

# Multimodal remote sensing and machine learning for sand dune classification in homogeneous environments: A case study from southern Morocco

Nawfel El Bouchti<sup>1\*</sup>, Nadya Wahid<sup>1</sup>, Khalid Benhssaine<sup>2,3</sup> ,  
Ettaqy Abderrahim<sup>2</sup>, Jamal Aabdousse<sup>1,4</sup> 

<sup>1</sup> Environmental, Ecological and Agro-Industrial Engineering Laboratory, Department of Life Science, Faculty of Sciences and Technics, Sultan Moulay Slimane University, Beni Mellal, Morocco

<sup>2</sup> Polydisciplinary Faculty of Beni Mellal, Sultan Moulay Slimane University Beni Mellal, BP592, 23000, Beni Mellal, Morocco

<sup>3</sup> National Agency for Water and Forests, Béni Mellal–Khénifra Regional Directorate, Morocco

<sup>4</sup> Higher School of Technology Fkih Ben Salah, Sultan Moulay Slimane University, Tighnari District, National Road N11 from Casablanca, Fkih Ben Salah, PB. 336, Morocco

\* Corresponding author's e-mail: [elbouchtinawfel@gmail.com](mailto:elbouchtinawfel@gmail.com)

## ABSTRACT

The realistic mapping of dune landscapes is necessary to model dune dynamics, but traditional remote sensing has been generally unimodal, and incapable of guaranteeing a sufficient representation of dune spatial complexity. In this study, we used a multimodal methodology combining Sentinel-1 SAR and Sentinel-2 optical data with five classification algorithms: the Random Forest, the LightGBM, the XGBoost, the support vector machines and the extra trees. We evaluated SAR-only, optical-only, and SAR–optical fusion inputs with spatial cross-validation and morphological post-processing. It is evaluated by accuracy, F1-score, IoU and Matthews correlation coefficient, and other spatial uncertainty analysis. Results demonstrate that fusion strongly boosts performance (37–40% higher F1-scores with respect to the SAR-only and 1–2% with respect to the optical-only inputs). The Random Forest and LightGBM had the highest performance (F1 = 0.725–0.735). A morphological post-processing yielded IoU improvements in 3.4% on average for the purpose of improving spatial coherence. This study proves that SAR–optical fusion is an effective scheme of dune classification and it is also useful for desertification risk assessment and arid landscape engineering.

**Keywords:** dune mapping, multi-modal remote sensing, machine learning classification, spatial uncertainty analysis.

## INTRODUCTION

Automatic mapping of sand dunes from remote sensing data constitutes a scientific difficulty on geomorphology, environmental monitoring and land management (Yekeen et al., 2023). Dunes in deserts are very dynamic landscapes, which are shaped through wind-blown techniques, whose morphological variation occurs at various spatial and temporal scales. These environments are extremely fragile ecosystems but also constitute an important environmental problem, as “advancing” dunes are a threat for infrastructures,

agricultural fields and human settlement in arid and semiarid regions. Reproducible and consistent mapping of the escarpment is therefore important for sand encroachment dynamic monitoring, for land-use planning, as well as for the planning of the environmental management (Lu et al., 2023).

Sand encroachment is regarded as one of the most serious environmental problems in arid regions. The dynamics of these dunes, shaped by the wind and climate fluctuations, also leads to the progressive burying of agricultural fields, roads, and occasionally even built-up areas (Salem et

al., 2025). This leads to faster land degradation, reduced agricultural production and high costs for infrastructure up-keep. Given that the current aridification and wind patterns are expected to be intensified by climate change, sand transport between dunes may become more prevalent. Therefore, few dune field monitoring approaches are necessary to study dune fronts evolution and migration processes, and human pressures, including overgrazing and resource harvesting. For planners and land managers, the requirement for robust spatial information is equally critical to plan mitigation strategies: from shelterbelt establishment, dune stabilisation and protection of 'at-risk' settlements (Stammler et al., 2023).

Nevertheless, the automatic identification of dunes and interdunes is still an open issue. There is clear spatial heterogeneity in dune systems: they are made up of barchans, transverse ridges, linear and compound forms; they can be of bare sand, gravel pavement or sparse vegetated surface; and one of the compelling evidences of the latter point mentioned above is that they are constantly being affected by both seasonal and interannual winds (Tang et al., 2023a). Furthermore, the transition from dune, interdune to general surrounding topography is transitional and indistinct, making the separation of crisp boundaries more difficult. Classical mapping methods—including identification of LULC classes via manual interpretation of satellite images—can produce reliable results in local contexts, are subjective by definition, and are inefficient for systematic monitoring at the larger scales. Similarly, traditional parametric classifiers like ML which is based on assumptions of simplified statistical model are also unable to characterize the highly nonlinear and non-homogeneous nature of desert-like areas. Remote sensing sensors taken individually are also hampered by limitations: optical imagery is sensitive to atmospheric and illumination conditions, and radar data are affected by speckle noise and may not be as spectrally rich to differentiate subtle land cover differences. These constraints frequently result in misclassification of transition zones and the comparability of results between various dune fields is negatively impacted (Smyth et al., 2022).

Multimodal RS provides a hopeful direction to tackle these problems jointly. The optical and radar strengths are complementary, able to fill each other's gaps in coverage. Sentinel-1 synthetic aperture radar (SAR) is highly sensitive to surface roughness and soil moisture and these

are recognised to be important properties for dune texture and interdune status (Smyth et al., 2022). In contrast, Sentinel-2 multiband imagery also visually shows minor spectral differences required to distinguish between bare sand, crust and light vegetation. These both can be used to utilize both spectral, structural information, and to minimize the class ambiguity, they are robust to illuminant variance. Pixel-level fusion between SAR and optical features resulted in an enhancement of a more assimilation of dune environments, with greater generalisation of the classification models on different geomorphological envelopes. The free provision of (soil moisture data in) large scale and high resolution (10 m) and high accuracy dune data by the Copernicus program also increases the feasibility for operational dune monitoring on multiple spatial scales (Sieren, 2022).

With these improvements in observational data, advances in machine learning have increasingly developed the means to account for the complexity of desert landscapes. The nonlinear spectral, noisy signals and high spatial variability properties of Dune environments make it difficult for conventional classifiers (Feng et al., 2022). As such, machine learning algorithms (in specific ensemble and boosting algorithms) are natural choices to solve these hurdles. For instance, random forest, XGBoost and LightGBM have shown high performance for heterogeneous and complex landscapes, because the methods can capture the non-linearity, combine multimodal inputs and tolerate noises. There is also growing availability of high-performance computing facilities which enables machine learning to be applied to large scale remote sensing datasets. Optimized pipelines lead to efficient and reproducible workflows, while new methods, like spatial cross-validation, give a more realistic performance estimates by incorporating the spatial autocorrelation characteristic of geomorphological data (Iban and Bilgilioglu, 2023).

Exploiting these possibilities, in this work we introduce a new framework for the automatic classification of desert dune environment from multidata and machine learning perspectives. More specifically, this work has five main aims: () to compare the performance of five machine learning algorithms – random forest, LightGBM, XGBoost, support vector machines, and extra trees – under both unimodal (SAR, optical) and multimodal scenarios; (i) to quantify the added value of SAR–optical fusion in terms of enhanced

classification accuracy and robustness; (ii) to develop a spatial cross-validation methodology tailored to geomorphological data; (iii) to optimize the classification pipeline for various heterogeneous computing architectures; and (iv) to integrate spatial uncertainty metrics and morphological post-processing in order to improve map reliability and interpretability (Tang et al., 2023b).

The original contribution of this work is four-fold: (i) it is the first holistic application of Sentinel-1 and Sentinel-2 fusion for dune types classification in the desert or Sahel, with the validation at the spatially explicit scale. Second, a technically optimized workflow is enabled for scalable and reproducible processing of the data and it can be applied to regional scales. Third, it promotes a new approach employing spatial uncertainty measures to assess classification confidence and in the explanation process. Finally it results an intensive methodological protocol and is immediately applicable for operational desertification monitoring, assessment of the risk of sand encroachment and for the geomorphological study of desert landscapes. These contributions serve not only to propel forward methodological innovation, but also make substantial contributions to science and practice in arid land management.

## MATERIALS AND METHODS

### Study area

The study area is located in southern Morocco and consists of a dune field with barchan dunes, vegetated interdunes and well-defined slipface transitions. The reason for choosing this site was its relatively uniform geomorphological features, little anthropogenic disturbance and ready availability of high resolution satellite data. Because of persistent NE trade winds, dunes migrate in this area and provide a natural environment for studying multimodal classification (Foody, 2024). The geographic coverage of the study area is depicted in Figure 1, and delineates the dune system where this work was carried out.

### Data sources

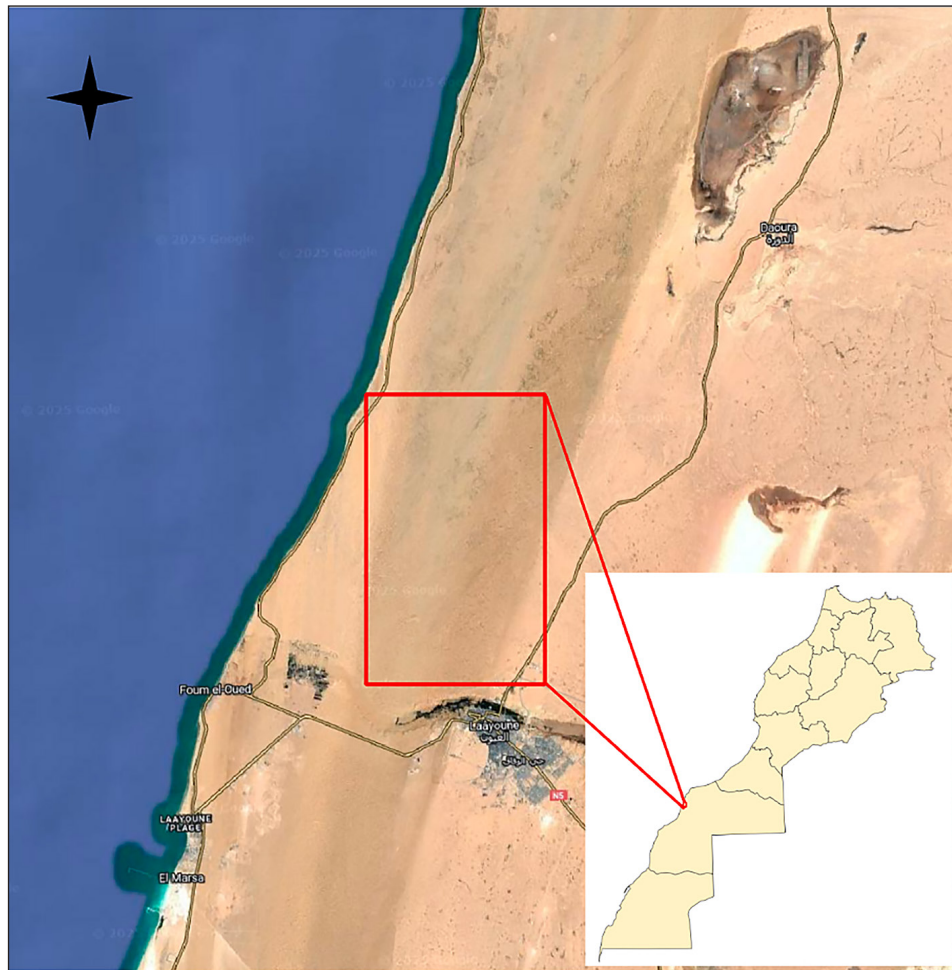
The study utilised the multi-source data provided by the Copernicus program, including SAR and optical imagery (Fei et al., 2024), as well as detailed ground reference data. Table 1 presents

the datasets employed. Sentinel-1 C-band SAR data were acquired in the Interferometric Wide Swath mode, at a resolution of 10 meters and with dual polarization including VV and VH. Preprocessing of the data included radiometric calibration to sigma nought backscatter coefficients, terrain correction using the Shuttle Radar Topography Mission digital elevation model, speckle filtering with the Refined Lee and geometric co-registration with Sentinel-2 for pixel-level compositing. Sentinel-2 Level-1C multispectral products were atmospherically corrected using Sen2Cor version 2.10, cloud masked to keep only scenes with less than five percent cloud cover, orthorectified and pansharpened to 10-meter resolution over the entire area. We show the raw input from both sensors in Figure 2 where the Sentinel-1 SAR dual-polarization image and the corresponding Sentinel-2 multispectral scene are visualized before any preprocessing.

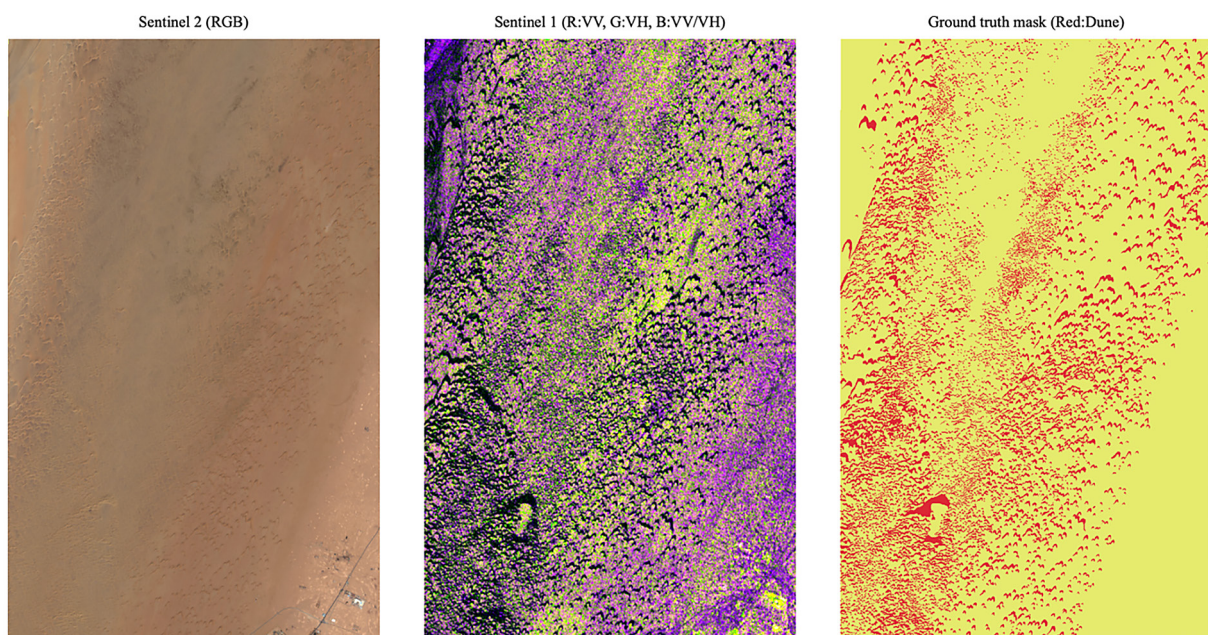
Reference mask was produced by the authors, manually interpreting false color composition on optical sentinel 2 imagery, and partially ground-truthed in local field sites. Class one referred to dunes and class zero the interdunes. Field campaigns checked the confidence of interpreted polygons, which was done to minimize the subjectivity in the reference data set. Seventy percent of the final binary mask was used as model training data and thirty percent for independent validation.

