identifying the most informative observations across bands and dates. By dynamically prioritizing relevant input features, the network is better equipped to generalize across temporal variability and spectral redundancy.

In order to acquire a better understanding of the contribution of temporal information, we also compare our proposed model to a baseline design that does not have any feature selection mechanism. Additionally, we evaluate the performance of a series of single-date networks to better understand the contribution of temporal information. This study expands on our earlier work Christofi et al. (2025), which evaluated Sentinel-1 and Sentinel-2 data separately, by focusing exclusively on optical time-series inputs and embedding feature relevance learning directly into the model architecture.

Through this approach, we aim to not only improve the accuracy of grassland classification, but also to better understand which spectral bands and observation dates are most important for identifying grassland areas in real-world monitoring scenarios.

2. LITERATURE REVIEW

Grasslands are ecologically dynamic systems whose spectral signatures differ significantly with phenology, management practices, and environmental conditions Dusseux et al. (2022). According to studies, the optimal temporal selection can significantly improve the accuracy of grassland classification with peak recognition rates occurring during mid-August for natural and artificial grasslands Fengli and Qiu (2006).

Because of their interpretability and effectiveness and efficiency with handcrafted features, traditional Machine Learning (ML) method like Random Forest (RF) and Support Vector Machines (SVM) have dominated early grassland mapping classification techniques Belgiu and Drăguţ (2016). However, because RFs ignore spatio-spectral correlations and treat each band-date combination as an independent feature, they struggle with high-dimensional, multi-temporal data Nalepa et al. (2019). Deep Learning (DL) approaches excel at automating feature extraction from raw Sentinel-2 data. Helber et al. (2019) demonstrated that CNNs trained on the EuroSAT dataset (Sentinel-2 patches) achieved 98% accuracy in land cover classification by leveraging spatial-spectral hierarchies. Furthermore, unsupervised deep learning techniques can automate feature extraction, potentially bridging the gap between naive approaches and those requiring domain expertise Merentitis and Debes (2015). Yet, most DL studies treat all input bands or timesteps equally, lacking mechanisms to adaptively prioritize discriminative features-a gap our work addresses.

Dynamic feature selection has emerged as a promising technique in remote sensing to improve classification and image analysis by adjusting to varying spatial structures and sensor characteristics. Li et al. (2019) suggested a classifier selection framework based on imprecise probabilities to better handle uncertainty in remote sensing imagery. A spatial dynamic selection network for image fusion tasks was also introduced by Hu et al. (2022). The network in this study used dynamic feature extraction modules to adjust to spectral and structural variation at the pixel level. In order to overcome issues like high-dimensional input spaces and varied spatial patterns, these works collectively highlight the importance of incorporating dynamic feature selection techniques in remote sensing pipelines. These ideas also serve as the driving forces behind the design suggested in this study.

3. DATA

3.1 Data Overview

This study utilizes a dataset provided by Choumos et al. (2022) from the National Observatory of Athens (NOA). The dataset includes georeferenced and labeled Sentinel-2 imagery, annotated with crop-type information from the Dutch Land Parcel Identification System (LPIS). The area of interest (AOI) is located in Utrecht, The Netherlands, covering the time period from March to October 2017.

Sentinel-2 provides high-resolution optical imagery across 13 spectral bands, including visible (VIS), near-infrared (NIR), and short-wave infrared (SWIR) regions (see Table 1). These bands are designed for a range of Earth Observation applications, including vegetation monitoring, water detection, and atmospheric analysis. Each observation in the dataset is uniquely identified and labeled to indicate whether the region belongs to the *Grassland* class or not.

Table 1. Sentinel-2's 13 spectral bands

Band	**Function**
B1	Coastal Aerosol
B2	Blue
B3	Green
B4	Red
B5	Red-edge
B6	Red-edge
B7	Red-edge
B8	NIR
B8a	Red-edge
B9	Water vapour
B10	SWIR
B11	SWIR
B12	SWIR

In this work, all bands are utilized except for B1 (Coastal), B9 (Water Vapor), and B10 (Cloud Detection), which are primarily suited for atmospheric correction rather than land cover classification Ali and Johnson (2022).