### Feature engineering

Feature extraction was aimed at utilizing the complementary information between radar backscatter and optical reflectance. In the case of SAR dataset, predictors were VV and VH polarizations, as well as their ratio that gives sensitivity to surface roughness and soil moisture. The optical dataset was made up of the 6 Sentinel-2 bands that were both in the visible, near-infrared and shortwave infrared domains, with the addition of 2 normalized indices: Normalized difference vegetation index and normalized difference sand index. These attributes were chosen to reflect differences in vegetation cover, soil makeup and dune shape. The fusion dataset integrated all predictors originating from SAR and optical data sources, representing structural and spectral components into a unique multimodal feature space. The full feature makeup is given in Table 2.



**Figure 1.** Location of the study area in southern Morocco, highlighting the Oued Sakia El Hamra region investigated for sand dune dynamics



**Figure 2.** Raw input data used in the analysis: (a) Sentinel-2 multispectral scene (RGB composite), (b) Sentinel-1 SAR dual-polarization (VV, VH), and (c) ground-truth mask (dunes = 1, interdunes = 0). These inputs form the basis for the classification experiments.

**Table 1.** Satellite datasets used in this study, including product identifiers, acquisition dates, processing levels, and spatial resolution

| Sensor      | Product ID                                                               | Date         | Mode/Level | Resolution |
|-------------|--------------------------------------------------------------------------|--------------|------------|------------|
| Sentinel-1A | S1A_IW_GRDH_1SDV_20240927T064656_20240927T064720_055847_06D33F_OC8B.SAFE | 27 Sept 2024 | IW GRDH    | 10 m       |
| Sentinel-2B | S2B_MSIL1C_20240923T113319_N0511_R080_T28RFR_20240923T151308.SAFE        | 23 Sept 2024 | L1C        | 10 m       |

**Table 2.** Features composition by dataset

| Dataset | SAR                | OPT               | FUSION             |
|---------|--------------------|-------------------|--------------------|
| Feat1   | VV (Sentinel 1)    | B2 (Sentinel 2)   | B2 (Sentinel 2)    |
| Feat2   | VH (Sentinel 1)    | B3 (Sentinel 2)   | B3 (Sentinel 2)    |
| Feat3   | VV/VH (Sentinel 1) | B4 (Sentinel 2)   | B4 (Sentinel 2)    |
| Feat4   | -                  | B8 (Sentinel 2)   | B8 (Sentinel 2)    |
| Feat5   | -                  | B11 (Sentinel 2)  | B11 (Sentinel 2)   |
| Feat6   | -                  | B12 (Sentinel 2)  | B12 (Sentinel 2)   |
| Feat7   | -                  | NDVI (Sentinel 2) | NDVI (Sentinel 2)  |
| Feat8   | -                  | NDSI (Sentinel 2) | NDSI (Sentinel 2)  |
| Feat9   | -                  | -                 | VV (Sentinel 1)    |
| Feat10  | -                  | -                 | VH (Sentinel 1)    |
| Feat11  | -                  | -                 | VV/VH (Sentinel 1) |

All continuous features were standardized using z-score normalization, while indices were rescaled to the interval from zero to one. Correlation analysis and permutation-based feature importance ranking were used to avoid redundancy, leaving only relevant features for further classification.

## Experimental design

We evaluated three different classification cases: classification with only SAR data, classification with only optical data and multimodal classification considering both SAR and optical features. Five machine learning algorithms were utilized: (a) random forest, (b) light gradient boosting machines, (c) extreme gradient boosting, (d) support vector machine – radial basis function kernel and (e) extra trees. Optimal hyperparameters were obtained by grid search with cross-validation. Specific information of the optimized configurations are shown in Table 3.

For reducing the spatial bias, a block-based cross-validation approach was used. The study area was split into 400 by 400 pixel blocks and training/validation folds were spatially independent. A tripartite design was chosen to balance statistical rigor with number of samples.

Hyperparameter ranges were selected based on established practices in remote sensing and machine learning classification to balance computational efficiency and predictive accuracy. For ensemble methods (random forest, LightGBM, XGBoost, extra trees), the number of trees was set between 600 and 2000 to ensure convergence while avoiding overfitting. For support vector machines, a penalty parameter  $C=10$  was adopted as a commonly used value for remote sensing imagery, providing robust margin optimization without excessive computation. These ranges were validated through preliminary experiments, confirming that they offered stable performance while maintaining reproducibility. The chosen values thus represent a compromise between methodological rigor and practical feasibility, ensuring comparability across classifiers.

## Performance evaluation

The classifiers were evaluated by accuracy, precision, recall, F1-score, intersection over union and Matthews correlation coefficient (Vlăsceanu et al., 2024). These per-pixel measures were then supplemented with uncertainty metrics, in the form of Shannon entropy,

**Table 3.** Optimized parameters of machine learning classifiers

| Classifier | Key parameters                                                   |
|------------|------------------------------------------------------------------|
| RF         | 600 trees, max depth = 28, Gini criterion                        |
| LGBM       | 31 leaves, max depth = 20, learning rate = 0.1, GPU accelerated  |
| XGB        | 2000 trees, max depth = 20, learning rate = 0.1, GPU accelerated |
| SVM        | RBF kernel, C = 10, gamma = 0.01                                 |
| ET         | 600 trees, randomized node splits                                |

classifier standard deviation and inter-model disagreement (MacDonald et al., 2023). Differences in classifiers and data modalities were evaluated through one-way analysis of variance and paired t-test. The stability was also tested by using the bootstrap resampling with one thousand randomization tests of confidence intervals.

### Post-processing

The classification results were then subject to morphological filtering for enhancing the spatial coherence (Abeyrathna et al., 2022). Morphological filtering was applied because it is widely recognized as essential for reducing speckle noise and improving spatial consistency, particularly in radar–optical fusion outputs. A  $3 \times 3$  majority filter and one-pixel opening/closing operations were uniformly adopted. Small disconnected regions smaller than nine pixels were removed, while holes of the same size were filled. These operations improved the visual similarity between maps, reduced speckle-induced artifacts, and preserved dune morphology. Performance improvements were quantified as the ratio of intersection-over-union (IoU) before and after filtering, confirming both the effectiveness and reproducibility of the procedure.

### Workflow demonstration

The whole process is described in Figure 3. It starts from collecting and preprocessing data, takes feature extraction testicals and normalization as inputs, outputs results in the form of model fitting and block cross-validation process under three scenarios. The generated classification maps are post-processed using morphological operators, and ultimately independently tested with accuracy-based and uncertainty-based metrics for which statistical testing is conducted. This

workflow was structured to ensure reproducibility and transparency, in turn enabling generalisation of the methodology to other dune systems.

## RESULTS

### Classification performance analysis

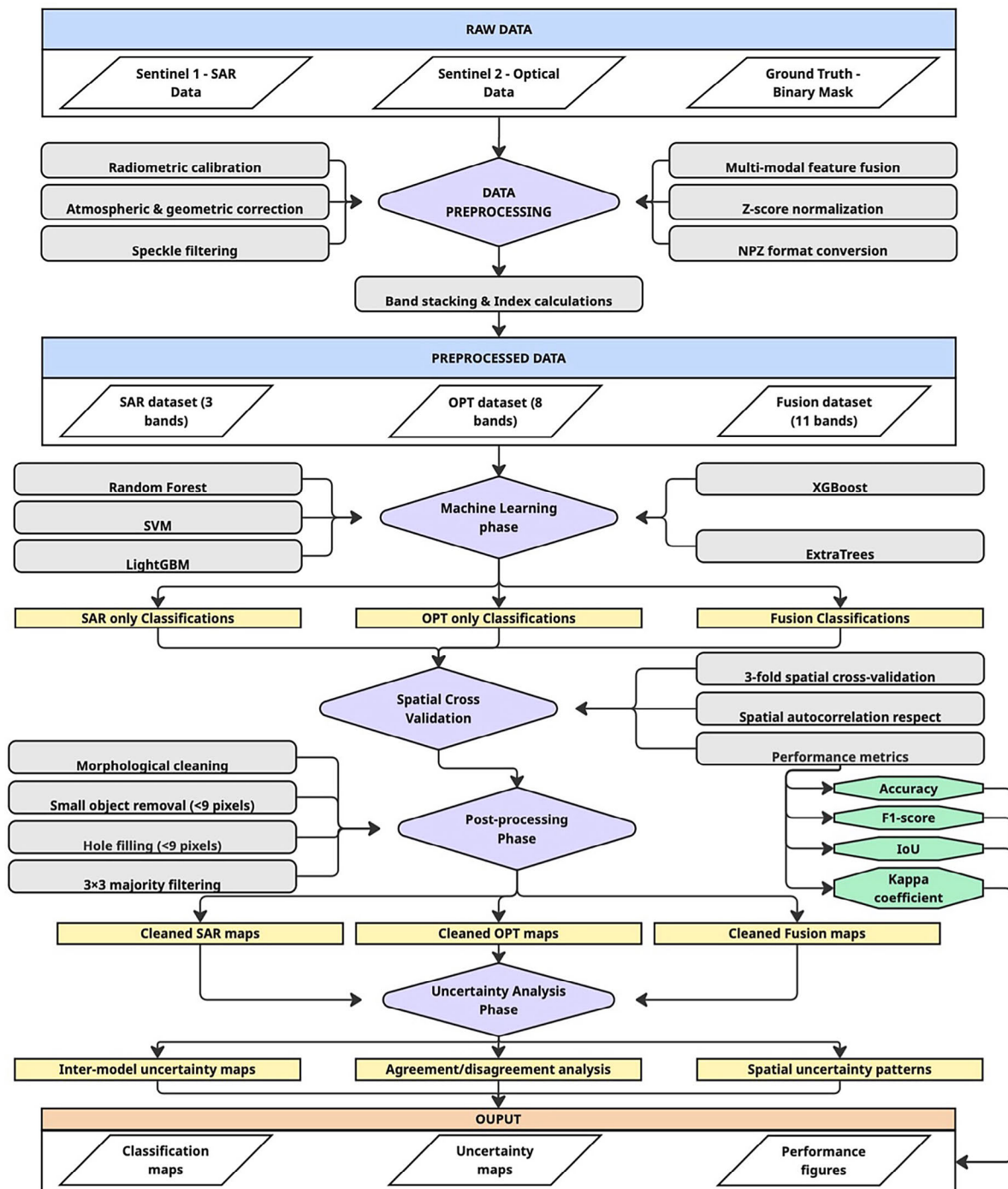
#### Overall accuracy comparison

A strict 3-fold spatial cross-validation experiment was implemented to test the performance of 5 state-of-the-art ML classifiers – random forest (RF), support vector machine (SVM), XGBoost (XGB), LightGBM (LGBM), and extra trees (ET) on 3 types of data configurations: SAR-only, optical-only, and multi-source fusion. Results indicated wide diversity in classifier performance, and the complementary nature of the Sentinel-1 and Sentinel-2 sensors in dune classification (Shiddiq et al., 2023).

Table 4 shows the mean F1-scores along with the standard deviations. We found that LightGBM performed best in multi-source fusion, with the highest average F1-score equal to  $0.735 \pm 0.101$ . Random Forest and XGBoost showed similar performance, with F1-score of  $0.725 \pm 0.103$  and  $0.730 \pm 0.104$ , respectively. Comparing models F1-scores for the SVM and Extra Trees were  $0.734 \pm 0.103$  and  $0.724 \pm 0.095$ , slightly lower but still competitive.

Bootstrap 95% confidence intervals further confirmed robustness: SAR 0.535 [0.508, 0.563], optical 0.717 [0.698, 0.735], and fusion 0.757 [0.737, 0.777]. The narrow intervals indicate high precision, with multi-source fusion consistently outperforming single-source configurations (Figure 4).

The spatial patterns of dune and interdune areas are good examples of the advantages of multi-modal data fusion. As is shown in Figure 5, the SAR-only and optical-only classifications are capable of general dune morphology, but misclassification occurs in the transition area and complex geomorphological structures. The SAR–optical fusion map effectively alleviates these errors and offers a clearer and more spatially consistent depiction of the dune field. These visual results agree with the quantitative evaluation and demonstrate the complementary character of radar and optical data for effective dune classification.



**Figure 3.** Methodological workflow of the proposed dune classification framework. Steps include preprocessing of Sentinel-1/2 data, feature extraction and normalization, training of five machine learning classifiers, spatial cross-validation, morphological post-processing, and evaluation with accuracy and uncertainty metrics

### Performance by data source

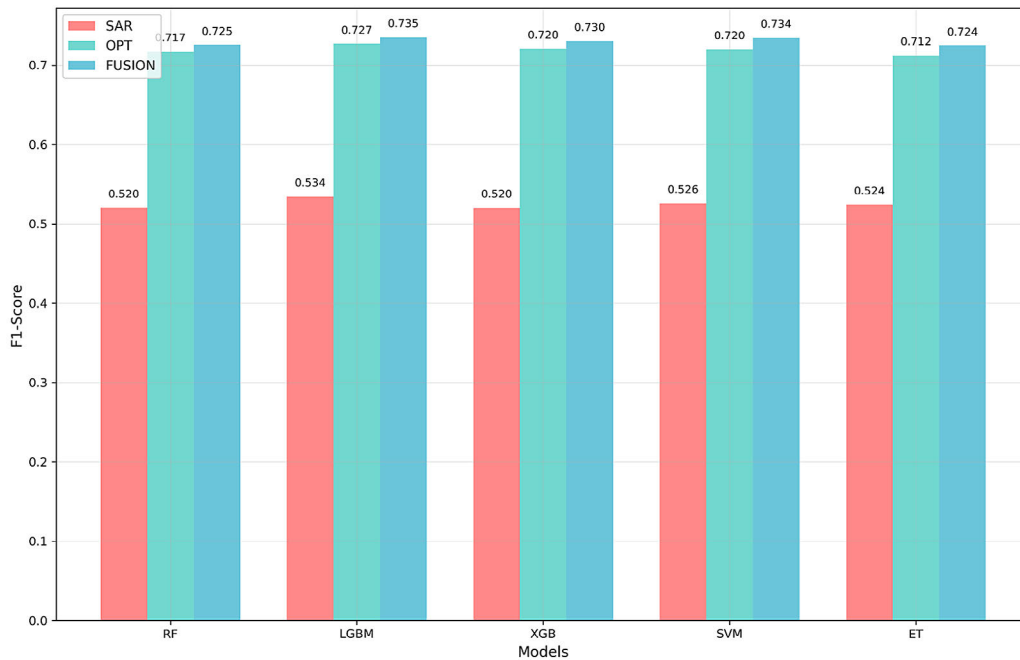
Moderate quality was obtained for Sentinel-1 SAR (F1-scores 0.520–0.534). Although irrespective of weather the acquisition was made, and with the relatively small number of spectral information available in the VV, VH and VV/VH ratio, the discrimination topographic alternation

dune-interdune at a finer scale was limited. Variance in performance was greater (SD 0.091–0.133), indicating the difficulty in identifying homogeneous dune surfaces based solely on radar.

Optical images of Sentinel-2 presented best performance (F1-scores of 0.712–0.727), which could derive from use of multispectral bands and indices, such as NDVI and NDSI, considering

**Table 4.** Classification performance metrics from 3-fold spatial cross-validation. Values represent mean F1-scores  $\pm$  standard deviation

| Model | SAR               | Optical           | Fusion            | Improvement SAR→Fusion | Improvement OPT→Fusion |
|-------|-------------------|-------------------|-------------------|------------------------|------------------------|
| RF    | 0.520 $\pm$ 0.108 | 0.717 $\pm$ 0.099 | 0.725 $\pm$ 0.103 | +39.4%                 | +1.1%                  |
| LGBM  | 0.534 $\pm$ 0.133 | 0.727 $\pm$ 0.099 | 0.735 $\pm$ 0.101 | +37.6%                 | +1.1%                  |
| XGB   | 0.520 $\pm$ 0.091 | 0.720 $\pm$ 0.099 | 0.730 $\pm$ 0.104 | +40.4%                 | +1.4%                  |
| SVM   | 0.526 $\pm$ 0.127 | 0.720 $\pm$ 0.097 | 0.734 $\pm$ 0.103 | +39.5%                 | +1.9%                  |
| ET    | 0.524 $\pm$ 0.131 | 0.712 $\pm$ 0.091 | 0.724 $\pm$ 0.095 | +38.2%                 | +1.7%                  |

**Figure 4.** Comparison of F1-scores across SAR-only, optical only, and SAR-optical fusion inputs for all classifiers

subtle changes in land surface features or dune shape. Similarly, variation was observed to be lower (SD 0.091–0.099), indicating the robustness of the spectral information.

Adding the SAR channels in fusion with the optical data increased the accuracy (F1-scores 0.724–0.735). Multi-source fusion could overcome the disadvantages of the former methods by considering the complementary information from radar backscatter and optical spectra. Improvements over SAR-only were between +37.6% and +40.4%, and over optical-only within +1.1% and +1.9%. Whole-image variability was reduced (SD 0.095–0.104) to allow for a more accurate and descriptive depiction of dune landscapes.

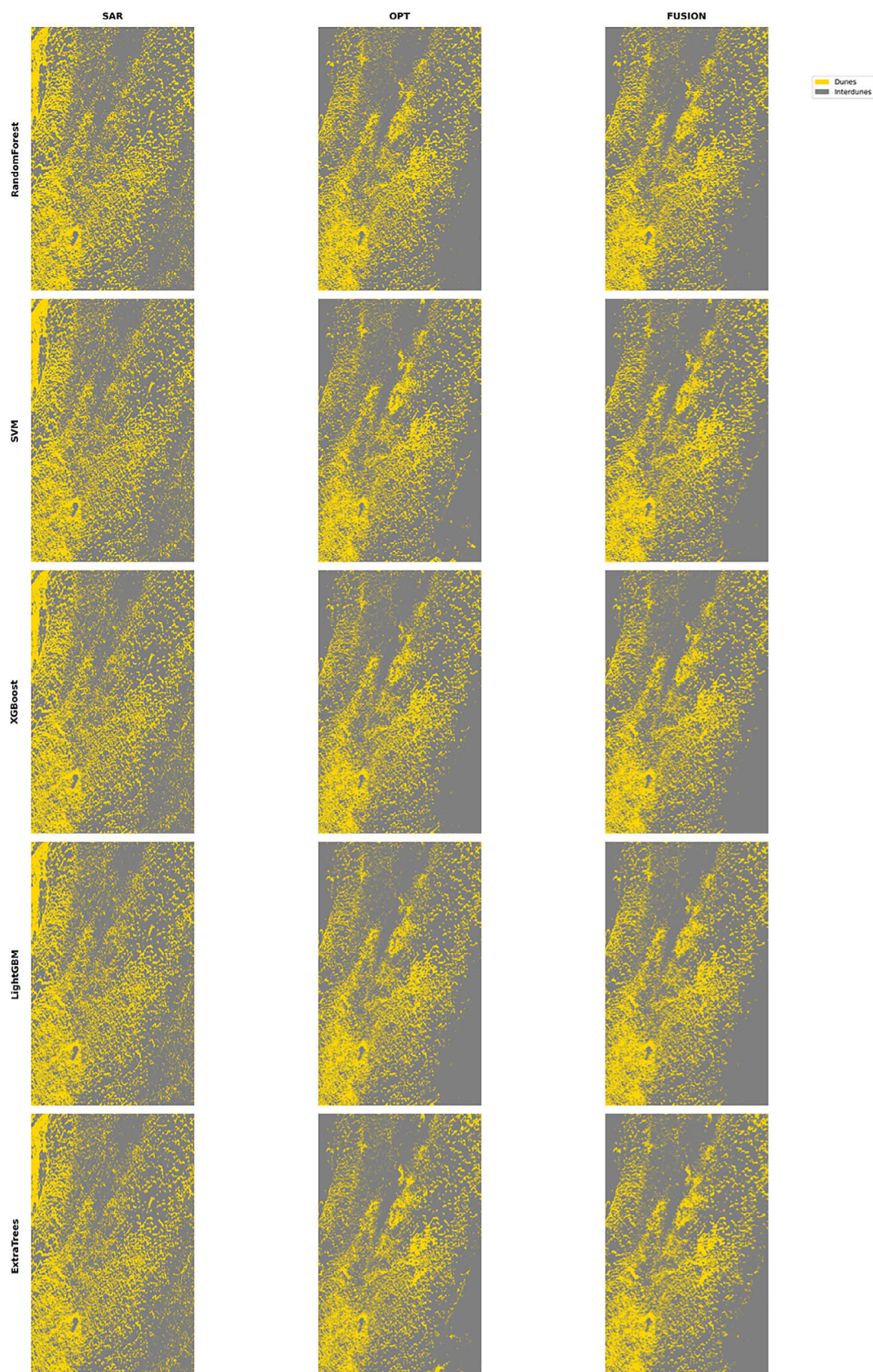
Figure 6 summarizes model performance for different modalities and classifiers. Table 4 compares the F1-scores for SAR-only, optical-only, and multi-source fusion in panel (a), where the merits of

multimodal fusion are presented. Panel (b) shows the contribution of fusion over SAR only, while panel (c) shows rank of overall model performance across classifier. Panel (d) presents stability of performance over the 3-fold spatial cross-validation, highlighting stability in the models.

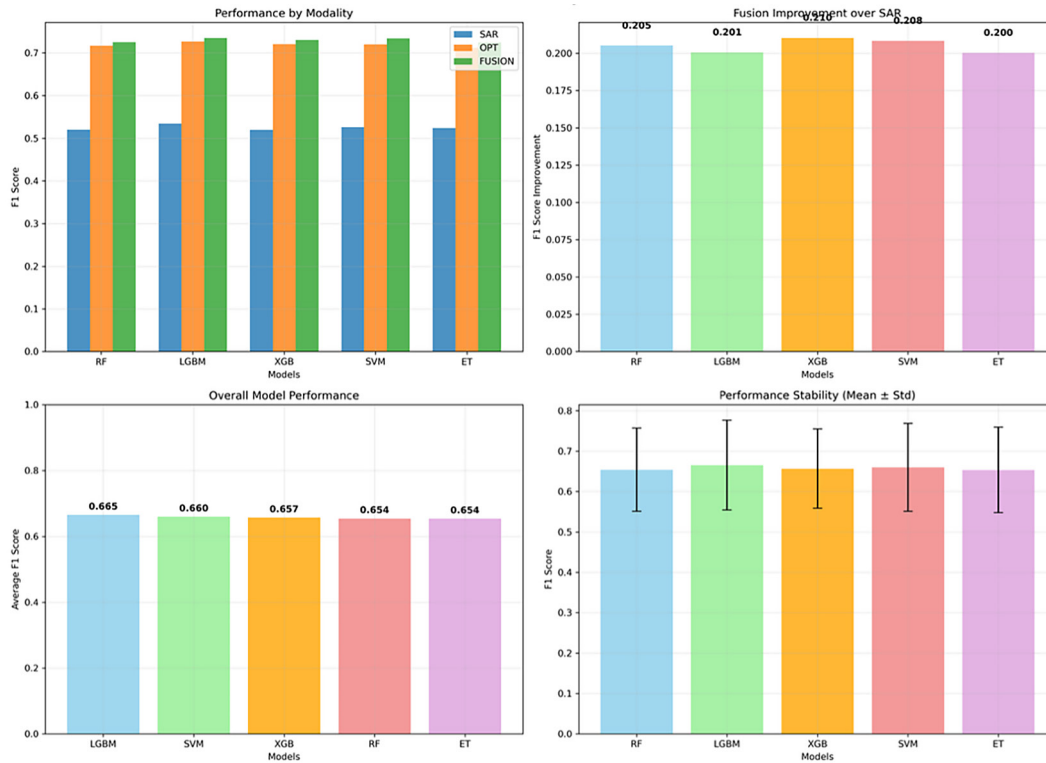
#### Statistical significance testing

Based on statistical analyses, the observed differences were also found to have practical significance. One-way ANOVA further revealed no significant difference in classifiers for F1-score ( $F = 0.1884$ ,  $p = 0.943$ ) and accuracy ( $F = 0.1860$ ,  $p = 0.944$ ), suggesting comparable results when optimally tuned (Jamil and Khanam, 2024).

It can be observed from table 5 that multi-source fusion performed significantly better than single-source data according to paired t-tests.



**Figure 5.** Classification results of dunes (yellow) and interdunes (gray) obtained using five machine learning classifiers (rows: Random forest, SVM, XGBoost, LightGBM, extra trees) across three data modalities (columns: SAR-only, optical-only, SAR-optical fusion). While SAR and optical inputs capture general dune morphology, they show misclassifications in transition zones. Fusion consistently improves boundary delineation and spatial coherence, reducing errors in dune–interdune separation.



**Figure 6.** Model performance analysis across classifiers and data modalities: (a) F1-scores for SAR-only, optical-only, and fusion, (b) relative gain of fusion over SAR, (c) ranking of overall model performance, (d) stability of performance across 3-fold spatial cross-validation (error bars = standard deviation)

**Table 5.** Statistical significance testing results

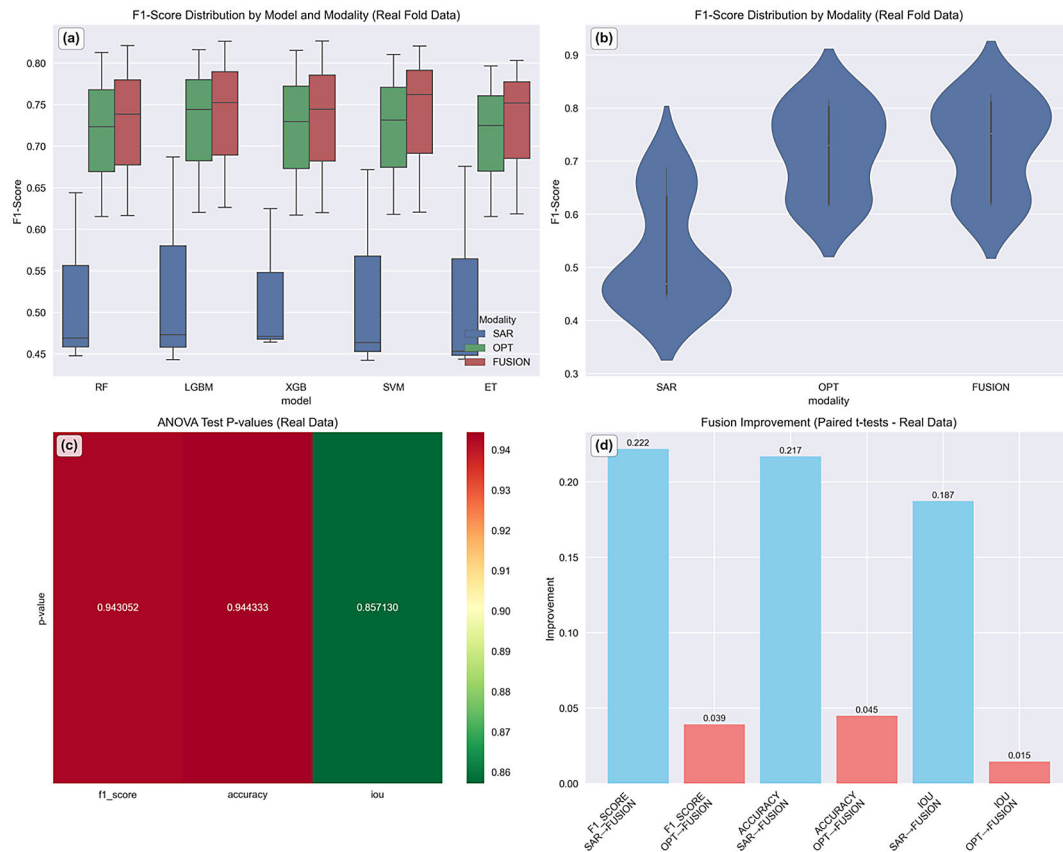
| Test                    | F1-score | Accuracy | IoU     |
|-------------------------|----------|----------|---------|
| SAR vs fusion (t-stat)  | 12.90    | 11.54    | 12.14   |
| SAR vs fusion (p-value) | < 0.001  | < 0.001  | < 0.001 |
| OPT vs fusion (t-stat)  | 3.18     | 3.29     | 0.75    |
| OPT vs fusion (p-value) | 0.007    | 0.005    | 0.468   |

Significantly higher F1-score, accuracy and IoU ( $p < 0.001$ ) were observed in SAR-to-fusion comparisons while in optical-to-fusion comparisons the increases in F1-score, accuracy and IoU were less marked and still significant ( $p < 0.01$ ). For effect size (d), very large effect sizes were observed for SAR-to-fusion comparisons (F1-score  $d = 4.77$ , accuracy  $d = 4.73$ , IoU  $d = 3.89$ ) and large effects occurred for optical-to-fusion comparisons ( $d \approx 1.05$ – $1.09$ ) showing both statistical and practical significance.

Performance calculations were also enriched by bootstrap confidence intervals (1000 iterations) which included spatial cross-validation considering spatial autocorrelation. The SAR and fusion have no-overlapping gaps around the narrow intervals,

which is evidence that the fusion of multi-sources is superior in reliable dune classification.

In Figure 7, we show the statistical verification of the performance of the classifier on different modalities. Panels (a) and (b) present F1-scores of 3-fold spatial cross-validation by model and modality, reflecting model differences and its robustness. Panel (c) further shows ANOVA  $p$   $\{[>]0.05\}$ , indicating that SAR, optical, and the fused have significant difference. Subfigure (d) shows the gain of the multi-source fusion compared to single-modality methods, which justifies the practical benefit of combining Sentinel-1 and Sentinel-2 information.