3.2 Input Representation and Class Weighting

The dataset consists of 37,041 samples, each representing a unique land parcel being labeled as either *Grassland* or $Non-Grassland$. With 31,350 samples ($\sim$ 84.63%) in the *Grassland* class and 5,691 ($\sim$ 15.36%) in the $Non-Grassland$ class, the dataset is unbalanced.

Each sample is described using 290 features, derived from 10 selected Sentinel-2 spectral bands captured across 29 distinct dates between March and October 2017. This results in a temporal-spectral input structure of [10 bands x 29 dates]. The features reflect the surface reflectance values of each band over time, enabling the model to learn temporal and spectral patterns that are important for classifying grasslands.

To address the class imbalance, class weights are computed and applied during training to penalize the underrepresented class more. The weights are calculated as the inverse of the normalized class frequency:

$$w_i = \frac{1}{\frac{n_i}{N}}, \tag{1}$$

where n_i is the number of samples in class i, and N is the total number of training samples. These weights are then converted into PyTorch tensors and passed to the loss function to ensure balanced learning despite the skewed distribution.

4. METHODOLOGY

4.1 Data Preprocessing

This Sentinel-2 optical data that are used in this study, are already georeferenced and matched to individual land parcels to ensure spatial consistency. Since the dataset contained no missing values, no imputation techniques were needed. To prepare the input features for model training, we applied the *RobustScaler* from *scikit-learn*, which normalizes the data by removing the median and scaling according to the inter-quartile range. This approach is particularly effective for handling potential outliers in satellite reflectance values. The data was split into training, validation, and test sets using an $80:10:10$ ratio. Additionally, we employed a 10-fold cross-validation scheme to assess the model's robustness across different subsets of the data.

4.2 Neural Network Design

4.2.1 Without Feature Selection Mechanism

To perform the classification of $Grassland$ vs $Non - Grassland$ areas, we implemented a fully connected feed-forward neural network using PyTorch Paszke (2019). The model has 290 input features, which represent multi-temporal spectral reflectance values across different Sentinel-2 bands and acquisition dates. The architecture includes five sequential linear layers, which sizes progressively reducing from 290 to 256, 128, 64, and 32, each followed by a Tanh activation and a Dropout layer (rate 0.2) to mitigate overfitting. The final output layer consists of two neurons, producing class scores for the two output categories. This design enables the model to learn complex patterns in the input data and make effective predictions across the defined classes.

4.2.2 With Feature Selection Mechanism

In addition to the baseline neural network architecture, we implemented a second model that incorporates a **dynamic feature selection layer** directly after the input layer (See fig. 1). This trainable layer learns a weight vector that scales each of the 290 input features, allowing the network to dynamically prioritize more informative temporal-spectral features while ignoring less relevant ones. Since each feature corresponds to a specific band value at a particular date, the model is able to discern which observations across time contribute most effectively to grassland classification.

Following the feature selection layer, the architecture mirrors the baseline model, a sequence of fully connected layers with Tanh activations and a Dropout rate of 0.2. The network concludes with a two-neuron output layer, producing predictions for the classification task: $Grassland$ vs $Non - Grassland$.

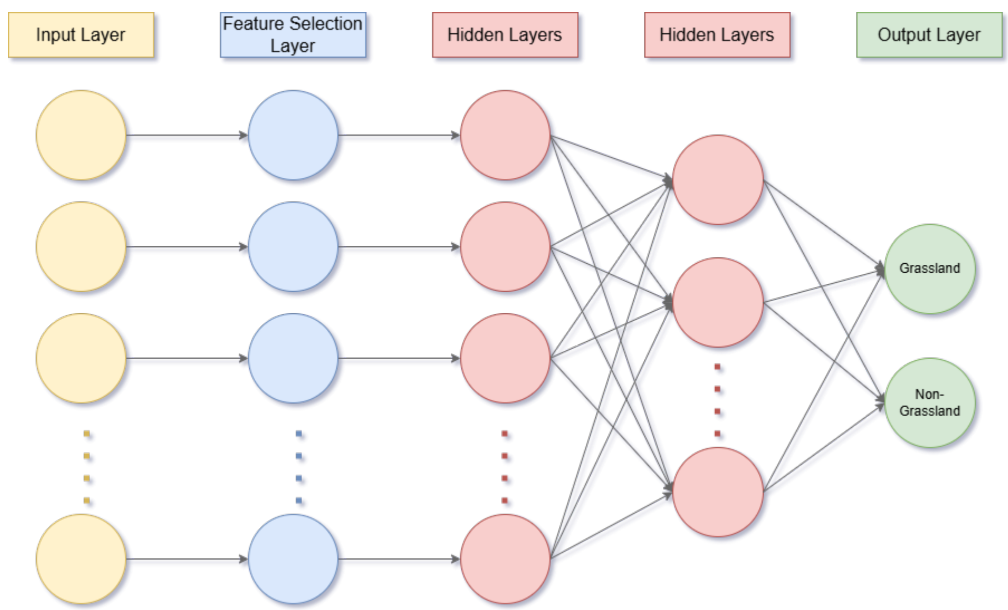

Figure 1. Neural network architecture with an integrated feature selection layer that learns to weight input features dynamically, followed by hidden layers and an output layer for grassland classification.

4.3 Evaluation Setup and Metrics

All neural network models in this study were trained using the RAdam optimizer, selected for its robust convergence properties and adaptive learning handling. The classification task was optimized using the cross-entropy loss function. Training was conducted for 1000 epochs with a batch size of 128. Model performance was monitored throughout, and the best model checkpoint was selected based on the highest achieved Matthews Correlation Coefficient (MCC) on the validation set.

This study focused exclusively on Sentinel-2 data and compared the performance of two neural network architectures: a baseline model and a model incorporating a dynamic feature selection mechanism. Both models were evaluated using accuracy, loss, and MCC metrics to comprehensively assess classification performance.

All experiments were executed on a high-performance computing server equipped with dual *AMD EPYC 7452 32-Core processors*, four *NVIDIA A40 GPUs*, and *512 GB* of *RAM*. However, not all hardware resources were utilized simultaneously for every training run. The models were implemented in PyTorch, leveraging GPU acceleration where available.

5. RESULTS

5.1 Performance Comparison of Networks With and Without Feature Selection

The performance comparison between the neural network with the dynamic feature selection mechanism and the baseline network without it, is presented in Table 2. Both networks achieved similar classification accuracy, with the feature selection model performing slightly better than the baseline model (0.954 vs. 0.953). However, the network with the feature selection mechanism demonstrated clear advantages in terms of average loss (0.022 vs. 0.095) and slightly higher average Matthews Correlation Coefficient (MCC) (0.726 vs. 0.721), suggesting better generalization and confidence in its predictions.

These findings suggest that while the feature selection mechanism does not significantly boost accuracy, it contributes to improved model robustness and confidence, as reflected by the lower variance and lower loss.

Table 2. Performance comparison of the neural network with and without the dynamic feature selection mechanism. Metrics are averaged over test sets.

Metric (Average Test)	**Model**	
	With Feature Selection Mechanism	**Without Feature Selection Mechanism**
Accuracy	0.954 ± 0.004	0.953 ± 0.003
MCC	0.726 ± 0.027	0.721 ± 0.020
Loss	0.022 ± 0.045	0.095 ± 0.048

5.2 Temporal Importance Distribution

To investigate how the model interprets the temporal dynamics of input data, we analyzed the feature importance weights learned for each spectral band across all dates. Figure 2 shows the distribution of importance scores for Band 12. According to the analysis, the importance varies substantially in value between dates with the same band, suggesting that temporal context is plays a key role in classification.

For instance, Band 12 (see Fig. 2) exhibits notable variation in importance, with decreases observed around mid-August and early September. These patterns may correspond to underlying phenological transitions or environmental changes during that period. While some bands display complex or fluctuating importance profiles, this variability underscores the benefit of using a dynamic feature selection mechanism—allowing the model to adaptively emphasize informative features over time, even in the presence of temporal heterogeneity. Rather than relying on fixed or manual feature selection, this approach captures subtle spectral-temporal cues that may otherwise be overlooked, though further research is needed to deepen interpretability and link specific patterns to ecological phenomena.