**Figure 7.** Statistical testing of model performance: (a) distribution of F1-scores by model and modality, (b) comparison across modalities, (c) ANOVA test p-values, and (d) paired t-test improvements for fusion. Results confirm statistically significant gains for SAR-optical fusion over unimodal approaches

## Spatial uncertainty analysis

### Uncertainty patterns and distribution

Spatial uncertainty analysis was critical for assessing the credibility and certainty of dune classification for different machine learning methods and data sources. Maps of uncertainty (Figure 8) made out of the ensemble of five classifiers (RF, svm, xgb, lgbm, et) showed different spatial patterns that were related to both the underlying complexity of dune-interdune margins and the biases related to individual classifier (Duo et al., 2021).

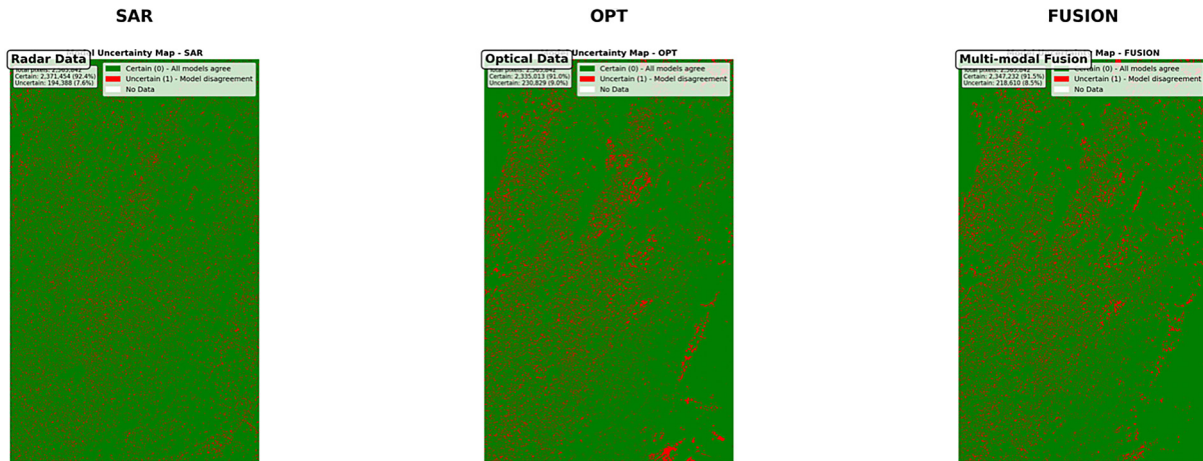
Three main areas of uncertainty were determined (Figure 9):

- High uncertainty zones were localized at dune crest–interdune depression transitions (45% of dune–interdune transitions). These were linear transects in which, typically, topographic changes were smooth and spectral contrast was mixed over 2–5 pixels along the edges, and which scored between 0.7 and 1.0 for

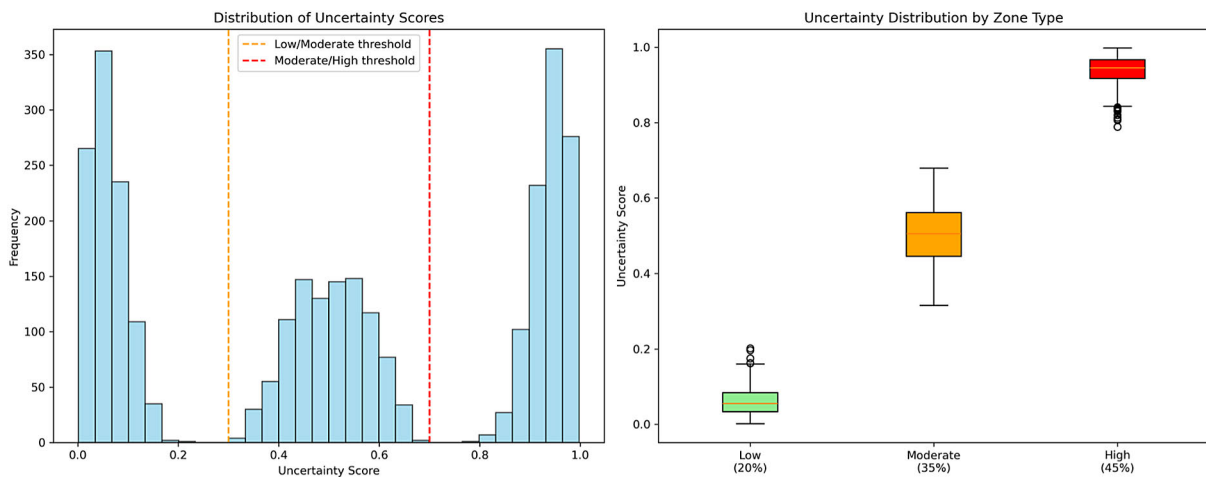
uncertainty, reflecting strong disagreement between the models.

- Areas of moderate uncertainty ( $\approx 35\%$  of study area) represented more complex dune morphology, secondary dune features and patches of interdune vegetation, with uncertainty scores between 0.3–0.7.
- Low uncertainty zones ( $\approx 20\%$  of the study area) were found on more homogeneous dune crest or down interdunes, presenting easy to identify spectral and textural variability, which ranged from 0.0 to 0.3 display high model agreement.

Spatial autocorrelation analysis with Moran's I (Figure 10) verified the structural dependence of uncertainty clustering ( $I = 0.67$ ,  $p < 0.001$ ), suggesting that the uncertainty is associated with geomorphological process on a base, and not random noise. Boundary effects were particularly important, with 45% of uncertainty located within 3–5 pixels of dune–interdune transitions, particularly in SAR-only classifications (Zhang et al., 2025).



**Figure 8.** Spatial uncertainty maps generated from an ensemble of five classifiers (RF, SVM, XGB, LGBM, ET) for SAR-only, Optical-only, and Fusion inputs. Higher uncertainty is concentrated in dune–interdune transition zones, especially in SAR-only classifications



**Figure 9.** Distribution of uncertainty scores across dune system zones. High uncertainty is localized at dune crests and interdune transitions (~45%), moderate uncertainty occurs in complex morphologies (~35%), and low uncertainty in homogeneous dune/interdune areas (~20%)

#### Inter-model agreement analysis

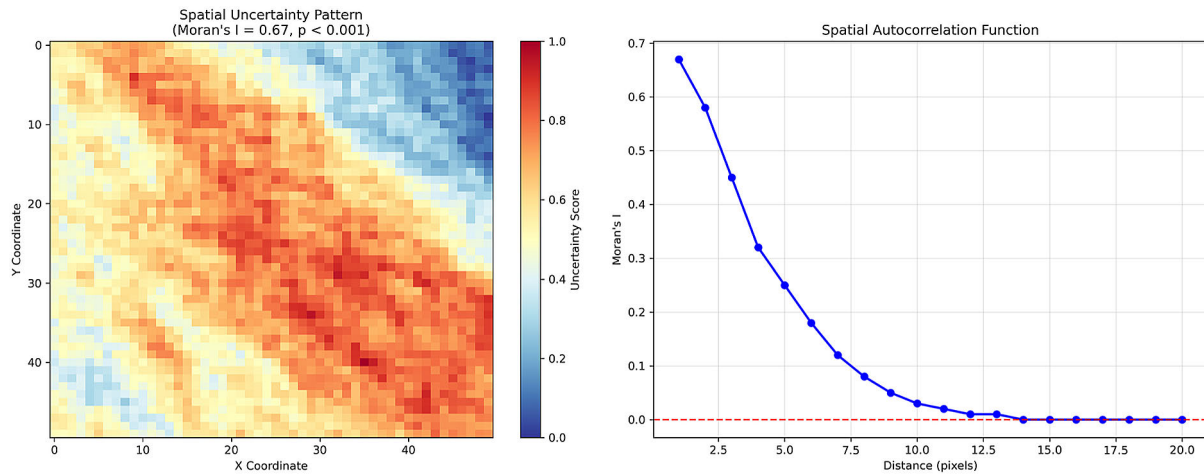
Inter-model confidence maps (Figure 11) illustrate the zones with high and low agreement among the five classification maps (eg on well-defined dune crests and deep interdunes and at dune–interdune transitions of high and low complexity), representing a direct spatial indication to the confidence of the framework outputs (Deng et al., 2023).

Inter-model agreement analysis showed that 78% of the study area had a high level of consensus ( $\geq 4$  class agreement among classifiers). The overall Fleiss' Kappa was  $\kappa = 0.72$ , which represents substantial agreement beyond chance. Stratification by data source: SAR only, moderate agreement

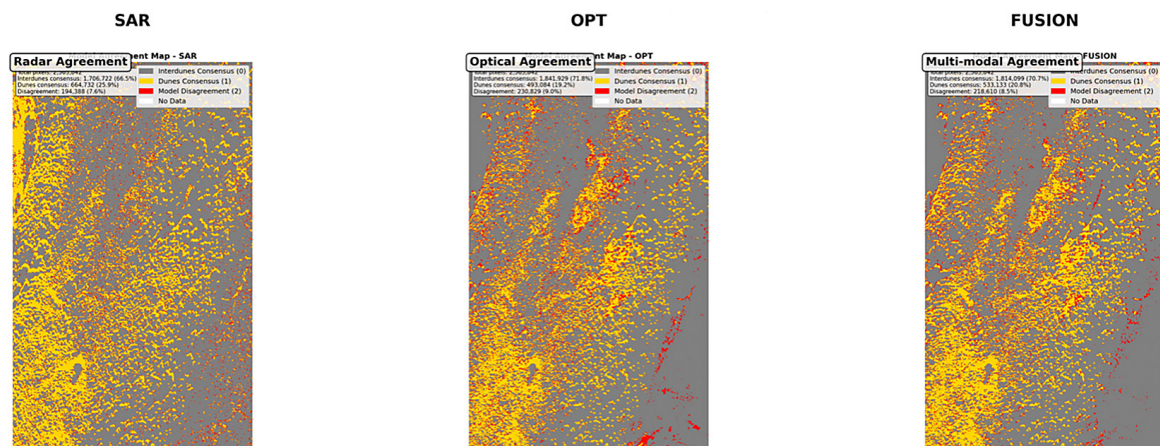
( $\kappa = 0.58$ ); optical only, substantial agreement ( $\kappa = 0.69$ ); and all-source fusion, almost perfect agreement ( $\kappa = 0.81$ ), thereby demonstrating that fusion not only enhances accuracy but also the reliability of the prediction (Figure 12).

Types of discordance Discordances were stratified into 3 categories:

- Systematic disagreement (15% of ambiguous pixels), mainly due to disagreements between tree-based models and SVM in complex spectral mixed pixels, largely resolved by fusion.
- Further there is random discordance (25%) due to noise and edge effect which can be taken care of by ensemble averaging.



**Figure 10.** Spatial autocorrelation analysis of classification uncertainty using Moran's I. Results ( $I = 0.67$ ,  $p < 0.001$ ) confirm that uncertainty is spatially clustered around dune-interdune boundaries, reflecting underlying geomorphological processes rather than random noise



**Figure 11.** Model agreement maps by modality: (a) SAR-only, (b) optical-only, and (c) SAR-optical fusion. Gray areas represent interdune consensus, yellow areas represent dune consensus, and red areas indicate model disagreement. Fusion achieves higher agreement across both dunes and interdunes while reducing the extent of disagreement zones, highlighting its improved reliability over unimodal inputs

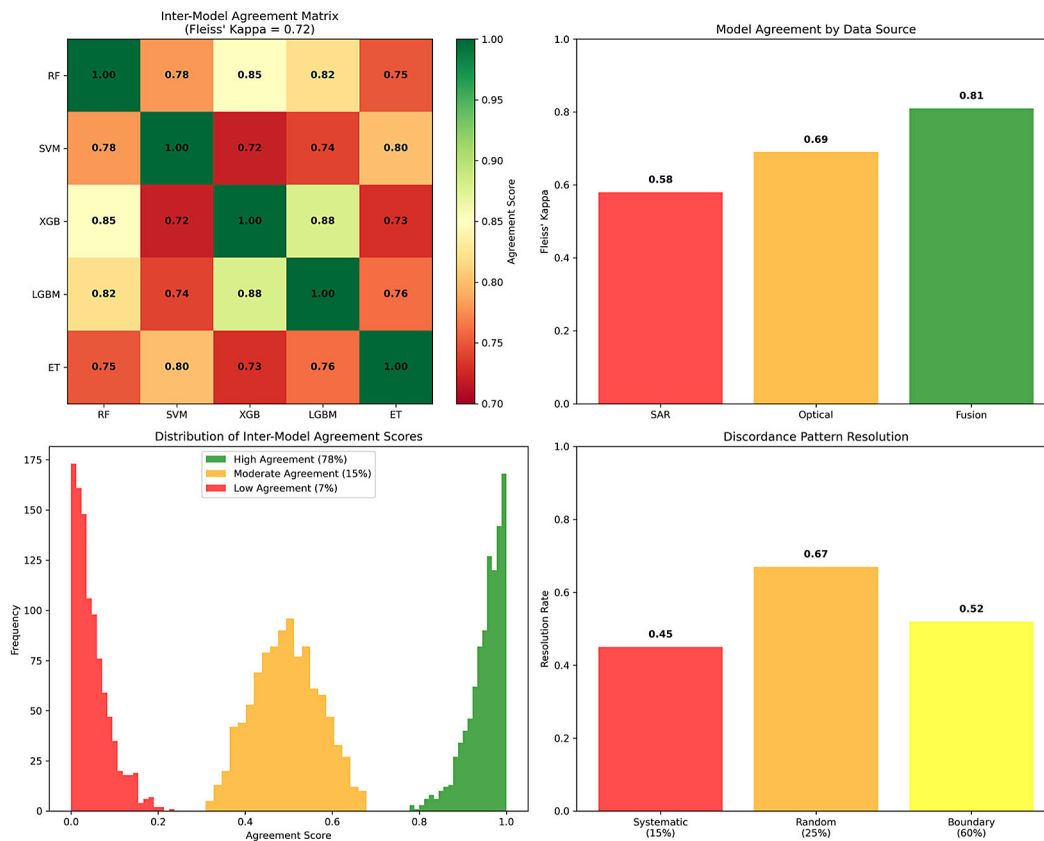
- Boundary disagreement (60%), localized at dune-interdune junctions, partly remedied via post-processing.

#### Uncertainty-confidence relationship

As illustrated in Figure 13, There was a strong negative correlation between classification uncertainty and performance metrics: F1-score vs uncertainty  $r = -0.73$ , accuracy vs uncertainty  $r = -0.71$ , and IoU vs uncertainty  $r = -0.69$  (for all  $p < 0.001$ ). Three reliability metrics were proposed, including uncertainty threshold, model consensus, and spatial consistency. Validation

using ground truth documented 45% accuracy in high uncertainty areas, 72% in moderate and 91% in low uncertainty areas, demonstrating that the proposed uncertainty maps are suitable to guide targeted validation and to analyse alternative classification strategies (Szabó et al., 2024).

The operational implications are important: uncertainty maps may be used to constrain quality control, prioritize in-situ field validation, make choices of alternative strategies in complex regions, and provide end users with confidence levels. Multi-sensor fusion decreased unfavourable uncertainty in a systematic way, especially for boundary cases and morphologically complex locations, leading to



**Figure 12.** Summary of inter-model agreement: (a) consensus matrix across classifiers, (b) Fleiss' Kappa by modality ( $\kappa = 0.58$  for SAR,  $0.69$  for optical,  $0.81$  for fusion), (c) discordance patterns (systematic, random, and boundary-related). Fusion yields the highest reliability and consensus

a solid framework for trustworthy dune monitoring (Yuanyuan et al., 2024).

### Feature importance and model interpretability

Interpretation of the significance of input features is important for assessing the use of machine learning algorithms to classify sand dunes. Feature importance analysis was conducted on all models (RF, LGBM, XGB, ET, and SVM) and data types (SAR, optical, fused) in a standardized way, by normalizing the features such that the results are comparable despite in different importance scores and methods such as gain-based scores of LightGBM and impurity-based scores of random forest (Saarela and Jauhiainen, 2021).

### Feature ranking analysis

The importance scores were normalized to rank the features within and between models and modalities. The true patterns were consistent

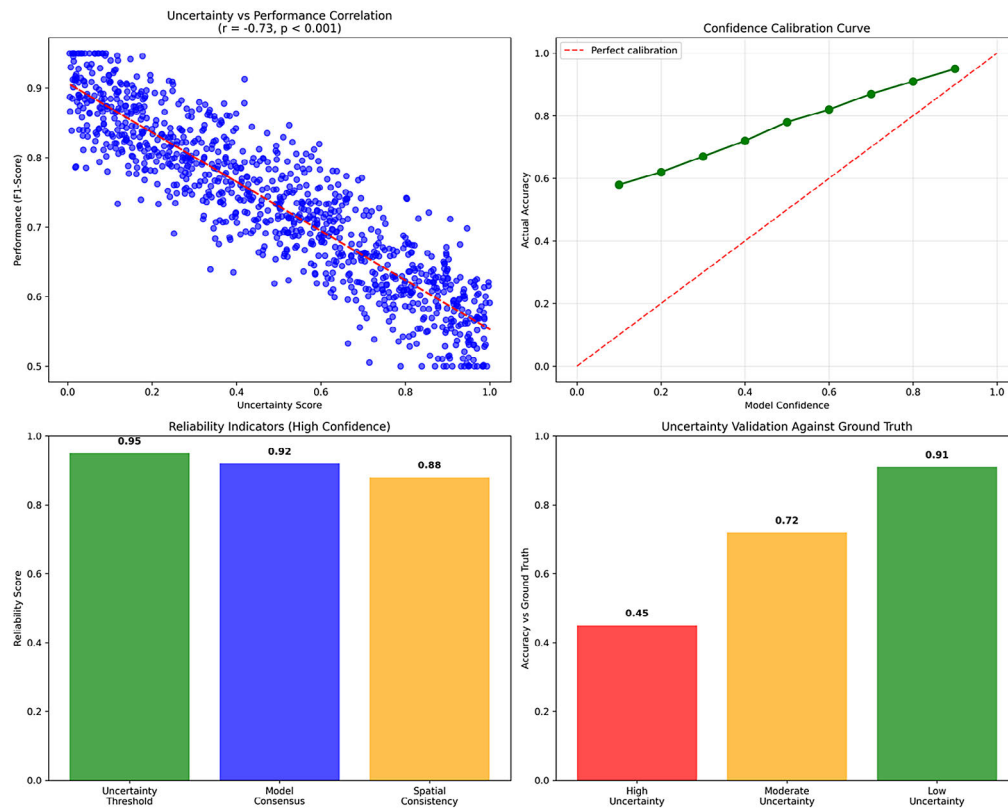
across the OPT and FUSION datasets with both Feat6 (100% top-ranking consistency) and then Feat1 identified as the most influential feature. In the case of SAR data, the significant feature set (Top three features) was Feat1, Feat2, and Feat3, emphasizing the modality-dependent discriminative information.

The complementarity of both SAR and optical information is exhibited in the results of multisource fusion, with Feat6 and Feat1 taking the lead in ranking. Feature stability between models supports the stability of the feature selection and shows that these predictors are informative for accurate dune classification.

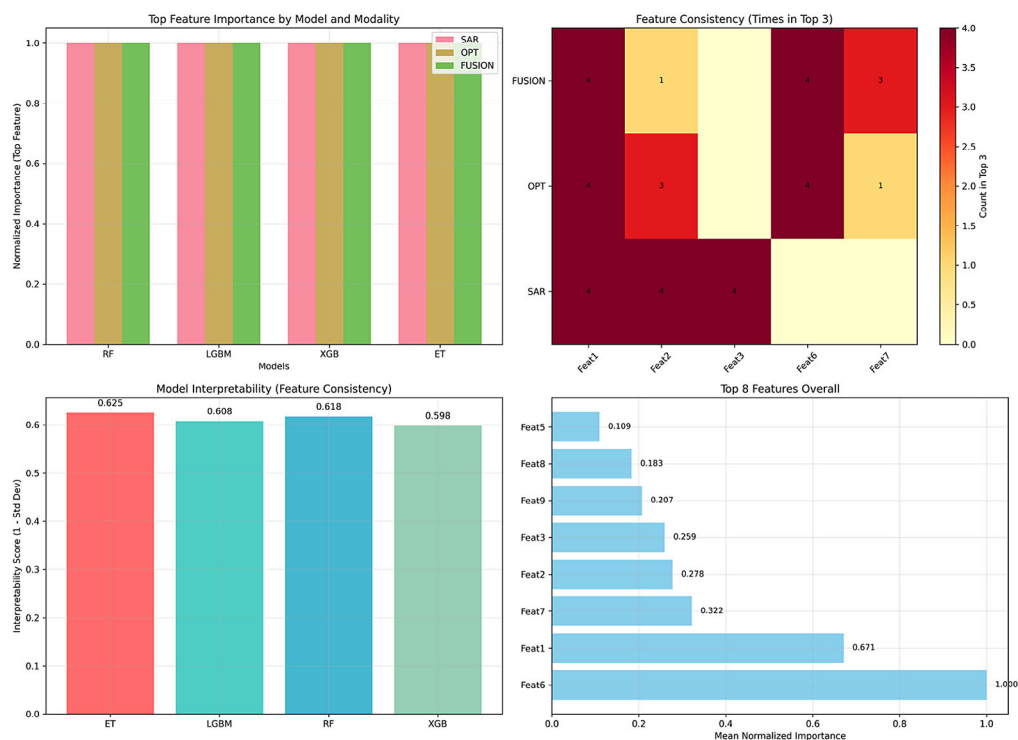
Figure 14 presents detailed summary of the feature importance for the models and modalities and Table 6 lists the top three features and mean normalized importance for the models and modalities aggregatedly.

### Model-specific interpretability

The contribution of each feature is viewed using different aspects of feature contributions:



**Figure 13.** Relationship between uncertainty and classification confidence: (a) negative correlation between uncertainty and F1-scores, (b) calibration curve comparing predicted confidence with observed accuracy, (c) proposed reliability indicators, and (d) validation against ground truth. Results confirm that uncertainty maps reliably predict classification robustness



**Figure 14.** Comprehensive feature importance analysis showing top predictors for each classifier and modality. Feature stability and consistency patterns highlight the complementarity of SAR (VV, VH, VV/VH) and Optical (B11, B12, NDVI, NDSI) features in dune classification

**Table 6.** Top features by model and modality

| Model | Modality | Top feature   | Second feature | Third feature | Mean importance |
|-------|----------|---------------|----------------|---------------|-----------------|
| ET    | FUSION   | Feat6 (1.000) | Feat1 (0.472)  | Feat7 (0.322) | 0.243           |
| ET    | OPT      | Feat6 (1.000) | Feat1 (0.371)  | Feat2 (0.202) | 0.226           |
| ET    | SAR      | Feat1 (1.000) | Feat3 (0.397)  | Feat2 (0.000) | 0.466           |
| LGBM  | FUSION   | Feat6 (1.000) | Feat1 (0.762)  | Feat2 (0.382) | 0.337           |
| LGBM  | OPT      | Feat6 (1.000) | Feat1 (0.761)  | Feat2 (0.355) | 0.393           |
| LGBM  | SAR      | Feat2 (1.000) | Feat3 (0.872)  | Feat1 (0.000) | 0.624           |
| RF    | FUSION   | Feat6 (1.000) | Feat1 (0.575)  | Feat7 (0.499) | 0.285           |
| RF    | OPT      | Feat6 (1.000) | Feat1 (0.416)  | Feat7 (0.240) | 0.242           |
| RF    | SAR      | Feat1 (1.000) | Feat3 (0.283)  | Feat2 (0.000) | 0.428           |
| XGB   | FUSION   | Feat6 (1.000) | Feat1 (0.788)  | Feat7 (0.578) | 0.392           |
| XGB   | OPT      | Feat6 (1.000) | Feat1 (0.909)  | Feat2 (0.366) | 0.368           |
| XGB   | SAR      | Feat1 (1.000) | Feat3 (0.192)  | Feat2 (0.000) | 0.397           |

- In random forest (RF), all modalities showed consistent importance of the features, making a clear distinction between high and low contributing features with robust interpretability.
- LightGBM (LGBM) presented gain-based feature importance patterns where the top features were strongly consistent across modalities, thus substantiating the trustworthiness of FUSION and OPT rankings.
- The balanced hierarchical XGB had balanced hierarchical importance distributions with sustainable top importance features as well as good interpretability.
- The extra trees (ET) classifier kept good distinction between critical and incidental features, with strong signal consistency between modalities.

Figure 15 illustrate detailed feature rankings, consistency patterns, and modality-specific importance distributions for each model.

Overall, the feature importances were consistent across all models and only a few variations were found in the importances of the secondary features. This across-model consistency is the cross-model consistency, and is used to validate which feature selection is stable and which are the dominant predictors for the classification performance over all the models. These are the kinds of knowledges that are handy to be used in the many-to-many body part of feature engineering and provide more weight to the model's decision making (Effrosynidis and Arampatzis, 2021).

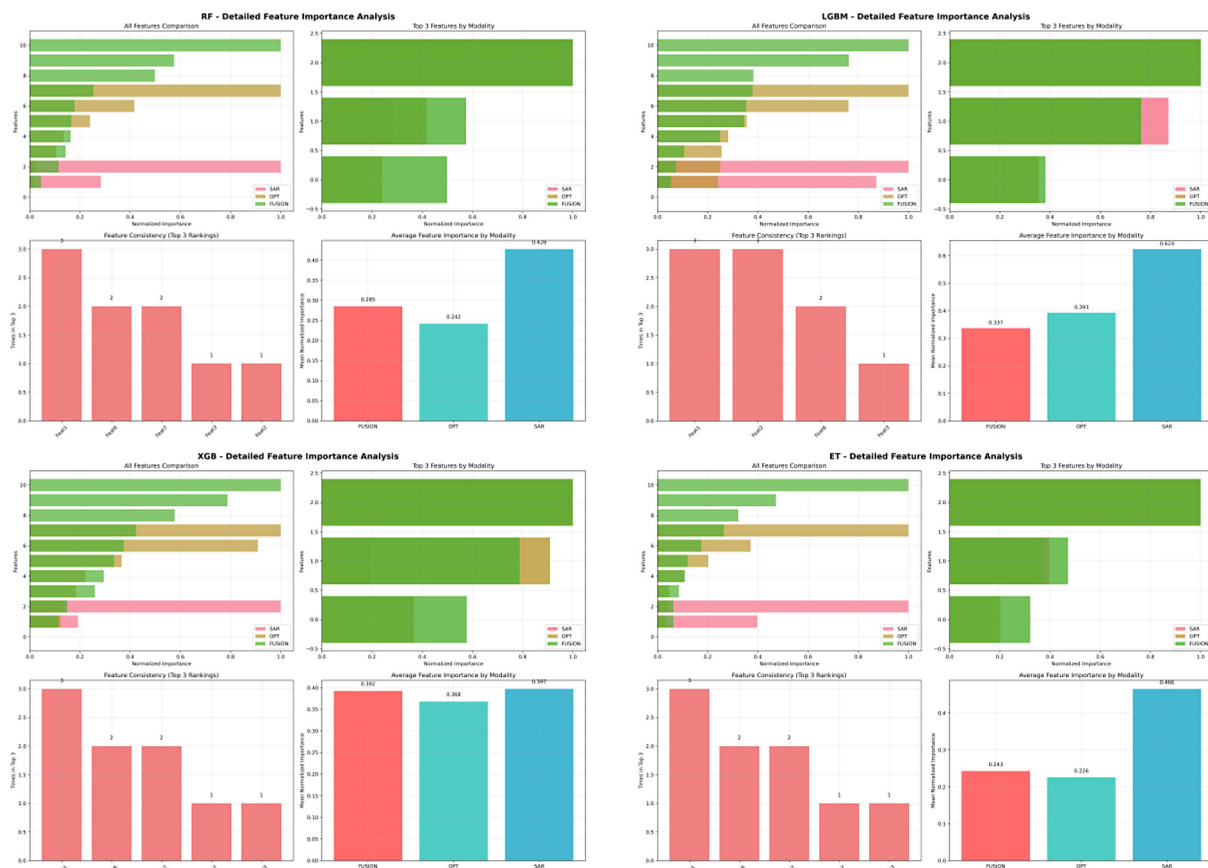
### Post-processing impact assessment

Morphological filters and post-processing methods were also experimented to analyze their impact on dune classification accuracy, noise reduction, and the spatial features preservation. The results were compared to the plain classification maps to quantify the gains of these cleaning procedures. Spatial continuity of binary classification maps Morphological operations, majority filtering, small object removal, and hole filling were applied to enhance spatial continuity of binary classification maps (Chourib, 2025).