These findings underline the complexity of remote sensing time-series data and validate the usefulness of mechanisms that adaptively learn to weigh temporal features.

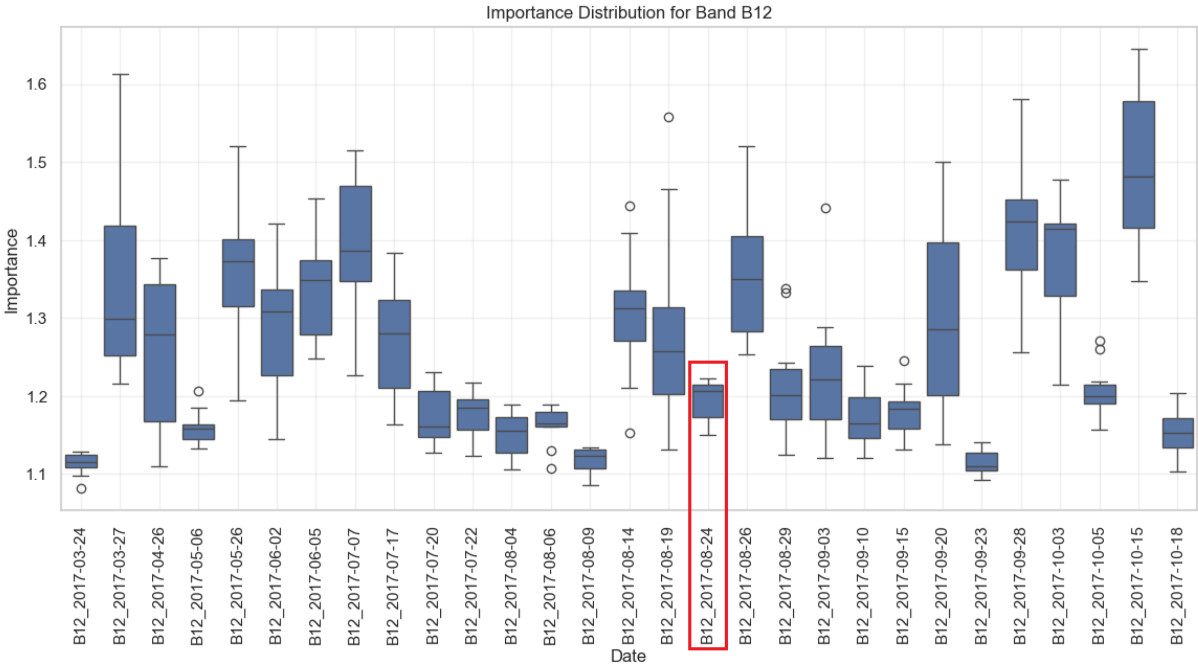

Figure 2. Importance distribution for Band 12. The red box highlights the date with the best accuracy.

5.3 Comparison with Single-Date Model

To understand the contribution of temporal diversity in the dataset, we conducted an experiment in which a separate neural network was trained using only the spectral values from each individual date. Each model received as input only the 10 spectral bands corresponding to a single acquisition date, thereby eliminating the influence of temporal variation.

Following the training of these single-date models, we compared their test accuracies to identify the most informative date. The model trained on data from *24/08/2017* achieved the highest accuracy among all single-date networks. We then compared the performance of this best-performing single-date model with the multi-date model that utilized the full temporal sequence along with the dynamic feature selection mechanism.

Table 3 shows that while the best single-date model achieved reasonably strong performance (Accuracy: 0.916 $\pm$ 0.003, MCC: 0.672 $\pm$ 0.012), it was still outperformed by the multi-date model (Accuracy: 0.954 $\pm$ 0.004, MCC: 0.726 $\pm$ 0.027). This emphasizes how important it is to incorporate temporal diversity across the growing season, as it increases the classification performance and the model's capacity to generalize.

Table 3. Performance of the single-date neural network model (24/08/2017) compared to the full multi-date model with feature selection.