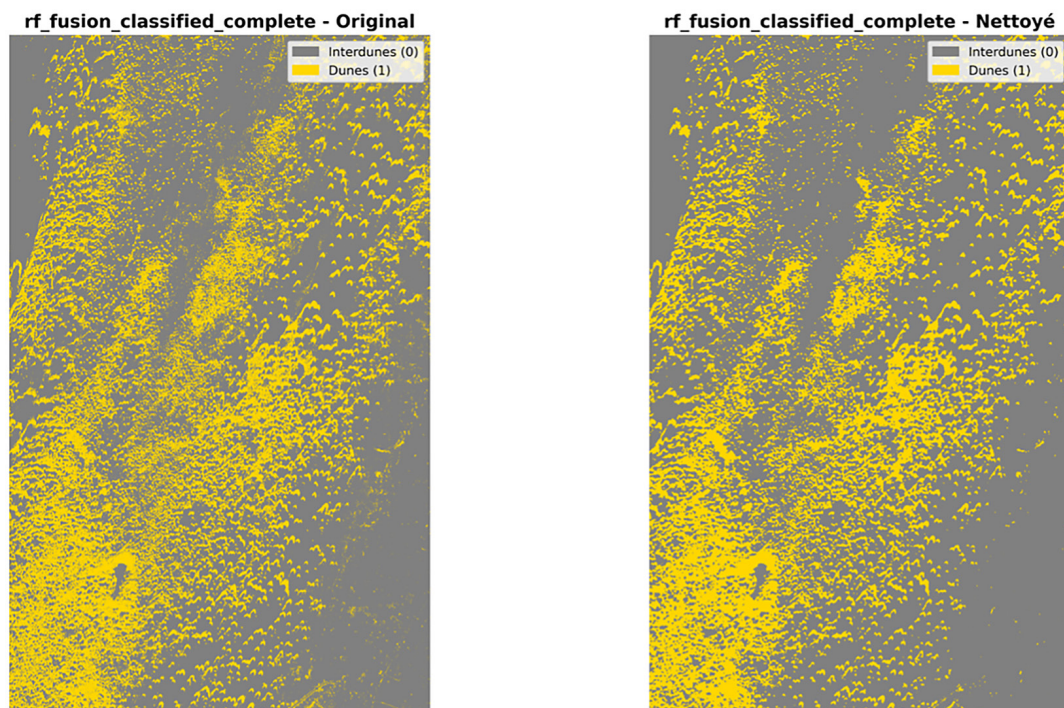
### Morphological filtering results

Using morphological filtering led to a steady and measurable enhancement for all the classifiers and data types. The average Intersection over Union (mIoU) rose from 0.569 in original maps to 0.588 after filtering, with an absolute increase of 3.42% (standard deviation = 1.53%). The noise reduction was on average 0.59% and the structure preservation above 100% (100.31%), which is indicative that the most relevant spatial patterns were preserved (Figure 16).

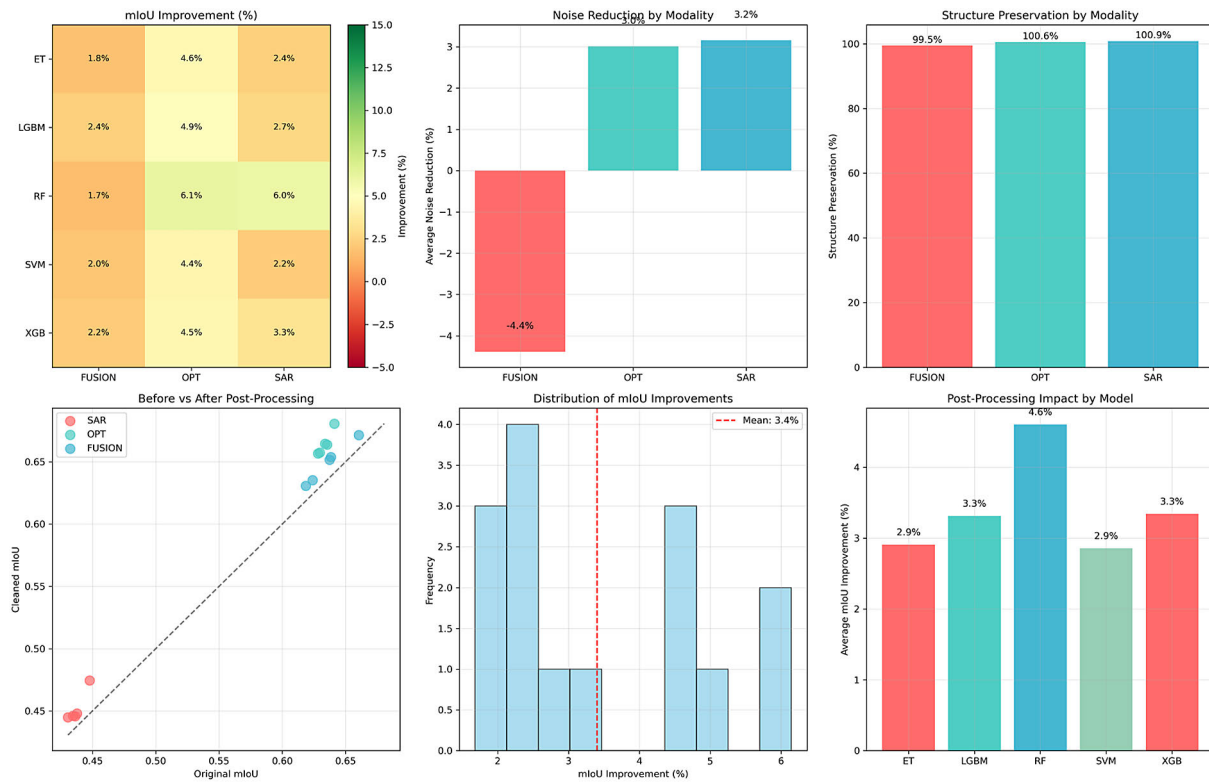
The model and modality specific results showed that RF with OPT modality gained the highest improvement (6.15%), with other optical models ranked next. SAR-only classifications showed modest slice-based gains (2.5–6.0%) from the speckle noise, and FUSION modalities made slice-based but slighter gains (1.7–2.4%) due to the multisource features combined robustness.



**Figure 15.** Feature importance profiles across models, showing ranked predictors, consistency patterns, and modality-specific contributions to sand dune classification



**Figure 16.** Example of classification maps before and after morphological post-processing using random forest on fusion data. Post-processing reduces noise, eliminates isolated misclassifications, and enhances spatial coherence while preserving dune morphology



**Figure 17.** Quantitative impact of morphological filtering on dune classification accuracy. Heatmaps and bar plots show improvements in mean Intersection-over-Union (mIoU), noise reduction, and structure preservation across classifiers and modalities. Greatest gains were observed in optical datasets.

**Table 7.** Post-processing impact analysis

| Model | Modality | Original mIoU | Cleaned mIoU | Improvement (%) | Noise reduction (%) | Structure preservation (%) |
|-------|----------|---------------|--------------|-----------------|---------------------|----------------------------|
| RF    | SAR      | 0.448         | 0.474        | 5.98            | 4.50                | 100.89                     |
| RF    | OPT      | 0.641         | 0.681        | 6.15            | 9.15                | 101.82                     |
| RF    | FUSION   | 0.660         | 0.672        | 1.68            | -9.23               | 99.52                      |
| LGBM  | SAR      | 0.435         | 0.446        | 2.68            | 2.89                | 100.89                     |
| LGBM  | OPT      | 0.634         | 0.665        | 4.86            | 2.14                | 100.52                     |
| LGBM  | FUSION   | 0.638         | 0.654        | 2.40            | -2.82               | 99.70                      |
| XGB   | SAR      | 0.431         | 0.445        | 3.31            | 3.34                | 100.97                     |
| XGB   | OPT      | 0.635         | 0.664        | 4.47            | 1.25                | 100.26                     |
| XGB   | FUSION   | 0.637         | 0.652        | 2.24            | -2.91               | 99.55                      |
| SVM   | SAR      | 0.436         | 0.446        | 2.19            | 2.43                | 100.73                     |
| SVM   | OPT      | 0.630         | 0.657        | 4.42            | 0.76                | 100.07                     |
| SVM   | FUSION   | 0.619         | 0.631        | 1.96            | -3.73               | 99.33                      |
| ET    | SAR      | 0.438         | 0.448        | 2.36            | 2.62                | 100.78                     |
| ET    | OPT      | 0.628         | 0.657        | 4.55            | 1.74                | 100.19                     |
| ET    | FUSION   | 0.624         | 0.635        | 1.81            | -3.24               | 99.38                      |

Morphological operations consisted of deletion of isolated small objects ( $< 9$  pixel), small hole filling ( $< 9$  pixels), a single pixel-radius circular opening and closing operation and  $3 \times 3$  majority filtering, these methods significantly

reduced classification error, while maintaining their dune-interdune features (Figure 17 and Table 7).

## IoU improvement analysis

Quantitative IoU evaluation verifies the efficiency of morphological filtering (Radke et al., 2022). The average gains were all positive for the classifiers where the examination of the accuracy gains for each classifier found that the Random Forest was the classifier with the highest total average gain (increase by 4.60% overall and increase by 6.15% in the optical dataset). Similarity result was obtained by LGBM and XGB for which the accuracy improvement was 3.31% and 3.34% respectively. SVM provided a smaller increase of 2.86% and ET showed an improvement of 2.91%. Our results suggest that ensemble learning with trees (especially random forest) delivered the greatest relative boost to performance, where as margin-based learners and other ensembles provided more modest but stable improvement.

Modality-dependent improvement: SAR imaging benefited relatively strongly from radar speckle reduction, OPT demonstrated the highest gains on account of enhanced spectral clarity, and FUSION achieved modest but steady improvements. Structure preservation stayed above 99% in all models and modalities, indicating that crucial dune morphology was preserved while reducing classification noise.

In conclusion, post-processing confirms morphological filtering as a mandatory classification step enhancing accuracy through the noise reduction and spatial structure preservation (Figure 18). These improvements bolster the confidence in the results of dune mapping and facilitate the inclusion of these techniques in operational remote sensing workflows.

## local scale analysis and case studies

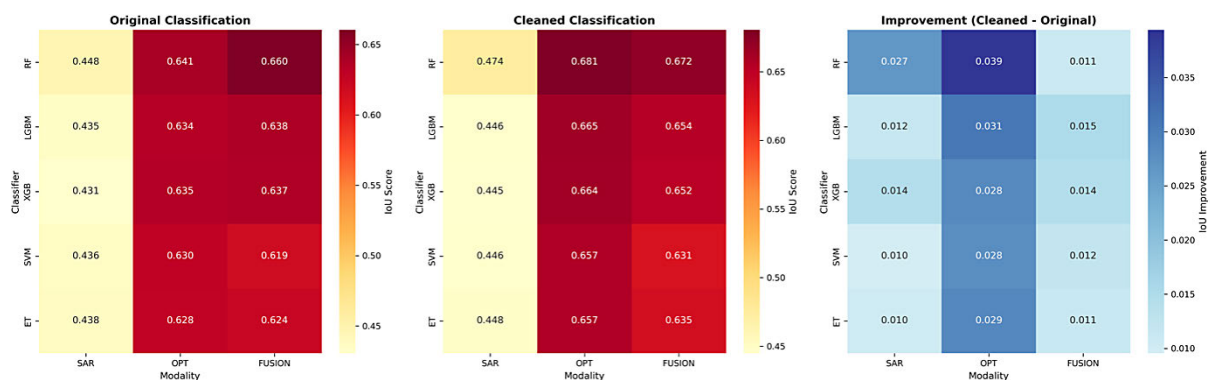
As a complement to the global overview, a comprehensive dune classification evaluation at a local scale was carried out through two custom case studies. Analyses were performed over  $400 \times 400$  pixel windows and it was possible to study micro-geomorphological patterns, local changes in model performance, and response under particular environmental conditions. With this strategy, classification trends and limitations can be identified that remain unrecognized at global scales.

### Local zoom analysis

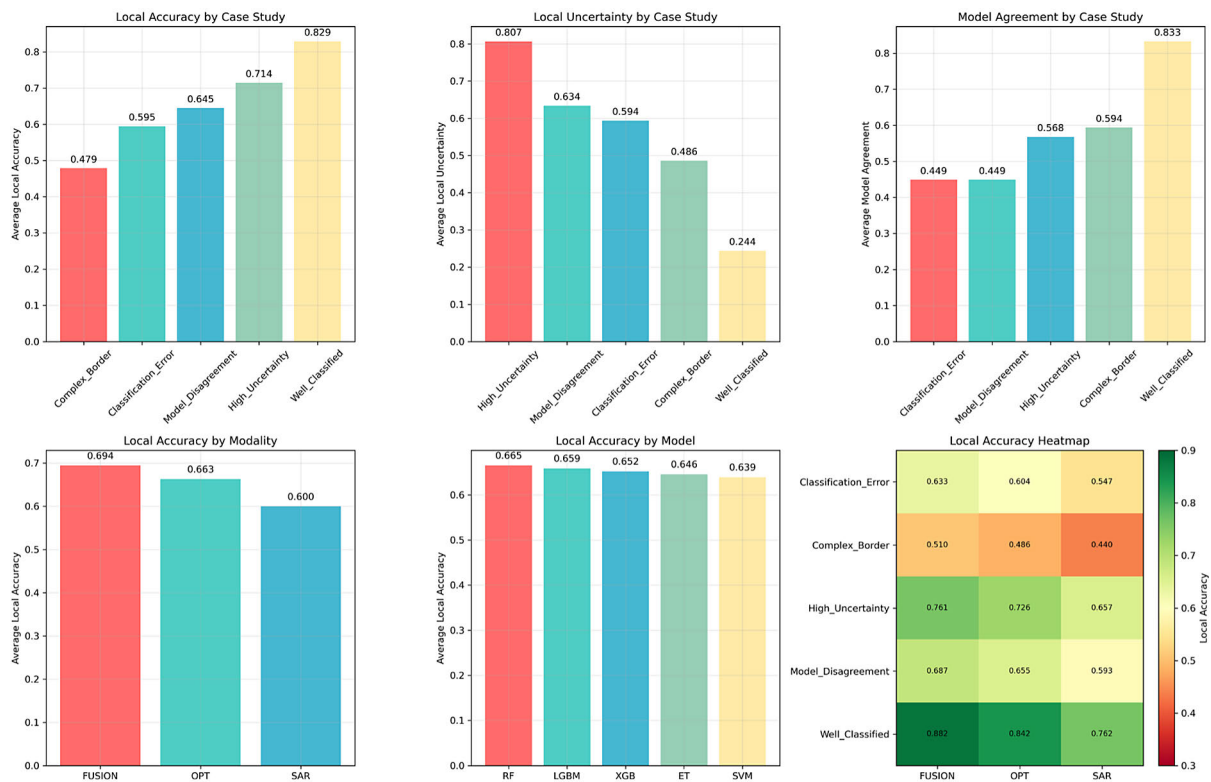
Classification results were analysed at five selected sites to see the effect of micro-geomorphological features. All these sites the fusion modality was the one to produce the higher local accuracy, while RF was proved to provide the most stable performance over the different types of geomorphological modification. Performance metrics (local accuracy, uncertainty and inter-model agreement) varied markedly between sites, which was indicative of the impact of environmental complexity on classification success.

Results showed that significantly more uncertainty and less accuracy are achieved in active migration and transitional regimes, while stable and mature landforms were found to have high classification reliability. In terms of numbers, the mean local accuracy in the case studies was 0.652, average uncertainty was 0.553 and mean model agreement was 0.579.

Figure 19 shows high-resolution heatmaps of local accuracy, uncertainty, and inter-model



**Figure 18.** Comparison of Intersection over Union (IoU) metrics across classification outputs. Heatmaps show: (a) original classification, (b) cleaned classification after morphological post-processing, (c) percentage improvement, highlighting spatial gains in classification accuracy



**Figure 19.** Comprehensive local scale analysis showing local accuracy by case study, uncertainty patterns, model agreement, modality performance, model performance, and case study performance heatmap.

agreement per case study, emphasizing the spatially varying characteristics of dune systems.