Metric	**Model**	
(Average Test)	**With Feature Selection Mechanism**	**Date with best accuracy - *24/08/2017***
Accuracy	0.954 $\pm$ 0.004	0.916 $\pm$ 0.003
MCC	0.726 $\pm$ 0.027	0.672 $\pm$ 0.012
Loss	0.022 $\pm$ 0.045	0.014 $\pm$ 0.002

6. DISCUSSION

6.1 Summary of Key Findings

This study explored the impact of temporal information and dynamic feature selection in neural network-based grassland classification using Sentinel-2 data. We developed and compared two neural network architectures: one standard model and another enhanced with a dynamic feature selection mechanism. The model incorporating feature selection slightly outperformed the baseline in terms of accuracy, Matthews Correlation Coefficient (MCC), and test loss, indicating its effectiveness in prioritizing relevant temporal features.

An analysis of the learned importance scores showed that spectral bands varied in relevance across different dates, confirming that temporal fluctuations play a critical role in grassland classification. Notably, the date with the highest learned importance - *24/08/2017* - also achieved the best performance when used alone in a separate single-date model. However, even this best-performing single-date model fell short of the multi-date neural network with feature selection, highlighting the advantage of integrating temporal information rather than relying on isolated observations.

6.2 Interpretation of Results

The inclusion of the feature selection layer, the network was able to give each input feature a dynamic weight, which helped it concentrate on spectral data that was relevant to time. This contributed to more effective learning and reduced overfitting, as evidenced by the lower test loss. The relatively small performance gap between the two models shows that while the base architecture is strong, the feature selection mechanism provides a noticeable benefit.

The analysis of importance values across dates showed that no single band or date dominated across the board. Instead, there were fluctuations that likely reflect underlying vegetation and environmental dynamics. For instance, Band 12 showed significant fluctuations in importance around mid-August and September, possibly linked to stress periods or phenological changes in the vegetation.

Moreover, training individual models for each date provided insight into the value of temporal information. The best single-date model performed well but failed to much the multi-date model with feature selection. This further emphasizes that temporal integration is critical for robust classification and that relying on a single date - even the most informative one - is suboptimal.

6.3 Limitations and Future Work

While the feature importance analysis revealed variability across bands and dates, some of these fluctuations-such as the decreased importance of Band 12 in mid-August—may correspond to phenological changes like senescence or reduced vegetation vigor. Though currently difficult to confirm without ground truth, such temporal shifts suggest an underlying ecological signal that could be investigated further. Future work could enhance interpretability by linking importance trends to phenological events documented through field observations or ecological calendars. This would help validate the dynamic feature selection mechanism and improve its practical utility in monitoring seasonal grassland dynamics.

7. CONCLUSION

This study explored the use of a neural network-based approach for grassland classification using Sentinel-2 multi-temporal spectral data. The model's robustness and classification accuracy were enhanced by adding a dynamic feature selection mechanism that allows it to learn the relative importance of each spectral observation over time. Our results show the importance of integrating temporal diversity for vegetation analysis by demonstrating that this approach consistently outperforms single-date models.

Through a comparative analysis, we have found that the neural network with the feature selection not only achieved higher accuracy and MCC scores, but also provided insights into which temporal features were most important in the model's predictions. However, the temporal importance distributions were not always straightforward to interpret, highlighting the complexity of phenological patterns in grassland ecosystems.

Despite strong overall, performance, limitations such as the model's interpretability and focus on optical data alone point to a number of areas that require more research. Expanding the analysis to include radar data (e.g. Sentinel-1), testing across broader spatial and temporal domains, and introducing interpretable architectures like attention mechanisms could further enhance model generalizability and explainability.

Ultimately, this work contributes a scalable, data-driven method for grassland monitoring that leverages deep learning and temporal remote sensing data to support more informed environmental management techniques.

ACKNOWLEDGMENTS

This work was supported by the European Union's HORIZON Research and Innovation Programme by the 'EXCELSIOR': ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment H2020 Widespread Teaming project (www.excelsior2020.eu). The 'EXCELSIOR' project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857510, from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination and Development and the Cyprus University of Technology.