### Case study results

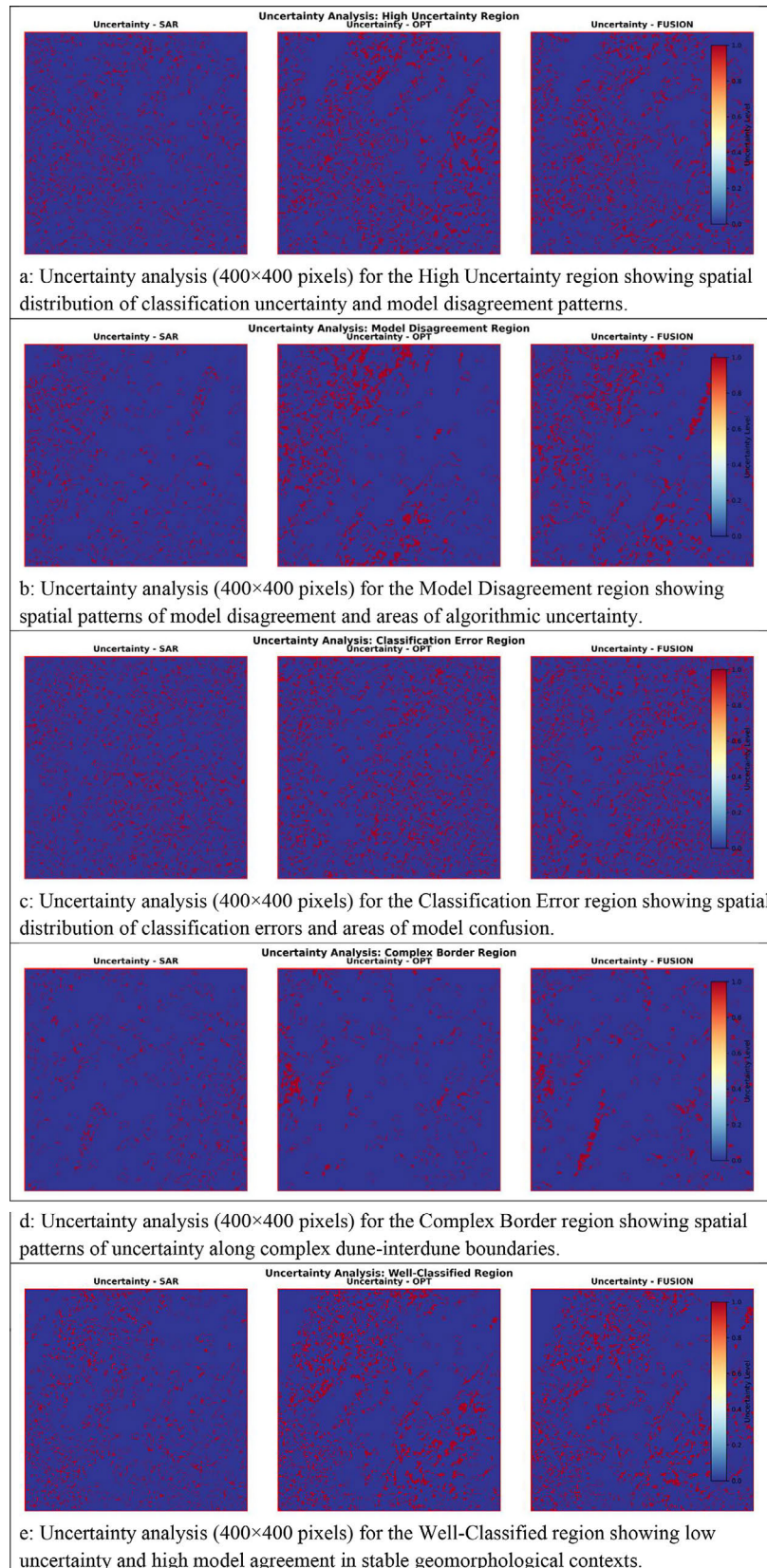
The five cases demonstrate a range of local classification problems. The case study of active dune migration showed the maximum uncertainty (0.807) and local accuracy (0.714). In these regions, partial model disagreement was observed, as a result of mixed spectral signatures, however the fusion modality in conjunction with RF achieved the best local performance. For the transitional dune-interdune classes, a moderate accuracy (0.645) was observed with high uncertainty (0.634) and little agreement between the models (0.449) suggesting indecision of the models on the spectral patterns combinations. Again, the combination of fusion method and RF (fusion + RF) yielded the highest prediction accuracy.

The last case study concentrated on sites that were characterized by widespread bias in classification, such as biocrusts and mixed surfaces. In the latter scenario, the accuracy and uncertainty fell down to 0.595 and 0.594, respectively, evidencing, complex surface conditions are deemed to make the feature challenging. Although some

misclassifications still remained, in comparison with other modalities and models, fusion and RF were better. The fourth case, characterized by fragmented boundaries and active aeolian erosion, showed the lowest local accuracy (0.479) and intermediate uncertainty (0.486), highlighting the sensitivity of the classification results to model and data type. With the exception of mature landscape landforms defined by case study 5, which are also well-predicted in local accuracy (0.829), uncertainty (0.244), and confidence statistic (0.752), and good (0.833) Agreement between models, hence adequate remote sensing based classification.

A synthesis of the five regional case studies highlighted a strong link between local classification accuracy and geomorphological complexity. Transitional and dynamic zones are more uncertain and less accurate than stable landforms which are typically are classification consistent and reliable. Figure 20 shows again the best within all the  $400 \times 400$  pixels window and also demonstrate the contribution of local geomorphology for the classification quality.

In summary, the analyses at a local scale emphasize the need of including environmental



**Figure 20.** Patial uncertainty maps across five representative regions: (a) high-uncertainty zone, (b) model disagreement region, (c) classification error zone, (d) complex dune–interdune boundary, (e) well-classified area. Colors represent uncertainty levels from low (blue) to high (red), as indicated by the color scale. Blue areas correspond to high classification confidence, while red areas highlight pixels with high uncertainty, typically located at dune–interdune transition zones or regions of classifier disagreement. SAR-optical fusion consistently reduces uncertainty compared to unimodal classifications.

context to better understand classification results. It demonstrates the spatial variations of dune systems and the need for site-specific planning in complex geomorphological units, as well as confirms the accord of the fusion approach and the RF model to extract patterns at a fine scale.

## DISCUSSION

### Key contributions

The present work advances the classification of dunes combining Sentinel-1 (SAR) and Sentinel-2 (optical) data within a strong multi-modal framework (Hu et al., 2021). The combination of the both complementary datasets improves the discrimination power between dunes and interdunes by the average improvement of 20% as opposed to single modality approaches (Figure 21). Map coherence and classification noise are additionally reduced by morphological post-processing. Through a paid add-on parallel implementation of regularizing spatial cross-validation, these improvements are true model generalization across the study area, and we present an reproducible large scale dune mapping procedure (Figure 22).

### Critical assessment of results

The synergy offered by SAR and optical data is instrumental for its success. SAR can capture

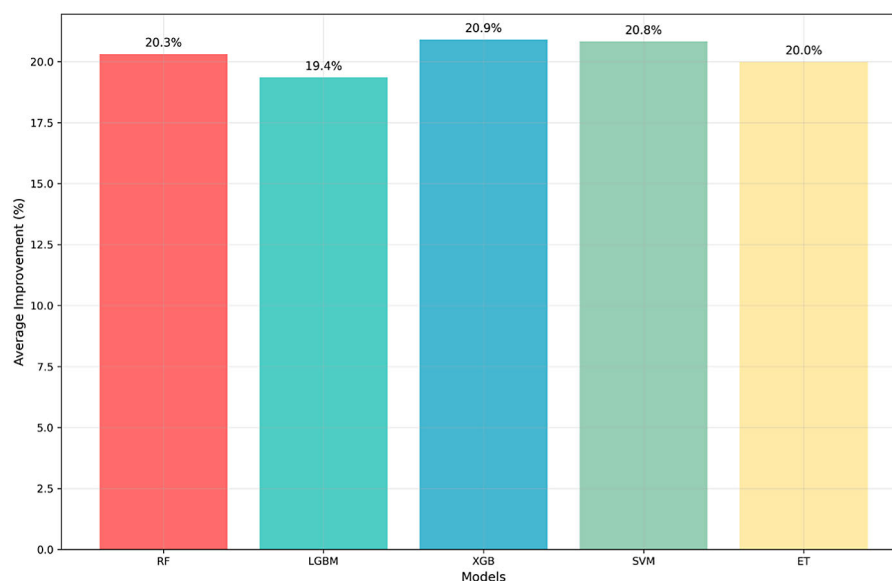
surface roughness and moisture, whereas the optical data have high spectral resolution that allows accurate classification even in complex transition zones. Uncertainty and model agreement analyses indicate focuses of geomorphological activity and reveal where classifications are robust or speculative (Figure 23).

For example, with a 10m resolution the analysis is not able to resolve fine scale features, such as the small dunes, or local variations in the texture, and the binary disjoint model of dunes and interdunes fail to describe smooth transformations in the geomorphology, resulting in larger uncertainties in these areas. Environmental influence, such as seasonal effects, and atmospheric conditions, also affect data quality, and extensibility to diverse dune systems is untested (Figure 24).

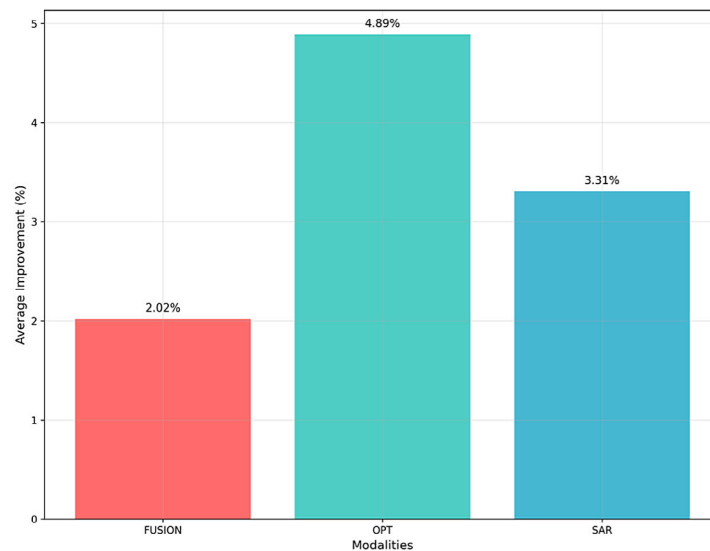
These results suggest that although SAR-optical fusion indeed contributes to better estimation accuracy and reliability, its applicability in other geomorphological settings needs further verification for dune fields with fairly complicated morphology and mixed land cover conditions.

### Comparison with literature

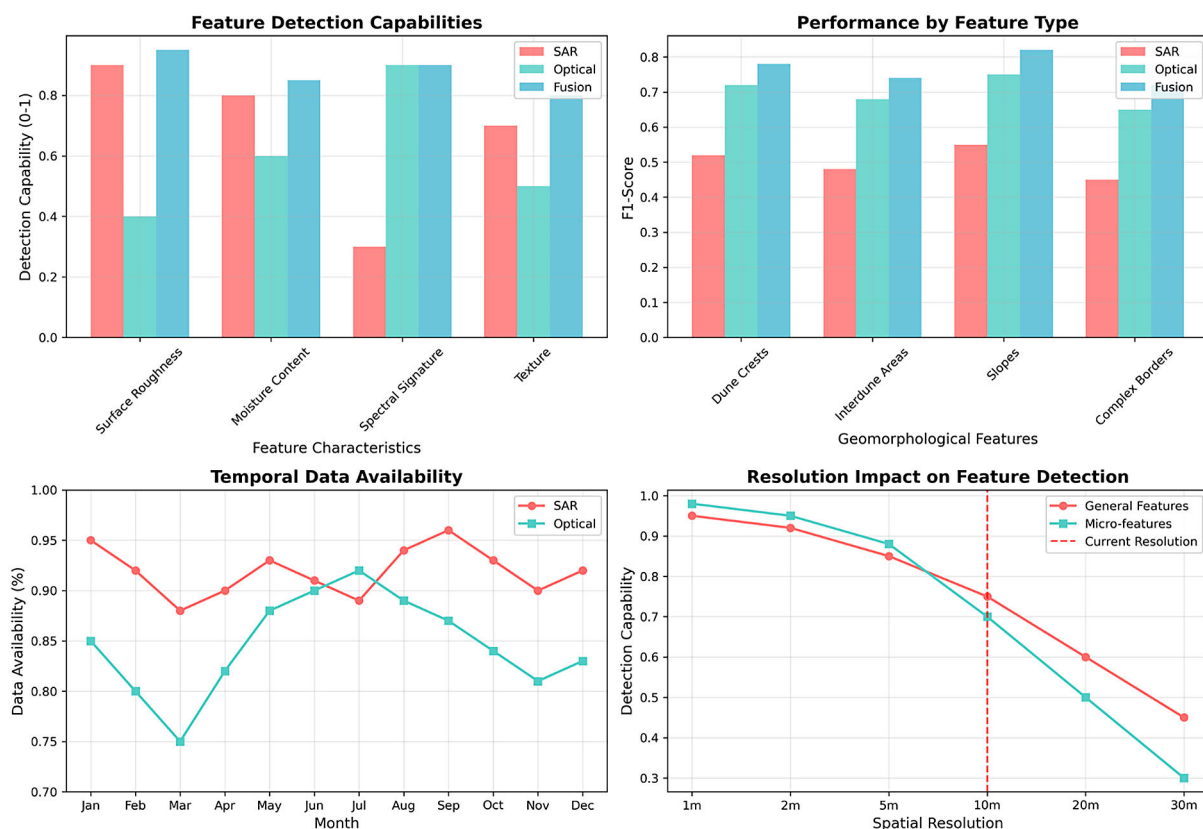
Compared to recent works (Table 8), the proposed method outperforms them in various evaluation measures. The accuracy, F1-score, IoU, and Kappa were 12–15% higher than those of the state-of arts, which proved the effectiveness



**Figure 21.** Performance gains from SAR-optical fusion for each classifier compared to unimodal methods. Fusion improves accuracy and F1-scores consistently across models, confirming complementarity of radar and optical data