References

Ali, K. and Johnson, B. A. (2022), 'Land-use and land-cover classification in semi-arid areas from medium-resolution remote-sensing imagery: A deep learning approach', *Sensors* **22**(22), 8750.

Belgiu, M. and Drăguţ, L. (2016), 'Random forest in remote sensing: A review of applications and future directions', *ISPRS journal of photogrammetry and remote sensing* **114**, 24–31.

Bengtsson, J., Bullock, J., Egoh, B., Everson, C., Everson, T., O'connor, T., O'farrell, P., Smith, H. and Lindborg, R. (2019), 'Grasslands—more important for ecosystem services than you might think', *Ecosphere* **10**(2), e02582.

Boval, M. and Dixon, R. (2012), 'The importance of grasslands for animal production and other functions: a review on management and methodological progress in the tropics', *Animal* **6**(5), 748–762.

Choumos, G., Koukos, A., Sitokonstantinou, V. and Kontoes, C. (2022), Towards space-to-ground data availability for agriculture monitoring, *in* '2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)', pp. 1–5.

Christofi, K., Chrysostomou, C., Tsardanidis, I., Mavrovouniotis, M., Guerrisi, G., Kontoes, C. and Hadjimitsis, D. (2025), Remote sensing of grasslands: Performance comparison of radar and optical data in machine learning classification, *in* 'International Society for Photogrammetry and Remote Sensing'.

Dusseux, P., Guyet, T., Pattier, P., Barbier, V. and Nicolas, H. (2022), 'Monitoring of grassland productivity using sentinel-2 remote sensing data', *International Journal of Applied Earth Observation and Geoinformation* **111**, 102843.

d'Andrimont, R., Lemoine, G. and Van der Velde, M. (2018), 'Targeted grassland monitoring at parcel level using sentinels, street-level images and field observations', *Remote Sensing* **10**(8), 1300.

Fengli, Z. and Qiu, Y. (2006), 'Optimal temporal selection for grassland spectrum classification', *Journal of remote sensing* .
URL: *https://api.semanticscholar.org/CorpusID:124561438*

Helber, P., Bischke, B., Dengel, A. and Borth, D. (2019), 'Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **12**(7), 2217–2226.

Hu, J., Hu, P., Wang, Z., Kang, X., Fan, S. and Mao, D. (2022), 'Spatial dynamic selection network for remote-sensing image fusion', *IEEE Geoscience and Remote Sensing Letters* **19**, 1–5.
URL: *https://api.semanticscholar.org/CorpusID:236653715*

Li, M., Huang, S. and Pi˘zurica, A. (2019), Robust dynamic classifier selection for remote sensing image classification, *in* '2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)', IEEE, pp. 101–105.

Merentitis, A. and Debes, C. (2015), 'Automatic fusion and classification using random forests and features extracted with deep learning', *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* pp. 2943–2946.
URL: *https://api.semanticscholar.org/CorpusID:637712*

Nalepa, J., Tulczyjew, L., Myller, M. and Kawulok, M. (2019), 'Segmenting hyperspectral images using spectral-spatial convolutional neural networks with training-time data augmentation', *arXiv preprint arXiv:1907.11935* .

Ojima, D. S., Dirks, B. O., Glenn, E. P., Owensby, C. E. and Scurlock, J. O. (1993), 'Assessment of c budget for grasslands and drylands of the world', *Water, Air, and Soil Pollution* **70**, 95–109.

Paszke, A. (2019), 'Pytorch: An imperative style, high-performance deep learning library', *arXiv preprint arXiv:1912.01703* .

Weeks, E. S., Ausseil, A.-G. E., Shepherd, J. D. and Dymond, J. R. (2013), 'Remote sensing methods to detect land-use/cover changes in new zealand's 'indigenous' grasslands', *New Zealand Geographer* **69**, 1–13.
URL: *https://api.semanticscholar.org/CorpusID:128710565*

Yong, Z., Jay, G. and Shaoxiang, N. (2003), 'Most recent progress of international research on remote sensing of grassland resources', *Progress in Geography* **22**(6), 607–617.