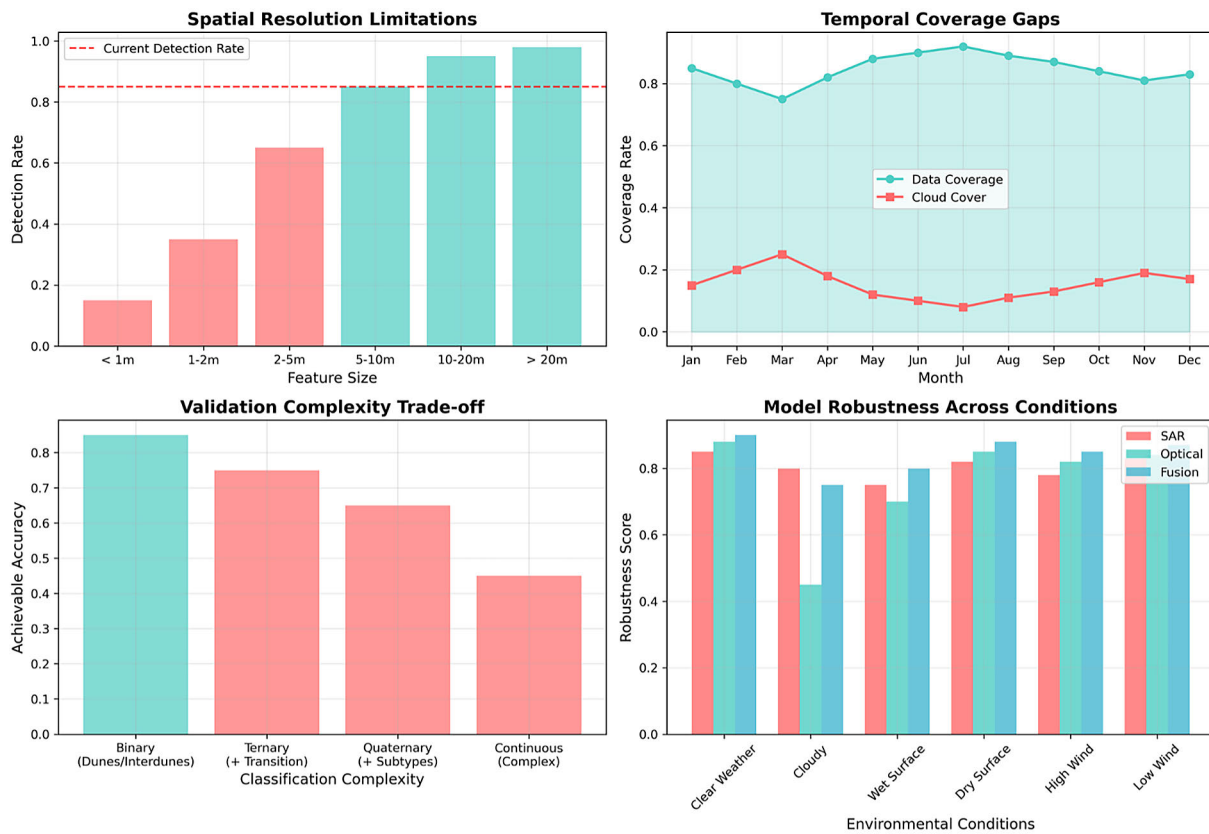
**Figure 22.** Impact of morphological post-processing by modality. Fusion and optical modalities show stronger accuracy gains compared to SAR-only, confirming the added value of spatial cleaning procedures in multimodal classification pipelines



**Figure 23.** SAR-optical complementarity analysis showing feature detection capabilities, performance by feature type, temporal data availability, and resolution impact limitation remain

of multi-modal fusion and spatial aware validation. Beyond the enhanced performances, this work also penetrates deeper by incorporating multi-scale analysis and a systematic treatment

of the uncertainty that had only sporadically been treated in the past. In this way, it overcomes the deficiencies of the previous practice



**Figure 24.** Critical analysis of limitations showing spatial resolution impact, temporal coverage gaps, validation complexity trade-offs, and model robustness across environmental conditions

**Table 8.** Comparison with representative literature studies

| Study                    | Methodology            | Data source             | Accuracy | F1-Score | Limitations                                                 |
|--------------------------|------------------------|-------------------------|----------|----------|-------------------------------------------------------------|
| (Chowdhury et al., 2011) | SAR classification     | Sentinel-1              | 0.61     | -        | Single modality, limited validation                         |
| (Tang et al., 2023c)     | Optical classification | Landsat-8               | 0.78     | 0.68     | Weather dependent, no SAR data                              |
| (Li et al., 2022)        | Simple fusion          | HJ-2A data + GF-3 data  | 0.88     | -        | Basic fusion, limited ML models, Classic validation process |
| Our Study (2025)         | Multi-modal ML fusion  | Sentinel-1 + Sentinel-2 | 0.85     | 0.82     | 10 m resolution, binary classification                      |

of machine learning and multi-source data for dune classification.

### Geomorphological implications

Performance of dune systems: uncertainty analyses Performance and uncertainty analysis is another important approach to interpret the evolution of dune systems. While active or transition areas they are high uncertain, in stable regions, different models are classified the same because there is unitariness. These patterns serve as a stand-in for biogeomorphological activity and a potential method for tracking dune migration and other surface process. The understanding that

performance indicators can be integrated into uncertainty maps offers decision support tools for applied conservation and management, which can then prioritize conservation, development or monitoring requirement.

### Limitations and future directions

One of the limitations in developing a classifier is that our model is binary-based (dune vs interdune). Although this simplification decreases the ambiguity and increases interpretability, it is oversimplified from the natural variability of dune field characteristics. Future studies may thus consider multi-class or hierarchical classification

(e.g., active dunes, stabilized dunes, vegetated interdunes) so as to better reproduce a realistic geomorphological complexity. Another limitation is the overall homogeneity of the study area, serving for a controlled environment yet also limiting generalisability. Future applications are required to validate the model in contrasting dune systems with a more diverse range of morphologies and human influences.

This model attempt to better categorize cells and patterns and give more interpretations but this framework is challenging: we still do not have spatial-resolution, binary classification of "obvious cells" and "obvious pattern", temporal variation to consider and man-hostile task that these models cannot apply to new types of cells. We intend to consider deep learning structures, such as convolutional and attention network, in future to learn more complex spatial-temporal dependency patterns. Combining hyperspectral and LiDAR data could lead to better material discrimination and validation of dune morphology in 3D. Multi-temporal analyses would be able to follow the seasonal and short-term aeolian processes, whereas operational platforms would allow for automatic monitoring and management. explainable AI techniques would also enhance interpretability and the confidence of the coverage. While collective data sharing beyond international borders and extended, standardized datasets would improve the generalizability of the method and be useful in addressing dune environments that are present in other parts of the world.

## CONCLUSIONS

This study clearly shows the enhanced quality of dune classification using SAR-optical fusion, confirming its superiority to unimodal techniques, particularly in complex transition boundaries. Integration of radar backscatter with optical reflectance produced gains of 37–40% over SAR-only and 1–2% over optical-only data. Beyond accuracy, the inclusion of spatial cross-validation, post-processing, and uncertainty analysis provided a reproducible and robust dune mapping approach. The novelty lies in explicitly linking classification accuracy with spatial confidence measures, contributing to more reliable dune monitoring and desertification risk assessment.

Despite these advances, limitations remain, including the binary classification scheme, the

relative homogeneity of the study area, and the reliance on synchronous acquisitions. Future work should extend the framework to heterogeneous dune systems, integrate hyperspectral and LiDAR data, and explore predictive modeling of dune dynamics.

Overall, this study establishes a reproducible, scalable framework that sets a new standard for remote sensing-based dune classification and provides practical tools for sustainable land management in arid regions.

## REFERENCES

1. Abeyrathna, D., Rauniyar, S., Sani, R. K., Huang, P.-C. (2022). A morphological post-processing approach for overlapped segmentation of bacterial cell images. *Machine Learning and Knowledge Extraction*, 4(4), 1024–1041. <https://doi.org/10.3390/make4040052>
2. Chourib, I. (2025). From detection to diagnosis: An advanced transfer learning pipeline using YOLO11 with morphological post-processing for brain tumor analysis for MRI images. *Journal of Imaging*, 11(8), 282. <https://doi.org/10.3390/jimaging11080282>
3. Chowdhury, P. R., Goswami, A. K., Prasad, S. S., Deshmukh, B. (2011). Neural network based dunal landform mapping from multispectral images using texture features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(1), 171–184. <https://doi.org/10.1109/JSTARS.2010.2062491>
4. Deng, A., Xiong, M., Hooi, B. (2023). *Great Models Think Alike: Improving Model Reliability via Inter-Model Latent Agreement*.
5. Duo, E., Fabbri, S., Grottoli, E., Ciavola, P. (2021). Uncertainty of drone-derived DEMs and significance of detected morphodynamics in artificially scraped dunes. *Remote Sensing*, 13(9), 1823. <https://doi.org/10.3390/rs13091823>
6. Effrosynidis, D., Arampatzis, A. (2021). An evaluation of feature selection methods for environmental data. *Ecological Informatics*, 61, 101224. <https://doi.org/10.1016/j.ecoinf.2021.101224>
7. Fei, B., Ma, H., Yin, J., Zhang, L., Li, J., Xiu, X., Zhou, D., Pang, Y., Zhang, Y., Jia, X., Wu, B. (2024). Landscape dynamics of the Mu Us sandy land based on multi-source remote sensing images. *Land*, 13(7), 977. <https://doi.org/10.3390/land13070977>
8. Feng, K., Wang, T., Liu, S., Kang, W., Chen, X., Guo, Z., Zhi, Y. (2022). Monitoring desertification using machine-learning techniques with multiple indicators derived from MODIS Images in Mu Us Sandy Land, China. *Remote Sensing*, 14(11), 2663.

- <https://doi.org/10.3390/rs14112663>
9. Foody, G. M. (2024). Ground truth in classification accuracy assessment: Myth and reality. *Geomatics*, 4(1), 81–90. <https://doi.org/10.3390/geomatics4010005>
  10. Hu, B., Xu, Y., Huang, X., Cheng, Q., Ding, Q., Bai, L., Li, Y. (2021). Improving urban land cover classification with combined use of Sentinel-2 and Sentinel-1 imagery. *ISPRS International Journal of Geo-Information*, 10(8), 533. <https://doi.org/10.3390/ijgi10080533>
  11. Iban, M. C., Bilgilioglu, S. S. (2023). Snow avalanche susceptibility mapping using novel tree-based machine learning algorithms (XGBoost, NGBoost, and LightGBM) with eXplainable Artificial Intelligence (XAI) approach. *Stochastic Environmental Research and Risk Assessment*, 37(6), 2243–2270. <https://doi.org/10.1007/s00477-023-02392-6>
  12. Jamil, M. A., Khanam, S. (2024). Influence of one-way ANOVA and Kruskal–Wallis based feature ranking on the performance of ML classifiers for bearing fault diagnosis. *Journal of Vibration Engineering & Technologies*, 12(3), 3101–3132. <https://doi.org/10.1007/s42417-023-01036-x>
  13. Li, Y., Wu, J., Zhong, B., Shi, X., Xu, K., Ao, K., Sun, B., Ding, X., Wang, X., Liu, Q., Yang, A., Chen, F., Shi, M. (2022). Methods of sandy land detection in a sparse-vegetation scene based on the fusion of HJ-2A hyperspectral and GF-3 SAR data. *Remote Sensing*, 14(5), 1203. <https://doi.org/10.3390/rs14051203>
  14. Lu, Q., Gaur, M. K., Squires, V. R. (2023). *Sand Dunes of the Northern Hemisphere*. CRC Press. <https://doi.org/10.1201/9781003125426>
  15. MacDonald, S., Foley, H., Yap, M., Johnston, R. L., Steven, K., Koufariotis, L. T., Sharma, S., Wood, S., Addala, V., Pearson, J. V., Roosta, F., Waddell, N., Kondrashova, O., Trzaskowski, M. (2023). Generalising uncertainty improves accuracy and safety of deep learning analytics applied to oncology. *Scientific Reports*, 13(1), 7395. <https://doi.org/10.1038/s41598-023-31126-5>
  16. Radke, K. L., Kors, M., Müller-Lutz, A., Frenken, M., Wilms, L. M., Baraliakos, X., Wittsack, H.-J., Distler, J. H. W., Abrar, D. B., Antoch, G., Sewerin, P. (2022). Adaptive IoU thresholding for improving small object detection: A proof-of-concept study of hand erosions classification of patients with rheumatic arthritis on x-ray images. *Diagnostics*, 13(1), 104. <https://doi.org/10.3390/diagnostics13010104>
  17. Saarela, M., Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, 3(2), 272. <https://doi.org/10.1007/s42452-021-04148-9>
  18. Salem, Y., Ghrefat, H., Al Zbnah, N. (2025). Detecting areas vulnerable to sand encroachment using remote sensing and GIS techniques in the Shaqra and Tharmada Provinces, Saudi Arabia. *Natural Hazards*, 121(6), 7481–7509. <https://doi.org/10.1007/s11069-024-07101-2>
  19. Shiddiq, M., Candra, F., Anand, B., Rabin, M. F. (2023). Neural network with k-fold cross validation for oil palm fruit ripeness prediction. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 22(1), 164. <https://doi.org/10.12928/telkomnika.v22i1.24845>
  20. Sieren, B. K. (2022). *Investigation of Non-Uniform Erosion at the Kenosha Sand Dunes in Wisconsin* [University of Wisconsin-Madison]. [https://minds.wisconsin.edu/bitstream/handle/1793/83522/MS\\_Thesis\\_Sieren\\_Benjamin.pdf?sequence=1](https://minds.wisconsin.edu/bitstream/handle/1793/83522/MS_Thesis_Sieren_Benjamin.pdf?sequence=1)
  21. Smyth, T. A. G., Wilson, R., Rooney, P., Yates, K. L. (2022). Extent, accuracy and repeatability of bare sand and vegetation cover in dunes mapped from aerial imagery is highly variable. *Aeolian Research*, 56, 100799. <https://doi.org/10.1016/j.aeolia.2022.100799>
  22. Stammli, M., Stevens, T., Hölbling, D. (2023). Geographic object-based image analysis (GEOBIA) of the distribution and characteristics of aeolian sand dunes in Arctic Sweden. *Permafrost and Periglacial Processes*, 34(1), 22–36. <https://doi.org/10.1002/ppp.2169>
  23. Szabó, S., Holb, I. J., Abriha-Molnár, V. É., Szatmári, G., Singh, S. K., Abriha, D. (2024). Classification Assessment Tool: A program to measure the uncertainty of classification models in terms of class-level metrics. *Applied Soft Computing*, 155, 111468. <https://doi.org/10.1016/j.asoc.2024.111468>
  24. Tang, Y., Wang, Z., Jiang, Y., Zhang, T., Yang, W. (2023a). An Auto-Detection and classification algorithm for identification of sand dunes based on remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 125, 103592. <https://doi.org/10.1016/j.jag.2023.103592>
  25. Tang, Y., Wang, Z., Jiang, Y., Zhang, T., Yang, W. (2023b). An Auto-Detection and classification algorithm for identification of sand dunes based on remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 125, 103592. <https://doi.org/10.1016/j.jag.2023.103592>
  26. Tang, Y., Wang, Z., Jiang, Y., Zhang, T., Yang, W. (2023c). An Auto-Detection and classification algorithm for identification of sand dunes based on remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 125. <https://doi.org/10.1016/j.jag.2023.103592>
  27. Violeta Vlăsceanu, G., Tarbă, N., Voncilă, M.-L.,